Why payment card fees are biased against retailers

Julian Wright*

June 2012

Abstract

I formalize the popular argument that retailers pay too much and cardholders too little to make use of payment card platforms, resulting in excessive use of cards. To do this, I analyze a standard two-sided market model of a payment card platform. With minimal additional restrictions, the model implies that the privately set fee structure is unambiguously biased against retailers in favor of cardholders, a result that continues to hold even if the platform can perfectly price discriminate on both sides. The market failure arising is primarily a regulatory problem and does not raise any competition concerns.

1 Introduction

MasterCard may be "Priceless" and Visa "Everywhere You Want To Be" but apparently only from the point of view of cardholders. These payment platforms have increasingly found themselves under fire from large retailers and policymakers, with regulatory and legal proceedings

*Department of Economics. National University of Singapore: email jwright@nus.edu.sg. For helpful comments and suggestions, I am thankful to Luca Anderlini, Lawrence Ausubel, Emilio Calvano, Sujit Chakravorti, Joshua Gans, Roger Lagunoff, Eric Ralph, Jean-Charles Rochet, Ding Rong, Marc Rysman, Marius Schwartz, Glen Weyl, seminar participants from the Bank of Canada, Federal Reserve Bank of Kansas City, Georgetown University and the University of Maryland, and the editor and two anonymous referees. Although I have consulted for Visa before, this article was not funded in any way by Visa and the views expressed here are strictly my own.

arising in more than 30 different countries.[1] Public authorities such as the Reserve Bank of Australia, the European Commission and the United States Government Accountability Office, together with a number of economists (e.g., Carlton and Frankel, 1995; Katz, 2001; Cabral, 2005; Vickers, 2005; Farrell, 2006) have argued that the fee structure in debit and credit cards are likely to be distorted, with retailers paying too much to accept payment cards and cardholders paying too little compared to the efficient fee structure, resulting in excessive usage of payment cards by consumers, a cost which is ultimately passed on to consumers paying by cash. Some have argued further (e.g., Vickers, 2005) that retailers accept cards that raise their costs as the alternative, rejecting payment cards, is not a viable option for retailers given customers expect to be able to use these cards, the so called "must-take" cards argument.

The rapidly expanding literature on multi-sided platforms (e.g., Rochet and Tirole, 2003; Armstrong, 2006; Rysman, 2009; Weyl 2010) has cast some doubt over these concerns, noting that skewed pricing often arises (e.g., in shopping malls, Yellow Pages directories, and Internet portals) and that this can simply reflect asymmetries in externalities across the multiple sides. In general models of interactions between cardholders and retailers (Schmalensee, 2002), consumers and websites (Laffont *et. al.*, 2003), buyers and sellers (Rochet and Tirole, 2003), and callers and receivers (Hermalin and Katz, 2004), in which a monopoly platform faces downward sloping demand from users on each side, each of the studies has shown there is no systematic bias towards one side or the other. Rochet and Tirole (2011) reach this conclusion using a fully fledged model of cards in which there is elastic demand from users on both sides, noting in this case (p. 485) "*... the price structure chosen by a monopoly platform, in the absence of a regulation, is no longer systematically biased in favor of cardholders.*" Recent surveys of the theoretical literature

[1]Recent or ongoing investigations include those in Argentina, Australia, Brazil, Canada, Chile, Colombia, Czech Republic, Denmark, the European Union, France, Germany, Honduras, Hong Kong, Hungary, Israel, Italy, Mexico, the Netherlands, New Zealand, Norway, Poland, Romania, Singapore, South Africa, Spain, Sweden, Switzerland, Turkey, Venezuela, the United Kingdom, and the United States. Bradford and Hayashi (2008) contain references for many of these investigations.

on payment cards come to a similar conclusion (e.g., see Bolt and Chakravorti, 2008, Prager *et al.*, 2009, and Verdier, 2010).

The point of this article is to challenge this conclusion in the case of payment cards. To proceed, I take a standard model of a card platform facing elastic demand for cards on each side of the market (i.e., the same model considered by Rochet and Tirole, 2011, in reaching their conclusion above) and show it actually implies an unambiguous bias against retailers, even when there is no possibility of price discrimination on either side of the market. Specifically, the interchange fee determined by the platform will be excessive.[2] Reducing the amount retailers pay and making cardholders pay more will, up to some point, raise welfare. I show conditions under which this result continues to hold even when the card platform can perfectly price discriminate on both sides.

Contrary to the claims in previous work on the topic (including some of my own), these strong results do not depend on the relative level of cardholder and retailer benefits from different payment instruments. Nor do they rely on there being any revenue shifting role for interchange fees, in which it is argued the interchange fee shifts revenues to the cardholder side of the business as this is the side where the revenues are competed away less (although the bias continues to hold when revenue shifting also arises). Rather, the results arise by exploiting the full implications of a phenomenon known in the literature as *merchant internalization* (Rochet and Tirole, 2002, Farrell, 2006) together with the property of price coherence (Frankel, 1998) in which retailers set a single price regardless of how consumers pay.

Merchant internalization is the property that when deciding whether to join a platform (e.g., whether to accept cards), sellers on one side (known as merchants in the cards literature) take into account the benefits buyers on the other side get from being able to interact with them on the platform (i.e., the benefits cardholders get from using cards including potentially, rewards and interest-free benefits that cardholders receive) given this allows sellers to charge a higher

[2] The interchange fee is an interbank transfer from the retailer's bank, known as the acquirer, to the cardholder's bank, known as the issuer, when card transactions involve a cardholder and retailer belonging to different banks.

price to buyers or to attract more business at the same price. As explained by Rochet and Tirole (2011) in the context of payment cards, merchant internalization holds for a wide range of different market structures. Merchant internalization together with price coherence imply that although consumers ignore the benefits retailers get from accepting cards, the reverse is not true.[3] These properties are assumed by Rochet and Tirole (2002, 2011), Wright (2004) and Guthrie and Wright (2007) among others. They can also arise for other types of platforms. One example is a content provider deciding whether to subscribe to an ISP that offers its customers faster downloading speeds for its content when they also subscribe to the same ISP, assuming the content provider has a single price for its content regardless of which ISP its customers use. Another example is a retail chain that adopts a "one price policy" and has to decide whether to locate in a shopping mall that offers customers a superior or more convenient shopping experience.

The reason for the result that the unregulated fee structure is biased against retailers (or privately determined interchange fees are excessive) is easiest to understand in the case of a monopoly platform that is able to perfectly price discriminate across the two sides. Perfect price discrimination implies the platform fully takes into account users' average willingness to pay on each side, in order to capture maximum profit. Merchant internalization implies the amount each retailer is willing to pay to accept cards takes into account the average surplus its customers expect to get from using cards. Provided consumers face the same retail price for goods regardless of how they pay, this surplus also determines what consumers are willing to pay to hold the card in the first place. Thus, the consumers' surplus from card usage gets counted twice, once when the platform extracts surplus from each retailer that accepts cards and once

[3] In contrast if items are sold through a process of negotiation between buyers and sellers rather than with posted prices, so that the resulting price can depend on whether buyers make use of the platform to interact with sellers, there is no reason to think there would be any asymmetry in the extent to which each side internalizes the others' benefits, or indeed, that the structure of platform fees across the two sides would matter at all (Gans and King, 2003).

when it extracts surplus from the consumers who hold cards. The resulting fee structure favors cardholders and is biased against retailers.

For a platform that cannot price discriminate on either side, the logic behind the bias is more subtle but holds more generally (i.e., even if consumers do not expect any positive surplus from holding a card). The logic can be seen most directly in the special case in which cardholders and retailers draw the convenience benefits of using cards from the same distribution. By symmetry, welfare maximization calls for an interchange fee that equalizes the benefits of the marginal user on each side. As retailers internalize the benefits of cardholders in their card acceptance decision, this requires retailers to be charged more than cardholders. However, a monopoly platform will want to shift even more charges from cardholders to retailers, beyond the point where the benefits of the marginal users are equalized. This reflects that retailers' participation on the card platform is also relatively insensitive to a change in the interchange fee, as a higher interchange fee raises the fees charged to retailers but it also lowers the cardholders' rewards, which the retailer takes into account.

More generally, the fact that the monopolist platform only focuses on marginal users and not average users means it ignores the effect of its choice of interchange fees on the surplus of the average user on each side. As this happens on both sides, without merchant internalization there is no particular bias. However, due to merchant internalization, the marginal retailer already internalizes the surplus of the average card user, so it is only the average retailer's surplus that is not taken into account by the monopolist. Taking this into account calls for the structure of fees to be rebalanced in favor of retailers (i.e., a lower interchange fee).

The importance of merchant internalization in understanding biases in payments has been emphasized by others since the seminal work of Rochet and Tirole (2002), who were the first to provide a fully fledged model of payments based on cardholders and price-posting retailers. However, merchant internalization works for quite different reasons in Rochet and Tirole (2002) than in the present article, reducing retailers' resistance to accepting cards, meaning an issuer controlled card platform that can perfectly price discriminate across retailers but not across cardholders ends up in a corner solution where the interchange fee is either inefficiently high (i.e., so fees are biased against retailers) or is constrained efficient.[4] The case in which fees are biased against retailers therefore arises in their model when there is perfect price discrimination on one side (due to their assumption retailers are all homogenous) and not the other.

Bedre-Defolie and Calvano (2010) also exploit an asymmetry in price discrimination possibilities to show a monopoly card platform's interchange fee is inefficiently high. Using a two-part tariff, a monopoly issuer is able to fully extract the surplus cardholders get from using cards, given consumers decide both whether to hold a card and independently, when to use it. On the other hand, with unobservable heterogeneity in retailers, no price discrimination is allowed on the retailer side. Given the authors shut down merchant internalization, a monopoly card platform will take into account the surplus card usage generates for the average cardholder due to price discrimination on the cardholder side, but not on the retailer side, resulting in a bias against retailers.

In contrast to the biases arising in Rochet and Tirole (2002) and Bedre-Defolie and Calvano (2010), I show the bias against retailers arises without having to assume price discrimination is possible on one side and not the other. Given in practice issuers offer a range of different non-linear packages to cardholders and acquirers set different fees to different types of retailers reflecting that interchange fees differ by industry category, assuming price discrimination possibilities that are completely asymmetric does not seem reasonable. My results show the bias arises when price discrimination possibilities are fully symmetric, either because the card platform cannot internalize the surplus of average users on either side, or can do so perfectly on both sides.

A critical assumption for the results, as with results in the existing literature, is the lack of

[4]In a quite different framework with elastic goods demand and heterogeneous card issuers, but without merchant internalization, Wang (2010) also finds the privately set interchange fee is either equal to the efficient level or is too high.

retailer surcharging (i.e., price coherence). Gans and King (2003) show in a very general way that if retailers can costlessly surcharge, then interchange fees will be neutral, meaning they have no real effects. The card platforms' own no-surcharge rule is one reason to consider the case retailers do not surcharge. In jurisdictions where such rules have been banned, many retailers still do not surcharge. My results continue to hold if the reason only some retailers surcharge whereas others do not is determined by some exogenous factors. At the end of Section 4, I discuss other more realistic cases.

The rest of the article proceeds as follows. Section 2 lays out some preliminary model details that apply throughout. Section 3 gives some insight into the article's main result by contrasting the lack of any systematic bias in fee structure in case of a symmetric two-sided platform setting (with no payment between the two sides) to the bias that arises in the same setting but in which merchant internalization and price coherence also hold. Section 4 then lays out the specific benchmark model of payment cards, following Rochet and Tirole (2011), showing the systematic bias that arises. It also shows the conclusion is robust to asymmetry in pass-through rates, imperfect merchant internalization, price discrimination, and a simple extension allowing some exogenous fraction of retailers to surcharge. Section 5 concludes with a discussion of policy implications, including implications for another multi-sided platform setting.

2 Preliminaries

Consider a monopoly platform (e.g., a card platform) which provides some interaction benefits to users on two sides if they use the platform to interact. These benefits are to be interpreted as the additional benefits of interacting on the platform as opposed to using an alternative means to interact (e.g., cash). In keeping with the existing literature starting with Rochet and Tirole (2002, 2003), I will use B (buyers) and S (sellers) to refer to the two types of users. In other contexts they may correspond to callers and receivers (Hermalin and Katz, 2004) or consumers and websites (Laffont *et. al.*, 2003). The cost to the platform of enabling an interaction is c.

A representative buyer and seller wish to interact, each of whom has access to a platform which they can use for the interaction. Users on each side i draw an interaction benefit of using the platform b_i on the interval $\left[\underline{b}_i, \overline{b}_i\right]$ from a continuously differentiable distribution function G_i, with density g_i, and decide whether to use the platform or not. If the parties on both sides want to use the platform, then the interaction occurs on the platform and both users get the associated benefits. If either party does not want to use the platform, then neither side will obtain these benefits (i.e., both obtain zero benefits). Neither side observes the other side's particular draw of interaction benefits, and b_B and b_S are independent draws.

Users B and S face per interaction prices of p_B and p_S for using the platform. For the moment, consider the case the prices p_B and p_S are set directly by the platform. The case in which they are not (e.g., as with Visa) will be discussed in Section 4. For each $i = B, S$, let $D_i(.) = 1 - G_i(.)$ be the survival function of user i's interaction benefit which is also known as its quasi-demand in the payments literature (as actual demand depends on the other sides' willingness to interact). Consistent with Rochet and Tirole (2003) and Bedre-Defolie and Calvano (2010), assume that $D_i(.)$ is log-concave (or equivalently, the hazard rate of G_i is increasing). This property holds for extreme-value, logistic, normal, and uniform distributions, among others (see Bagnoli and Bergstrom, 2005). Let $\widehat{b}_i$ denote the interaction benefits of a marginal user on side i. The difference between the average and marginal user's interaction benefit, which I call i's inframarginal surplus per interaction, is defined as $v_i\left(\widehat{b}_i\right) = E\left(b_i | b_i \geq \widehat{b}_i\right) - \widehat{b}_i$. It plays a critical role in the analysis that follows. If $\widehat{b}_i = p_i$, then $v_i(p_i)$ is just the user surplus to i per interaction, and expected user surplus to a user on side i from being able to use the platform to interact with a user on the other side $V_i(p_i) = D_i(p_i)\, v_i(p_i)$ can equivalently be written as consumer surplus $V_i(p_i) = \int_{p_i}^{\overline{b}_i} D_i(z)\, dz$.

Consider now three different settings for how B and S interact, the last of which is the focus of this article.

Case (i): No payment between the two sides Suppose there is no payment between B and S associated with the interaction. The platform will be used for interactions if and only if $b_B \geq p_B$ and $b_S \geq p_S$. In comparison, the first-best outcome requires any interaction for which $b_B + b_S \geq c$ be carried out. Deadweight loss will arise given some possible interactions involving very high interaction benefits to one side and low interaction benefits to the other will not be carried out. Efficient prices p_B and p_S will then minimize such deadweight loss. This model is essentially the same as the one-way calling model of Hermalin and Katz (2004), the model without micropayments between websites and consumers in Laffont *et. al.* (2003), or the model of payments without merchant internalization as in Schmalensee (2002). There is no particular bias in the structure of prices towards one side or the other under these assumptions, as will be shown in Section 3.

Case (ii): Frictionless payments between the two sides Suppose payments between B and S are allowed, without any frictions. Then as shown by Laffont *et. al.* (2003) when they allow for micropayments, by Gans and King (2003) in the payments literature, and by Rochet and Tirole (2006) when discussing two-sidedness more generally, how much the platform charges to one side versus the other will become irrelevant. In the case of payments, this is true if the seller can costlessly add a surcharge when buyers pay using cards. In such a setting it makes no sense to say the price structure is biased against one side or the other — the interchange fee will be neutral. Interactions will arise if and only if $b_B + b_S \geq p_B + p_S$, so the first-best outcome can be achieved if $p_B + p_S = c$.

Case (iii): Uniform pricing by sellers and merchant internalization Now suppose there are payments between B and S as above but B faces the same price from S whether B interacts on the platform or off the platform. In the case of a card platform, this could be due to the platform's no-surcharge rule that prevents sellers from adding a surcharge when buyers pay using cards. Other possible reasons are discussed at the end of Section 4. Uniform pricing (or price coherence) has been the standard assumption in the majority of the payments literature. Facing a single price, B will want to use the platform to interact with S if and only if $b_B \geq p_B$. Assume a particular form of *merchant internalization* — that S will adopt the platform (e.g., accept cards) if and only if

$$v_B(p_B) \geq p_S - b_S. \tag{1}$$

Rochet and Tirole (2011) refer to this as Assumption 1. I will call it full merchant internalization since B's expected surplus per interaction is fully taken into account by S in its decision of whether to adopt the platform. It implies the seller compares the additional expected user surplus per interaction that is created $v_B(p_B)$ with the additional cost $p_S - b_S$ per interaction – provided this is non-negative, S will want to adopt the platform. The interaction benefit of the marginal seller is $\widehat{b}_S = p_S - v_B(p_B)$. This can arise, for example, if by adopting the platform S is able to capture B's expected user surplus from being able to interact with it on the platform through a higher price (or higher market share at the same price). Rochet and Tirole (2011) show (1) arises when sellers are perfectly competitive or when they compete in Hotelling-Lerner-Salop differentiated products competition. Wright (2010) shows the assumption hold with Cournot competition with elastic goods demand. Appendix A establishes (1) for several new settings. Farrell (2006, p.37) argues full merchant internalization applies as an approximation to the behavior of optimizing sellers much more generally.

I close this section by introducing the various objective functions that will be relevant in the article. Whether the platform sets the prices p_B and p_S directly, or determines them indirectly through an interchange fee (as in Section 4), its objective is assumed to be to maximize profit

$$\Pi = (p_B + p_S - c)\, D_B\left(\widehat{b}_B\right) D_S\left(\widehat{b}_S\right). \tag{2}$$

The social planner's objective is similarly to maximize expected welfare

$$W = \int_{\widehat{b}_B}^{\bar{b}_B} \int_{\widehat{b}_S}^{\bar{b}_S} (b_B + b_S - c)\, dG_S(b_S)\, dG_B(b_B), \tag{3}$$

as social surplus is generated whenever a buyer and seller obtain joint interaction benefits that exceed the joint cost of using the platform. Finally, define expected user surplus as

$$U = \int_{\widehat{b}_B}^{\bar{b}_B} \int_{\widehat{b}_S}^{\bar{b}_S} (b_B + b_S - p_B - p_S)\, dG_S(b_S)\, dG_B(b_B). \tag{4}$$

3 A symmetric case

Before turning to the formal analysis of a payment card platform that sets an interchange fee (Section 4), this section briefly considers the case of a generic monopoly platform that enables interactions between two users for the special case when the interaction benefits are drawn from the same log-concave distribution across the two sides (i.e., the symmetric case). This sheds some light on the role played by merchant internalization in generating a bias in the structure of prices charged by the platform. In the case of payments, this setting can capture a three-party card platform like American Express which deals directly with cardholders and retailers.

To start with, consider case (i) of Section 2 with no payment between the two sides. Welfare can be written as

$$W = D_S(p_S) V_B(p_B) + D_B(p_B) V_S(p_S) + \Pi.$$

Using that $V_i'(p_i) = -D_i(p_i)$ and assume there is a binding breakeven constraint, welfare maximization arises when

$$D_S'(p_S) V_B(p_B) - D_B'(p_B) V_S(p_S) = 0, \tag{5}$$

as emphasized by the existing literature (see, for example, equation (2) in Hermalin and Katz, 2004). The increase in p_S reduces welfare because it reduces the probability S will want to interact, which is multiplied by the expected surplus of B from interacting with S. The corresponding decrease in p_B increases welfare because it increases the probability B will want to interact, which is multiplied by the expected surplus of S from interacting with B. Welfare is maximized subject to a breakeven constraint when these two effects exactly balance, which in the symmetric case implies $p_B = p_S = c/2$. This will maximize the probability of an interaction between B and S, and equate the surpluses obtained by the two sides. Because in the symmetric case the probability of an interaction is maximized at equal prices, the equality of prices across the two sides will also be a feature of the profit maximizing platform's choice of price structure, which for a given total price is characterized by

$$D_S'(p_S)\,D_B(p_B) - D_B'(p_B)\,D_S(p_S) = 0. \tag{6}$$

Thus, without merchant internalization, in the symmetric case the optimal price structure of equal prices on both sides is the same whether one maximizes welfare subject to a breakeven constraint or profit.

How does the previous logic change once merchant internalization is added? Recall that due to merchant internalization, $\widehat{b}_S = p_S - v_B(p_B)$, so that it will no longer be socially optimal to simply equate prices across the two sides even under symmetry. Rather, social optimality requires equating the benefits of the marginal users so that $p_B = \widehat{b}_S$. Thus, for the symmetric case, constrained welfare maximization requires the price to B to be lower than to S by $v_B(p_B)$, the extent to which S takes into account the surplus of B per interaction.

The profit maximizing platform does not take into account the impact of its prices on the relative surpluses of the two sides. For any given total price, it just seeks to maximize the probability of an interaction between B and S, as before. The first order condition for this is

$$D_S'\left(\widehat{b}_S\right) D_B(p_B)\left(1 + v_B'(p_B)\right) - D_B'(p_B)\,D_S\left(\widehat{b}_S\right) = 0, \tag{7}$$

which is a modified version of (6). The difference arises from the fact that an increase in p_S and equal decrease in p_B changes the net price (or effective price) faced by S by only $(1 - |v_B'(p_B)|)$ times the change in p_S, where note $0 < |v_B'(p_B)| < 1$. The user S is relatively insensitive to a change in the price structure given it takes into account B is better off when S pays more of the total price. In the symmetric case, (7) requires $p_B < \widehat{b}_S$ so that the benefits of the marginal users are no longer equalized across the two sides as they were for social optimality. Thus, at least in the symmetric case, the result suggests there is a systematic distortion between the price structure set by a social planner and a monopoly platform, with the monopoly platform loading too much of the total price on S relative to B, beyond the point where the benefits of the marginal users are equalized. The next section will formalize this insight in the context of a card platform and generalize the result beyond the simple symmetric case considered in this section.

4 A card platform

In order to obtain formal results on interchange fees to contrast with the existing literature on payment cards, in this section I start by putting the previous generic model of a two-sided platform into the context of a four-party (or open) card platform like Visa. Interactions on the platform are card transactions between buyers (i.e., cardholders) and sellers (i.e., retailers). The interaction benefits b_B and b_S are the convenience benefits to each of the two parties of transacting with a payment card relative to the alternative (say cash).

Suppose there is a continuum of subsectors or industries each corresponding to a particular draw of b_S. Industries can be modelled as having a single seller or multiple competing sellers. Goods are independent across industries, and the buyers' distribution of convenience benefits G_B is independent of the industry they buy in. Sellers observe their industry b_S before deciding whether to accept cards. Buyers, which all hold the payment card, must decide which seller to buy from (if there is more than one) in each industry before knowing their particular draw of b_B which is only realized at the point of sale. They wish to buy one unit of the good from each industry. Assume case (iii) of Section 2, that sellers set uniform prices regardless of how consumers pay and (1) holds.

A card platform such as Visa does not set the fees p_B and p_S directly but rather influences these fees by setting an interchange fee a which is a fee paid by the retailer's bank (the acquirer) and is received by the cardholder's bank (the issuer). Assume equilibrium card fees set by issuers and acquirers are p_B and p_S which are continuously differentiable functions of a, with $p_B' < 0$,

$p_S' > 0$ and $p_B + p_S > c$. The card platform is assumed to choose a to maximize the collective profit of issuers and acquirers, as given in (2).[5] These are the usual assumptions adopted in the payments literature (see Rochet and Tirole, 2002, 2011).

The relevant range of interchange fees for analysis is defined by the interval $[\underline{a}_B, \overline{a}_S]$, where $p_B(\underline{a}_B) = \overline{b}_B$, $p_B(\overline{a}_B) = \underline{b}_B$, $\widehat{b}_S(\underline{a}_S) = \underline{b}_S$, $\widehat{b}_S(\overline{a}_S) = \overline{b}_S$ and $\underline{a}_S \leq \underline{a}_B < \overline{a}_S \leq \overline{a}_B$. This is the simplest way to ensure that the privately and socially optimal interchange fee does not involve a corner solution (at a point where a user on one side will always use the platform). It requires that users on each side are sometimes sufficiently resistant to using the platform.

Denote the total number of card transactions as T, so

$$T = D_B(p_B) D_S\left(\widehat{b}_S\right).$$

Assume throughout that the possible objective functions T, Π, U, and W are log-concave in a over the relevant range of interchange fees $[\underline{a}_B, \overline{a}_S]$, so that the first order conditions with respect to a characterize the respective maximums. Given the log-concavity of $D_i(.)$, the appendix establishes sufficient conditions for log-concavity of these objective functions. An example satisfying these conditions arises if issuing and acquiring margins are linear in interchange fees (i.e., constant pass-through rates), and b_B and b_S are drawn from generalized Pareto distributions with increasing hazard rates (see Appendix B).

Comparison to existing interchange-fee results

Existing models that allow for seller heterogeneity such as Wright (2004) and Rochet and Tirole (2011) use essentially the same set of assumptions as the model above. Despite this, they obtain ambiguous results on whether privately set interchange fees are too high or too low. I start by noting their ambiguous findings. I then show that reinterpreting their results for the case

[5] This reflects that in practice card platforms levy various network fees on issuers and acquirers, and the amount that they can extract from these fees should be closely related to issuers' and acquirers' profit from card transactions.

considered in Section 3 in which users on both sides draw benefits from the same distribution implies interchange fees are actually too high. The subsequent section will then show how the same result can be obtained even after relaxing this strong assumption.

Wright (2004) and Rochet and Tirole (2011) focus on the benchmark case in which the pass through of interchange fees is symmetric so that $p'_S = -p'_B > 0$, and the sum of fees does not change when the interchange fee changes. I will initially adopt the same assumption, which will be relaxed subsequently. Under this assumption, a card platform that maximizes profit also maximizes the total number of card transactions T. Proposition 11 of Rochet and Tirole (2011) states that the profit maximizing interchange fee will be too high if and only if at the profit maximizing interchange fee, the buyers' inframarginal surplus per card transaction is more than that of sellers at the profit maximizing interchange fee (i.e., $v_B\left(p_B\right) > v_S\left(\widehat{b}_S\right)$). They claim this shows (p.485) "*the price structure is no longer systematically biased in favor of cardholders.*" Wright (2004) shows an equivalent result and makes a similar claim.

In the symmetric case of Section 3, a card platform that maximizes the number of card transactions will choose an interchange fee such that $p_B < \widehat{b}_S$, as the card platform prefers to load more of the fees on sellers, beyond the point where the benefits of the marginal users are equalized. Log-concavity of the quasi-demand functions thus implies $v_B\left(p_B\right) > v_S\left(\widehat{b}_S\right)$ at the profit maximizing interchange fee, given symmetry implies $v_B = v_S$. Proposition 11 of Rochet and Tirole (2011) therefore implies that the privately set interchange fee is necessarily too high. The result is stated here formally (all proofs of propositions are in Appendix B).

Proposition 1 *If the issuing margin is constant and acquiring is perfectly competitive (as assumed in Proposition 11 of Rochet and Tirole, 2011), and the convenience benefits for buyers and sellers are drawn from the same distribution, then the privately set interchange fee strictly exceeds the interchange fees maximizing total user surplus and welfare.*

The bias in interchange fees induced by merchant internalization can be large. For example, consider the symmetric setting in which quasi-demand functions are linear (i.e., transaction benefits are drawn from the same uniform distribution over $\left[\underline{b}, \overline{b}\right]$ for buyers and sellers), issuing and acquiring costs are equal (i.e., both equal to $c/2$), and profit margins $\widetilde{\pi}$ are constant and arbitrarily close to zero for both issuers and acquirers. The relevant range of interchange fees is defined by $\left[c/2 - \overline{b} + \widetilde{\pi}, 3\left(\overline{b} - c/2\right) - 4\widetilde{\pi}\right]$, with the requirement that $4\overline{b} > 2c + 5\widetilde{\pi} > \underline{b} + 3\overline{b}$ for an interior solution. Then instead of $a^W = a^\Pi = 0$ in the limit as margins $\widetilde{\pi}$ go to zero, which would be true if there was no merchant internalization, $a^W = a^\Pi/3 > 0$, so that moving to the welfare maximizing interchange fee requires privately set interchange fees be cut by two-thirds.[6] This result does not depend on particular values of the parameters $\overline{b}$ and c. To see this note that $a^\Pi = \overline{b} - c/2 - \widetilde{\pi}$ and $a^W = a^\Pi/3 + 4\left(\left(\overline{b} - c/2 + \widetilde{\pi}\right) - \sqrt{\left(\overline{b} - c/2\right)^2 + 3\widetilde{\pi}^2}\right)/3$. As $\widetilde{\pi} \to 0$, $a^W \to a^\Pi/3$ and $a^\Pi > 0$ given $\overline{b} > c/2$. The fact that the socially optimal interchange fee is positive despite both sides being completely symmetric is another implication of merchant internalization, and follows from the discussion in Section 3. Compared to the profit maximizing solution, there are 11.11% less transactions at a^W, but each transaction generates 33.33% more welfare, with overall welfare from cards increasing by 18.52%.

New benchmark result

Results from the existing literature suggest that whether interchange fees are distorted upwards is an empirical matter that depends on comparing the distributions of convenience benefits across the two sides. The discussion above already suggests that when the distribution of convenience benefits across the two sides is the same, then interchange fees will be distorted upwards. I now relax this symmetry assumption.

Without merchant internalization, a distortion in price structure may arise as the monopoly

[6] This compares to a regulated reduction in interchange fees by about one-half in Australia and in the U.S. In 2003, the Reserve Bank of Australia regulation resulted in credit card interchange fees being reduced from an average of 95 basis points to a cap of 50 basis points. The 2011 Federal Reserve ruling on the Durbin Amendment resulted in debit card interchange fees being reduced from an average of 44 cents to a cap of 21 cents plus 5 basis points.

platform does not account for the relative surpluses on the two sides. With merchant internalization, the average buyer surplus per transaction is already taken into account by a monopoly platform when it attracts marginal sellers given these sellers internalize the average buyer surplus per transaction when they decide whether to accept cards. Therefore, it is only the difference between the average and marginal seller surplus that drives a wedge between the profit maximizing and welfare maximizing interchange fee. This is the key insight I exploit. Reflecting this, welfare can be written as

$$W = D_B(p_B)\, D_S\left(\widehat{b}_S\right) v_S\left(\widehat{b}_S\right) + \Pi. \tag{8}$$

Proposition 2 establishes (8) and shows that provided pass-through rates of interchange fees are assumed to be symmetric, as was assumed by Wright (2004) and Rochet and Tirole (2011), then the bias towards excessive interchange fees holds more generally.

Proposition 2 *If the pass through of interchange fees by issuers and acquirers is symmetric at the privately determined interchange fee, then the privately set interchange fee strictly exceeds the interchange fees maximizing total user surplus and welfare.*

Lowering interchange fees from the privately set level unambiguously raises total user surplus and welfare. Note this bias does not depend on some asymmetry between issuing and acquiring competition, or on the particular distribution of convenience benefits by buyers and sellers. Instead, it requires only merchant internalization, price coherence and that the quasi-demand function of sellers is log-concave. To see why this bias arises, note that due to merchant internalization, the only difference between welfare maximization and profit maximization arises from the fact sellers' inframarginal surplus per card transaction is ignored by the platform (i.e., the difference between the average and marginal seller's surplus per card transaction). The average buyer's surplus per transaction is already taken into account by the marginal seller. Starting from the privately determined interchange fee, a slightly lower interchange fee increases buyer fees p_B and decreases seller fees p_S by the same amount due to symmetric pass-through rates, and does not change the total number of card transactions at the margin. But the change in fee structure in favor of sellers is equivalent to an improvement in the terms facing sellers,[7] which leads to an increase in sellers' inframarginal surplus per card transaction under the log-concavity of sellers' quasi-demand function.

In the special case when quasi-demand functions are linear on both sides but not necessarily equal, and issuer and acquirer margins are constant but possibly unequal, the card platform's private choice of interchange fee is $a^{\Pi} = \bar{b}_S - p_S$. As $\bar{b}_S$ is the convenience benefit of the seller with the maximum convenience benefit of accepting cards, this has the stark implication that no seller would be willing to accept cards if not for merchant internalization. Thus, an extreme form of Vicker's must-take argument holds with linear demands. If not for the ability to offer a valuable service to their customers, all sellers would reject cards at unregulated interchange fees.

The assumption that the pass through of interchange fees by issuers and acquirers is symmetric at the privately determined interchange fee does not necessarily rule out asymmetric market structures. For instance, with linear quasi-demand functions, it can easily be checked that perfectly competitive acquirers that pass through costs one-for-one with no margin, imply a monopoly issuer will pass through interchange fees one-to-one into higher rewards (or lower fees) at a^{Π}. A monopoly issuer will take into account how changing the card fee affects the proportion of sellers accepting cards due to merchant internalization, and this makes the issuer's card fee more responsive to the interchange fee than otherwise would be the case.[8] Thus, despite the asymmetric market structure of issuers and acquirers in this simple example, pass-through rates are symmetric at a^{Π} and Proposition 2 continues to apply.

[7] A lower interchange fee improves the average quality of buyers coming to sellers as buyers, on average, obtain higher convenience benefits from using cards reflecting the higher fees they face, or reduced rewards.

[8] Without merchant internalization, the pass through would be only 1/2 as is usual for a monopolist facing linear demand.

Partial merchant internalization and asymmetric pass-through rates

Thus far, the analysis has assumed full merchant internalization and equal pass-through rates for issuers and acquirers, at least at the privately determined interchange fee. This section relaxes both assumptions.

The possibility of partial merchant internalization is considered in Rochet and Tirole (2002, 2011), in which sellers may not take into account all of their customers' surplus from card transactions when deciding whether to accept cards, possibly because some fraction of buyers are not aware of whether the seller accepts cards when deciding whether to go to the seller. Following their notation, the convenience benefit of the marginal seller thus becomes

$$\widehat{b}_S = p_S - \alpha v_B\left(p_B\right), \tag{9}$$

with $0 < \alpha \leq 1$ being the extent of merchant internalization.

The possibility of asymmetric pass-through rates has been allowed for in many of the theoretical studies of interchange fees. In the seminal work of Rochet and Tirole (2002), they assumed a 100% pass through on the acquiring side but cost absorption on the issuing side. Katz (2005, p.132) writes "*Industry wisdom is ... that acquirers generally pass through a higher percentage of fee changes to their customers than do issuers.*" There is indeed some limited empirical evidence that is suggestive of an asymmetric pass through. In an empirical study of issuing and acquiring in Europe, the European Commission (2006) found a pass-through rate of 0.4 for acquiring versus -0.25 for issuing (based on their preferred fixed-effects specification). Studies of the regulatory experiment in Australia in 2003 in which interchange fees were reduced by 40 basis points in 2003 suggest a roughly 100% pass through into lower merchant fees, but that cardholder fees (net of rewards and interest free benefits) have increased by a more modest amount (Chang *et. al.,* 2005).

If this asymmetry in pass-through rates exists, then a card platform can increase the total profit margin of issuers and acquirers by setting higher interchange fees thereby shifting revenue to the issuing side where it is "competed away" less (Wright, 2004). At the same time, if issuers do not pass through interchange fees as much as acquirers, then any change in interchange fees will have a possibly larger effect on sellers' willingness to accept cards as the total fees across the two sides will now be affected by interchange fees. Allowing for partial merchant internalization and asymmetric pass-through rates, the following proposition establishes sufficient conditions for the bias established in Proposition 2 to remain.

Proposition 3 *With full merchant internalization, a sufficient condition for the privately set interchange fee to strictly exceed the interchange fees maximizing total user surplus and welfare is that the pass-through rate on the issuing side is no higher (in magnitude) than on the acquiring side at the privately determined interchange fee. With partial merchant internalization, a sufficient condition for the privately set interchange fee to strictly exceed the interchange fees maximizing total user surplus and welfare is that the pass-through rate on the issuing side must be sufficiently small (in magnitude) relative to that on the acquiring side at the privately determined interchange fee.*

Proposition 3 generalizes the bias found in Proposition 2 by showing it continues to hold even if pass-through rates are asymmetric in that issuers pass through interchange fees less than do acquirers at the profit maximizing interchange fee, indeed it is stronger in this case, and that partial merchant internalization can be accommodated in this case. For instance, with linear quasi-demands, the bias arises for any degree of merchant internalization $\alpha > 0$ provided the pass-through rate in acquiring p_S' is not too small relative to the pass-through rate in issuing $|p_B'|$, namely $p_S' > (1 - \alpha/2)\,|p_B'|$ at a^{Π}. If the presumption that the pass-through rate is less on the issuing side holds, this requirement always holds.

The logic behind Proposition 3 is similar to that in Proposition 2. The effects that drive a bias against sellers in Proposition 2 are stronger when issuers pass through interchange fees less than do acquirers. This reflects that the improvement in sellers' terms as a result of a lower interchange fee is now stronger because seller fees will decrease by more than buyer fees increase, and that there is now an additional effect, which is the increase in transactions resulting from the total fee charged to the two sides being lower, which also increases welfare. Given an asymmetry in pass-through rates strengthens the previous result that the privately set interchange fee is set inefficiently high, it is therefore not surprising that general results (i.e., irrespective of the relative shapes of the quasi-demand functions) can still be obtained with partial merchant internalization in such cases.

Whereas it is clear from the proof of Proposition 3 that regulating lower interchange fees will result in more card transactions at the margin (given the interchange fee is set beyond the volume maximizing level in order to shift revenues to the side where they are competed away less), it is less clear whether reducing the interchange fee all the way to the welfare maximizing interchange fee will end up increasing or decreasing card transactions. The next proposition compares welfare and output maximizing interchange fees to address this question allowing for the possibility pass-through rates are asymmetric and merchant internalization is partial.

Proposition 4 *Provided the pass-through rate on the acquiring side is not too much higher (in magnitude) than on the issuing side at the output maximizing interchange fee, welfare maximization requires an interchange fee lower than the interchange fee maximizing output (i.e., the total number of card transactions). In other words, maximizing the total number of card transactions will still result in an excessive interchange fee.*

The proposition shows that unless the pass-through rate of interchange fees is particularly low on the issuing side, the welfare maximizing solution requires an interchange fee below that which maximizes the number of card transactions. For instance, with linear quasi-demands, the condition in the proposition holds for any degree of merchant internalization $\alpha > 0$ provided the pass-through rate in acquiring p_S' is not too large relative to the pass-through rate in issuing $|p_B'|$, namely $p_S' < (1 + \alpha/2)\,|p_B'|$ at a^T.

Initially, lowering interchange fees from the unregulated level will increase the number of card transactions, reflecting that with asymmetric pass-through rates the card platform sacrifices some transactions in order to shift revenues to the side where they are competed away less.

However, under the condition of the proposition, eventually card transactions will decrease before the welfare maximizing interchange fee is reached. Taken together, the conditions (12) and (13) from the proofs of the two propositions imply $a^W < a^T \leq a^\Pi$. Lowering interchange fees to the welfare maximizing level may imply more or less card transactions compared to the unregulated private solution. This provides a possible explanation for why despite interchange fees being cut in half following regulations in Australia in 2003, there was no discernable impact on the overall level of card transactions (see Hayes, 2010). The implication would be that the volume of card transactions remained roughly unchanged in Australia despite a drop in the average use of cards at any given retailer, as more retailers accepted cards following the change, with the two effects roughly offsetting.[9]

Another reason Proposition 4 is of interest is that card platforms' have stated their objective is actually to maximize the total number of card transactions and not member profit (see p.35 of Prager *et. al.*, 2009). Maximizing the total number of card transactions is the same as maximizing profit if pass-through rates are equal, in which case Proposition 2 still applies. However, if the pass-through rate is higher for acquiring, then Proposition 2 can no longer be used to show the privately set interchange fee (that maximizes card transactions) will be set too high. Instead, Proposition 4 is required. It shows that provided the condition (13) from the proof of Proposition 4 holds, the privately set interchange fee a^T is still too high.

Perfect price discrimination

So far, the analysis has assumed that the card platform cannot price discriminate on either side. In this section, I consider instead the opposite extreme, in which there is perfect price discrimination possibilities on both sides. To proceed, I suppose there is a single acquirer that can observe b_S for each subsector and set a different p_S accordingly, and a single issuer that can set a two-part tariff (i.e., a fixed membership fee and a per transaction usage fee) to the ex-ante

[9] Another possible explanation is that, as proposed by Gans (2006), retailing in Australia was sufficiently competitive that interchange fees were effectively neutral.

identical buyers who have to decide first whether to hold the payment card or not. This setting captures perfect price discrimination. The assumed timing of the players' moves is as follows.

- In stage 1, the monopoly platform sets its interchange fee to maximize the joint profit of the issuer and acquirer.
- In stage 2, the issuer and acquirer observe this and set their fees.
- In stage 3, buyers decide whether to hold cards and sellers decide whether to accept cards.
- In stage 4, buyers decide which seller to go to in each subsector. They then draw their convenience benefit of using cards and decide whether to use the card or cash for the purchase (or not to purchase at all).

Assume buyers get no surplus from holding the payment card other than the surplus they expect to get from possible usage. In addition to merchant internalization, assume that in each subsector there is a single equilibrium retail price regardless of how consumers pay, and that at this retail price buyers retain their surplus from using cards. These assumptions hold with perfectly competitive sellers or Hotelling-Lerner-Salop sellers (see Rochet and Tirole, 2011), amongst other settings. Finally, as before, the relevant objective function for the discriminating platform is assumed to be log-concave (see Appendix B for sufficient conditions). With these assumptions, the interchange fee that is set by the monopoly platform is still excessive.

Proposition 5 *A platform that can perfectly price discriminate will set an interchange fee strictly exceeding the interchange fee maximizing total user surplus and welfare.*

The bias given in Proposition 5 reflects that buyers' surplus from using cards gets counted twice in the determination of the monopoly interchange fee — once from the platform extracting buyers' surplus from using cards and once from the platform extracting sellers' surplus from accepting cards, given individual sellers already internalize the buyers' surplus from using cards when deciding how much to pay to accept cards. As a result, the price structure remains biased in favor of buyers (and against sellers). Buyers' interests are over-weighted. With linear quasi-demands, the interchange fee determined in Proposition 5 is given by

$$a^{PD} = a^W + \frac{2\left(\bar{b}_B + \bar{b}_S - c\right)}{3}\left(\sqrt{3} - 1\right),$$

which is even higher than the interchange fee $a^\Pi = a^W + \left(\bar{b}_B + \bar{b}_S - c\right)/3$ that would be set by a monopoly platform in the absence of price discrimination on either side. Price discrimination results in a greater bias towards buyers in this special case.

The double counting of buyers' surplus from using cards only arises if buyers indeed expect to obtain a surplus from being able to use cards. In the case of a monopoly seller facing buyers with a common unit demand in each subsector (see Appendix A), the seller fully extracts the buyers' surplus through a high price and the buyer does not expect any surplus from holding cards (indeed, the buyer does not expect to get any positive surplus from purchasing goods). In this particular setting, no bias in fee structure arises under a platform that can perfectly price discriminate. Similarly, if there are different prices for cash buyers and card buyers in each subsector (as in the setting with a single competitive card accepting seller in Appendix A) such that buyers who do not hold the card can always buy from cash sellers for a correspondingly lower price, then again buyers will not be willing to pay anything to the issuer to hold the payment card. As a result, their surplus will only be counted once, on the seller's side, and the platform will set the efficient interchange fee.[10] More generally, buyers can expect to retain some of their surplus from using cards, and to the extent they do, their surplus will still be over-weighted by the platform.

A similar logic to that underlying Proposition 5 can be expected to hold even if price discrimination is not perfect (either because issuing and acquiring is not monopolized, or because the heterogeneity in user benefits cannot be perfectly captured), provided the platform puts some

[10]In contrast, the bias towards excessive interchange fees remains in these cases in the absence of price discrimination as Proposition 2 does not require that buyers expect to obtain a positive surplus from being able to use cards.

weight on the inframarginal surplus of both sides and buyers retain some surplus from being able to use cards. Guthrie and Wright (2007) establish a somewhat similar result for perfectly competing platforms. If sellers accept both cards and buyers choose one of the competing cards to hold, they show a competitive bottleneck arises, in which each platform sets interchange fees at the same level as would be set by a single (i.e., monopoly) card platform.[11] This provides one justification for my focus on the interchange fees set by a single card platform.

Allowing sellers to surcharge

A fundamental assumption in most of the existing literature is that sellers cannot pass on the costs of accepting cards to their customers (e.g., through a surcharge for using cards). One reason for this assumption is that such surcharging is banned by card platforms in many countries. If, on the other hand, surcharging is allowed, existing theories imply interchange fees will be completely passed through to buyers and so the level at which they are set becomes irrelevant from a policy perspective, as well as from the perspective of buyers, sellers and card platforms (see Gans and King, 2003 for a very general neutrality result, and also Rochet and Tirole (2002), Schwartz and Vincent (2006) and Wright (2003) for comparisons of welfare with and without surcharging). However, even in countries where card platforms have been required to allow sellers to surcharge (e.g., Australia, New Zealand, and a number of European countries), many sellers continue not to add any surcharge for buyers paying by card.[12] Moreover, in most countries sellers are free to discount for cash but do not often do so. This suggests that surcharging is not perfect, in the sense some (or perhaps many) sellers are reluctant to surcharge for card payments even when it is allowed. Possible reasons include (i) the costs of setting differential prices; (ii) that surcharges

[11] Rysman (2007) provides empirical evidence that consumers concentrate their spending on a single card network.

[12] In Australia, where surcharging seems to be far more prevalent than other countries, surcharging rates have been reported in one study to be around 30 per cent of all retailers accepting cards (as at December 2010) and increasing. See “Review of Card Surcharging: A Consultation Document” June 2011, Reserve Bank of Australia. In other countries, surcharging rates are typically found to be less than 10 per cent.

for using cards may be a lot more salient to buyers than small differences in the sellers' cash prices; or (iii) that in certain contexts, buyers may view surcharges as unfair and so reputational considerations stop sellers from adding surcharges.

Suppose, when allowed, some fraction of sellers surcharge. If this fraction is representative of all sellers and does not depend on the interchange fee, then the previous analysis will not change. For instance, suppose some fraction of sellers do not face any cost to surcharging whereas the remaining sellers face relatively high costs and are always worse off surcharging. Neutrality (Gans and King, 2003) implies the number of transactions for which cards are used with surcharges in each subsector will be independent of a, and the profit maximizing interchange fee will still maximize profit $(p_B + p_S - c)\,T$ as it makes no difference that this is now multiplied by the exogenous fraction of sellers that do not surcharge. The interchange fee that maximizes welfare will also be the same as the one arising if no sellers surcharge as the contribution to welfare from industries where there is surcharging does not vary with the interchange fee. The only difference to the analysis is that the welfare effects of changing interchange fees will be reduced because the welfare effects only apply to the group of sellers that do not surcharge. Thus, the previous results continue to apply.

Things can obviously change if the fraction of sellers surcharging increases with the interchange fee, or if the distribution of sellers that surcharge is not representative of all sellers. In an earlier working paper version of this article, I worked out the case in which sellers that set a surcharge for card transactions incur a positive cost of doing so. The unobserved heterogeneity in sellers is assumed to be in terms of this cost. When issuing and acquiring margins are arbitrarily small, the efficient interchange fee leads to no surcharging as sellers are indifferent between accepting cash and cards at this interchange fee. Relative to this benchmark, a small increase in interchange fees has no first-order effect on the degree of surcharging but increases card transactions and platform profits by making buyers use cards more frequently. In deciding how much higher to set interchange fees, the profit-maximizing card platform trades off higher card transactions at sellers that set uniform prices with a shift towards more sellers choosing to surcharge. The result is that the card platform always set an interchange fee that is too high, inducing some inefficient surcharging.

The emerging empirical evidence on surcharging paints a more complicated picture. The evidence from Australia suggests large sellers are more likely to surcharge than small sellers, but that there is some surcharging in all size categories. Furthermore, evidence suggests some sellers may be adding surcharges in order to extract an additional fee from buyers who may not be aware of the surcharge until the point of purchase. This has led to a government inquiry in Australia, where the *average* surcharge level amongst sellers surcharging is reported to be around twice that of the merchant fee faced by these sellers.[13] Similar issues have arisen in the U.K. where surcharges are allowed, with excessive surcharges noted in certain sectors such as air travel and booking agencies. This recently led the U.K government to announce the banning of excessive surcharges on card payments by the end of 2012.[14] An important direction for future research is to develop a theory which is capable of explaining why some sellers do not add any surcharge for cards or discount for cash while at the same time other sellers seem to surcharge card payments excessively. A more complete understanding of whether unregulated interchange fee are too high or not allowing for surcharging will therefore need to await the development of such a theory.

5 Policy implications

Having established that with uniform retail pricing there is an underlying bias in the structure of fees for payment cards, with retailers charged too much and cardholders too little, in this final section I address the case for government intervention.

The modelling in this article suggests excessive interchange fees can be fully explained (along

[13]See http://www.rba.gov.au/publications/consultations/201106-review-card-surcharging/pdf/201106-review-card-surcharging.pdf for recent empirical evidence on surcharging in Australia.

[14]See http://www.hm-treasury.gov.uk/press_148_11.htm

with the other concerns of policymakers) without relying on any anticompetitive behavior on the part of the card networks or their members. This is important as numerous antitrust cases have been filed against the card networks MasterCard and Visa (most of them settled out of court), including that interchange fees represent a horizontal price fix as they put a floor under the price that competing acquirers charge to retailers. The findings in this article suggest a regulatory approach rather than an antitrust approach is likely to be more appropriate.

Should governments then intervene and set lower interchange fees? There are several reasons to exercise caution in this regard. The bias I emphasized depends on several key properties — most importantly, merchant internalization. Farrell (2006) has argued that merchant internalization is likely to hold quite generally. Without merchant internalization, it is difficult to make sense of the must-take cards logic of Vickers (2005). Why would retailers want to accept American Express cards with higher merchant fees when they already accept MasterCard and Visa cards if not to enable them to attract additional sales, or set higher retail prices, taking as given the behavior of their rivals? Nevertheless, the extent to which merchant internalization holds is an open empirical question.

Another concern arises from determining what is the right level to regulate interchange fees at. There is near unanimity amongst economists that setting interchange fees based on the costs faced by issuers, as has happened in Australia (for credit cards) and in the U.S. (for debit cards) has no good theoretical basis. Instead, extending the insights of Baxter (1983), recently Rochet and Tirole (2011) and Rochet and Wright (2010) have proposed standards for setting interchange fees that are based on retailers' avoided cost from not having to accept cash, or in the case of credit cards, not having to provide credit themselves. The idea is to get cardholders to internalize the cost savings (or additional costs) their payment choices impose on retailers. However, measuring avoided costs involves calculating the costs to retailers of accepting different types of payments across the many different types of retailers accepting cards. This is a major undertaking and one which authorities such as the European Commission have only just started to embark on. If done correctly, not only does this provide a benchmark interchange fee that can be used to guide regulation, it potentially also provides a direct way of testing whether unregulated interchange fees are excessive.

A third major difficulty with regulating interchange fees is that doing so only deals with the bias in fee structure arising from four-party payment networks, leaving three-party networks that avoid interchange fees to continue to set high merchant fees and high rewards to cardholders if doing so gives them a competitive advantage. In the long-run, this could shift card transactions to three-party schemes, which are arguably less efficient as they do not exploit existing relationships between banks and end-users. An alternative approach of directly imposing caps on each network's average merchant fees would avoid this problem as it would apply across the board regardless of whether card networks make use of competing acquirers or deal with retailers directly.

Finally, to the extent that retailers can steer consumers to their preferred means of payment through surcharges and discounts, the bias against retailers found in this article is likely to be less of an issue. However, even if retailers are allowed to surcharge, the analysis in this article suggests the rationale for regulating lower interchange fees may remain. Not all retailers will surcharge if there are some costs to retailers of surcharging. On the other hand, removing constraints on surcharging may also come at its own cost. Some retailers may surcharge opportunistically by using the surcharge as a hidden add-on price that is not fully taken into account by customers until after the customer has already committed to make the purchase. Governments in Australia and U.K. have been investigating exactly this situation. Regulations that allow for surcharging may need to consider putting caps on the level of surcharges. Alternatively, removing restraints on surcharging may only be necessary for card networks that would otherwise escape the regulation of interchange fees (e.g., three-party networks, in case they do not have their merchant fees regulated), in order to maintain a more level playing field across three-party and four-party payment networks.

Although the main interest has been in the bias arising in the pricing of card platforms, similar implications will apply to other types of two-sided platforms in which one side internalizes the surplus of the other when deciding whether to interact on the platform. One case that seems particularly relevant for policy at present is the issue of net neutrality. For instance, should policymakers accommodate a high speed tier on the Internet in which traffic on this tier is prioritized over the basic tier, with the platform providers charging fees to end users (i.e., consumers and websites) for the interactions that they enable on the high speed tier (or equivalently determine settlement fees between the platform providers for such upgraded interactions)? In case a website's services are paid services, then like retailers, they may be reluctant to use differential pricing to consumers to reflect the differential fees they face from the platform providers for different speeds of service. Merchant internalization would then imply websites would take into account the additional surplus their customers get in using the high speed tier when deciding whether to pay for it, but not vice-versa. The theory presented in this article then suggests that the structure of fees for the high speed Internet service would be biased against websites, taking as given the existence of the basic tier as it is now. This seems like a fruitful direction for future research.

Appendix A: Merchant internalization

Consider some additional models of seller behavior which also give rise to the assumption in (1). Suppose the cost of each good to the seller is d and denote the seller's retail price p.

Monopoly seller (unit demand and elastic demand) Suppose the seller is a monopolist. Rochet and Tirole (2011) consider (in their Proposition 1) the case the buyer has known value from the good, denoted r here. They propose the seller will set the price $p = r + v_B(p_B) D_B(p_B)$ if it joins the platform (i.e., accepts cards) given the buyer only learns b_B after he has chosen to go to the seller. That is, the buyer's expected surplus is fully extracted. However, this cannot be the case because if a buyer draws $b_B < p_B$ at the point of sale such that they prefer not to interact on the platform (i.e., they prefer to use cash), their surplus from purchasing will be less than the price they face and so they will not want to buy the good.[15] Thus, in this setting, the seller cannot charge more than r without losing some demand (those consumers with low draws of b_B). However, (1) can still be derived for a monopoly seller if buyers face a sufficiently large travel cost τ to get to the seller. Assuming $\tau > v_B(p_B) D_B(p_B)$ is sufficient. This implies $p = r + v_B(p_B) D_B(p_B) - \tau$ if the seller accepts cards and $p = r - \tau$ if it does not. Once at the seller, this travel cost is sunk, and the buyer can always get a positive surplus buying with cash. The seller will therefore accept cards if and only if the extra margin $v_B(p_B) D_B(p_B)$ exceeds the extra cost $(p_S - b_S) D_B(p_B)$ giving rise to (1).

This result can be extended to the case in which there is a unit mass of buyers with each buyer drawing a value of r from some continuously differentiable distribution function G_r, prior to deciding whether to go to the seller or not. Buyers continue to draw b_B ex-post. A buyer will go to the seller provided $r + v_B(p_B) D_B(p_B) - \tau \geq p$, so $Q = 1 - G_r(p + \tau - v_B(p_B) D_B(p_B)))$ is the measure of buyers that will purchase from the seller. The profit of the seller is $(P(Q) - (p_S - b_S - v_B(p_B)) D_B(p_B) - \tau - d) Q$, where $P(Q) = G_r^{-1}(1 - Q)$ is the standard inverse demand function for a monopolist (i.e., ignoring the expected surplus to buyers from card usage and the transport cost). Written in this form shows this is a standard monopoly problem where the term $(p_S - b_S - v_B(p_B)) D_B(p_B)$ is equivalent to an additional constant marginal cost facing the seller. Given the seller's profit is decreasing in its own constant marginal cost, the seller will accept cards if and only if this term is non-positive, giving rise to (1).

Competitive sellers Suppose there are many (at least two) identical sellers (with the same b_S) that simultaneously choose whether to accept cards or not. After their acceptance decisions are known, they compete in retail prices. Buyers have to decide which seller to frequent before

[15] Indeed, even if $b_B > p_B$ but $b_B - p_B < v_B(p_B) D_B(p_B)$, they will prefer not to buy the good.

knowing their draw of b_B. If $v_B(p_B) \geq p_S - b_S$, then in equilibrium, at least one seller will accept cards as the increased convenience benefit of buyers $v_B(p_B) D_B(p_B)$ is at least as high as the additional cost to the seller $(p_S - b_S) D_B(p_B)$. If only one seller accepts cards, then the card price will be $d + v_B(p_B) D_B(p_B)$ and the cash price is d. Consumers are indifferent between the sellers and are assumed to buy from the card seller. Although the card seller obtains a profit of $v_B(p_B) D_B(p_B)$, there is no incentive for another seller to accept cards as then they will each obtain no profit if they do so.[16] If multiple sellers accept cards, then $p = d + (p_S - b_S) D_B(p_B)$. Card sellers will attract buyers as $v_B(p_B) \geq p_S - b_S$. Any cash-only seller that sets a price of d will not attract any buyers. If $v_B(p_B) < p_S - b_S$, no seller will want to accept cards as the amount buyers are willing to pay to be able to use cards is less than the cost of accepting card transactions. With multiple sellers accepting cash only, the retail price will be d.

As a result of these two cases, (1) always holds (i.e., regardless of whether only one or multiple sellers accept cards when $v_B(p_B) \geq p_S - b_S$). The case with only one seller accepting cards is similar to the specific "Single Credit-Card Merchant case" of Gans and King (2001). The case with multiple sellers accepting cards corresponds to Rochet and Tirole's (2011) analysis of perfectly competitive retailers.

Appendix B: Proofs

Log-concavity of objective functions

I show sufficient conditions for the log-concavity of the objective functions T, Π, W and U. Suppose D_i is log concave in $\widehat{b}_i$, and p_i'' and v_i'' are arbitrarily close to 0 (for $i = B, S$). Ignoring terms that depend on p_B'', p_S'', v_B'' and v_S'' (which add a term which can be made arbitrarily close to zero),

$$\frac{d^2(\log D_i)}{da^2} = \frac{d^2(\log D_i)}{d\left(\widehat{b}_i\right)^2}\left(\frac{d\widehat{b}_i}{da}\right)^2 < 0$$

[16]With any small fixed cost of accepting cards, the only equilibrium would involve one seller accepting cards.

for $i = B, S$ implying D_B and D_S are log-concave in a, and so $T = D_B D_S$ is log-concave in a. Profit $\Pi = (p_B + p_S - c) T$ is also log-concave in a given p_B'' and p_S'' are arbitrarily close to 0. To show W is log-concave in a, recall it can be written as $W = wT$, where $w = ((1-\alpha) v_B + v_S + p_B + p_S - c)$ and $0 < \alpha \leq 1$ depending on the extent of merchant internalization. Ignoring terms that depend on p_B'', p_S'', v_B'' and v_S'' (which add a term which can be made arbitrarily close to zero),

$$\frac{d^2}{da^2} \log(w) = -\frac{1}{w^2} \left(\frac{dw}{da} \right)^2 < 0$$

so that W is log-concave in a, as is $U = ((1-\alpha) v_B + v_S) T$ by the same argument. Similarly, for the case with perfect price discrimination, the objective function is $\Pi_{PD} = (v_B + v_S + p_B + p_S - c) T$ so this is log-concave by the same argument as for W.

If convenience benefits are drawn from the generalized Pareto distribution (GPD) with increasing hazard, and issuers and acquirers have margins that are linear in interchange fees, these sufficient conditions hold. To see this, note for the case of b_i, the GPD assumption implies the quasi-demand

$$D_i\left(\widehat{b}_i\right) = \left(1 - \frac{\varepsilon_i\left(\widehat{b}_i - \mu_i\right)}{\sigma_i}\right)^{\frac{1}{\varepsilon_i}},$$

where $\varepsilon_i > 0$, $\mu_i \leq b_i \leq \mu_i + \frac{\sigma_i}{\varepsilon_i}$ which is required for the quasi-demand function to be log concave. (Note if $\varepsilon_i = 1$ we get the uniform distribution for which quasi-demand is linear.) Then

$$v_i\left(\widehat{b}_i\right) = \frac{\sigma_i - \varepsilon_i \widehat{b}_i + \varepsilon_i \mu_i}{1 + \varepsilon_i},$$

where $\mu_i \leq \widehat{b}_i \leq \mu_i + \frac{\sigma_i}{\varepsilon_i}$. This implies $v_i'' = 0$. Margins that are linear in interchange fees imply $p_B'' = p_S'' = 0$. As shown above, these conditions are sufficient for the log-concavity of the objective functions T, Π, W and U.

Proofs of propositions

Proof of Proposition 1. Symmetry implies $D \equiv D_B = D_S$, $v \equiv v_B = v_S$, $\underline{b} \equiv \underline{b}_B = \underline{b}_S$ and $\overline{b} \equiv \overline{b}_B = \overline{b}_S$. Let a^{Π} be the interchange fee maximizing Π. Given the assumptions underlying

Proposition 11 in Rochet and Tirole (2011) hold, their result implies that if $v(p_B) > v\left(\widehat{b}_S\right)$ at a^{Π}, then a^{Π} is higher than the socially optimal level. Given $v' < 0$, this requires showing $p_B < \widehat{b}_S$ at a^{Π}.

Under the assumption of constant margins, the platform's profit maximization problem is the same as maximizing $T = D(p_B) D\left(\widehat{b}_S\right)$. Note $T(a) = 0$ for $a = \underline{a}_B$ and $a = \overline{a}_S$, with $T(a) > 0$ for $\underline{a}_B < a < \overline{a}_S$. Log-concavity of T with respect to a implies a^{Π} solves

$$\frac{\partial \log T}{\partial a} = \frac{D'\left(\widehat{b}_S\right)(1+v')}{D\left(\widehat{b}_S\right)} - \frac{D'(p_B)}{D(p_B)} = 0.$$

Now consider the unique interchange fee $\widetilde{a}$ such that $p_B = \widehat{b}_S$. This exists as p_B is decreasing in a with $p_B(\underline{a}_B) = \overline{b}$ and $p_B(\overline{a}_B) = \underline{b}$, whereas $\widehat{b}_S$ is increasing in a with $\widehat{b}_S(\underline{a}_S) = \underline{b}$ and $\widehat{b}_S(\overline{a}_S) = \overline{b}$, where recall $\underline{a}_S \leq \underline{a}_B < \overline{a}_S \leq \overline{a}_B$. Given $\partial^2 \log T / \partial a^2 < 0$ and $0 < 1 + v' < 1$, then $\partial \log T / \partial a > 0$ at $\widetilde{a}$. Thus, a^{Π} involves a higher interchange fee than $\widetilde{a}$, which implies $p_B < \widehat{b}_S$ at a^{Π}. As a result, a^{Π} is higher than the socially optimal level. Q.E.D.

Proof of Proposition 2. Let a^{Π} be the interchange fee maximizing Π and a^W be the interchange fee maximizing W. Given the assumptions made, these are uniquely defined by $d\Pi/da = 0$ and $dW/da = 0$ respectively. From (2), (3) and $\widehat{b}_B = p_B$, welfare can be written as

$$\begin{aligned} W &= D_B(p_B) D_S\left(\widehat{b}_S\right)\left(v_B(p_B) + E\left(b_S | b_S \geq \widehat{b}_S\right) - p_S\right) + \Pi \\ &= v_S\left(\widehat{b}_S\right) T + \Pi, \end{aligned}$$

given $v_S\left(\widehat{b}_S\right) = E\left(b_S | b_S \geq \widehat{b}_S\right) - \widehat{b}_S = E\left(b_S | b_S \geq \widehat{b}_S\right) - p_S + v_B(p_B)$ from (1). This explains (8) in the text. Differentiating W with respect to a and evaluating at a^{Π} implies that

$$\frac{dW}{da} = v_S'\left(\widehat{b}_S\right)\left(-1 - v_B'(p_B)\right) p_B' T. \tag{10}$$

To see this, note that $d\Pi/da = 0$ implies $dT/da = 0$ given symmetric pass-through rates. Note the expression in (10) comes from differentiating $v_S\left(\widehat{b}_S\right)$ with respect to a, taking into account $\widehat{b}_S = p_S - v_B(p_B)$ and $p_S' = -\ p_B'$. By construction $v_i'(.) > -1$, so the second term is negative. Lemma A1 of Hermalin and Katz (2004) or Theorem 3 of Bagnoli and Bergstrom (2005) establish that the log-concavity of $D_i(.)$ implies log-concavity of the right-hand integral of $D_i(.)$ (i.e., the surplus function). Lemma 2 of Bagnoli and Bergstrom (2005) establishes that the log-concavity of the right hand integral of $D_i(.)$ implies $v_i'(.) < 0$ so that $v_S'\left(\widehat{b}_S\right) < 0$ in particular. Given $p_B' < 0$, the expression in (10) is therefore negative. From the log-concavity of W, this implies $a^W < a^\Pi$. The same result applies with respect to total user surplus, which can be written as $U = W - \Pi$. The result follows given $dU/da = dW/da$ at a^Π. Q.E.D.

Proof of Proof of Proposition 3. Due to (9), $v_S\left(\widehat{b}_S\right) = \alpha v_B(p_B) + E\left(b_S | b_S \geq \widehat{b}_S\right) - p_S$, so welfare can be written as

$$W = \left((1-\alpha) v_B(p_B) + v_S\left(\widehat{b}_S\right)\right) T + \Pi. \tag{11}$$

Starting from a^Π, the equivalent of (10) is now

$$\begin{aligned}\frac{dW}{da} &= \left((1-\alpha) v_B(p_B) + v_S\left(\widehat{b}_S\right)\right) \frac{dT}{da} \\ &\quad + \left((1-\alpha) v_B'(p_B) + v_S'\left(\widehat{b}_S\right)\left(\frac{p_S'}{p_B'} - \alpha v_B'(p_B)\right)\right) p_B' T.\end{aligned}$$

The results established below will also apply to total user surplus $U = W - \Pi$ given $dU/da = dW/da$ at a^Π. To sign the first term, note the first order condition from profit maximization implies

$$\left(p_B' + p_S'\right) T + \left(p_B + p_S - c\right) \frac{dT}{da} = 0$$

or

$$\frac{dT}{da} = -\left(\frac{p_B' + p_S'}{p_B + p_S - c}\right) T.$$

This will also be non-positive provided $p_S' \geq |p_B'|$ at a^Π. Now consider the second term evaluated at a^Π. Provided

$$\frac{p_S'}{|p_B'|} > -\alpha v_B'(p_B) + (1-\alpha) \frac{v_B'(p_B)}{v_S'\left(\widehat{b}_S\right)}, \tag{12}$$

then $dW/da < 0$ at a^Π so that $a^W < a^\Pi$. Thus, it remains to show when the inequality in (12) holds. Given the log concavity of $D_B(.)$, $v_B'(p_B) < 0$ and so the right hand side of (12) is positive. If $\alpha = 1$, then the inequality in (12) just requires $p_S' > -v_B'(p_B)|p_B'|$. Since

$v_B'(p_B) > -1$, a sufficient condition for this to be true is $p_S' \geq |p_B'|$ at a^{Π}. If $0 < \alpha < 1$ and $|v_B'(p_B)| \leq \left|v_S'\left(\widehat{b}_S\right)\right|$, the right hand side of (12) is still less than 1, so $p_S' \geq |p_B'|$ is again sufficient for the inequality in (12) to hold. However, if $0 < \alpha < 1$ and $|v_B'(p_B)| > \left|v_S'\left(\widehat{b}_S\right)\right|$ then depending on how low is α, the inequality in (12) may require p_S' exceeds $|p_B'|$ by some critical amount. Q.E.D.

Proof of Proposition 4. Let a^T be the interchange fee maximizing the number of card transactions T. Evaluating dW/da at a^T instead of a^{Π} implies

$$\frac{dW}{da} = \left(\left(p_B' + p_S'\right) + (1-\alpha)\, v_B'(p_B)\, p_B' + v_S'\left(\widehat{b}_S\right)\left(p_S' - \alpha v_B'(p_B)\, p_B'\right)\right) T.$$

Welfare can be improved initially by lowering the interchange fee below the output maximizing level if at a^T

$$\frac{p_S'}{|p_B'|} < \frac{1 - \alpha v_B'(p_B)\, v_S'\left(\widehat{b}_S\right) + (1-\alpha)\, v_B'(p_B)}{1 + v_S'\left(\widehat{b}_S\right)} \tag{13}$$

and by raising the interchange fee if the inequality in (13) is reversed. Note since $v_i'(.) > -1$ for $i = (B, S)$ and $v_S'\left(\widehat{b}_S\right) < 0$ from the log-concavity of $D_S(.)$, the right hand side of (13) is positive. Moreover, if $\alpha = 1$, it necessarily exceeds unity. Q.E.D.

Proof of Proposition 5. For a given interchange fee a, the monopoly acquirer will want to set the highest p_S for each subsector b_S such that the seller(s) in the subsector still accept cards provided this enables the acquirer to cover its cost. Denote issuing costs to buyers c_B and acquiring costs to sellers c_S, so $c = c_B + c_S$. From (1), this implies $p_S = b_S + v_B(p_B)$ provided $b_S + v_B(p_B) \geq c_S + a$. Then the marginal seller that the acquirer attracts is defined by $\widehat{b}_S = c_S + a - v_B(p_B)$. The monopoly issuer's optimal two-part tariff is to set $p_B = c_B - a$ and a fixed fee equal to the resulting surplus each buyer expects to get from using cards $v_B(p_B)\, T$ given each buyer faces the same retail price in each subsector regardless of how they pay. With this two-part tariff, all buyers will hold a card and use their card efficiently from the issuer's perspective, so the issuer extracts maximal possible surplus from buyers. The joint profit of the issuer and the acquirer will be

$$\Pi_{PD} = v_B(p_B)T + \int_{\widehat{b}_S}^{\bar{b}_S} (p_B + b_S + v_B(p_B) - c)\, D_B(p_B)\, dG_S(b_S).$$

The card platform will therefore set its interchange fee to maximize this joint profit, which can be rewritten as $v_B(p_B)T + W$. Let the resulting interchange fee be denoted a^{PD}. Since $\widehat{b}_S = c_S + a - v_B(p_B)$, this is equivalent to a problem in which a platform sets a single interchange fee to maximize $v_B(p_B)T + W$ but in which it faces perfectly competitive acquirers that result in $p_S = c_S + a$ since in this case (1) also implies $\widehat{b}_S = c_S + a - v_B(p_B)$. Since the pass-through rates for this problem are symmetric (recall that $p_B = c_B - a$), Proposition 2 implies $dT/da > 0$ at a^W. Log-concavity of $D_B(.)$ implies $v_B'(p_B) < 0$. Therefore, $d\Pi_{PD}/da = v_B(p_B)\, dT/da - v_B'(p_B)T > 0$ at a^W, so that the log-concavity of Π_{PD} implies $a^{PD} > a^W$. With equal pass-through rates, Proposition 2 implies $a^W < a^T = a^{\Pi}$. So at a^W, $d\Pi/da > 0$. Since $U = W - \Pi$, this implies $dU/da = -d\Pi/da < 0$ at a^W. Since the interchange fee maximizing total user surplus is less than a^W it must also be less than a^{PD}. Q.E.D.

References

Bagnoli, M. and Bergstrom, T. "Log-Concave Probability and its Applications," *Economic Theory*, Vol. 26 (2005), pp. 445-469.

Baxter, W. "Bank Interchange of Transactional Paper: Legal Perspectives," *Journal of Law and Economics*, Vol. 26 (1983), pp. 541-588.

Bedre-Defolie, Ö. and Calvano, E. "Pricing Payment Cards" ESMT Working Paper 10-005, 2010.

Bolt, W. and Chakravorti, S. "Economics of Payment Cards: A Status Report," *Economic Perspectives*, Vol. 32 (2008), pp. 15-27.

Bradford, T. and Hayashi, F. "Developments in Interchange Fees in the United States and

Abroad." Working paper, Federal Reserve Bank of Kansas City, 2008.

Cabral, L. "Market Power and Efficiency in Card Payment Systems: A Comment," *Review of Network Economics*, Vol. 5 (2005), pp. 15-25.

Carlton, D. and Frankel, A. "The Antitrust Economics of Credit Card Networks," *Antitrust Law Journal*, Vol. 63 (1995), pp. 903-915.

Chang, H., Evans, D., and Garcia Swartz, D. "The Effect of Regulatory Intervention in Two-Sided Markets: An Assessment of Interchange-Fee Capping in Australia," *Review of Network Economics*, Vol. 4 (2005), pp. 328-358.

Farrell, J. "Efficiency and Competition between Payment Instruments," *Review of Network Economics*, Vol. 5 (2006), pp. 26-44.

Frankel, A. "Monopoly and Competition in the Supply and Exchange of Money," *Antitrust Law Journal*, Vol. 66 (1998), pp. 331-361.

Gans, J. "Review of the Reserve Bank & Payments System Board Annual Reports 2005" submission to the House Standing Committee on Economics, Finance and Public Administration, Parliament of Australia House of Representatives, 2006.

Gans, J. and King, S. "The Neutrality of Interchange Fees in Payment Systems," Working Paper, University of Melbourne, 2001.

Gans, J. and King, S. "The Neutrality of Interchange Fees in Payment Systems," *Topics in Economic Analysis & Policy*, Vol. 3 (2003), Article 1.

Guthrie, G. and Wright, J. "Competing Payment Schemes," *Journal of Industrial Economics*, Vol. 55 (2007), pp. 37-67.

Hayes, R. "Is Price Regulation of Payment Card Associations Effective? Evidence from a Dramatic Policy Experiment," mimeo, Melbourne Business School, 2010.

Hermalin, B. and Katz, M. "Senders or Receiver: Who Should Pay to Exchange an Electronic Message?" *RAND Journal of Economics*, Vol. 35 (2004), pp. 423-448.

Katz, M. "Reform of Credit Card Schemes in Australia II: Commissioned Report," Reserve Bank of Australia. Public Document. August 2001.

Katz, M. "What Do We Know About Interchange Fee and What Does It Mean For Public Policy?" Proceedings of Payments System Research Conference, Federal Reserve Bank of Kansas City, May 2005.

Laffont, J-J., Marcus, S., Rey, P., and Tirole, J. "Internet Interconnection and the Off-Net-Cost Pricing Principle," *RAND Journal of Economics*, Vol. 34 (2003), pp. 370-390.

Prager, R., Manuszak, M., Kiser, E., and Borzekowski, R. "Interchange Fees and Payment Card Networks: Economics, Industry Developments, and Policy Issues," Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System, 2009-03, 2009.

Rochet, J.-C. and Tirole, J. "Cooperation Among Competitors: Some Economics of Payment Card Associations," *RAND Journal of Economics*, Vol. 33 (2002), pp. 549-570.

Rochet, J.-C and Tirole, J. "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, Vol. 1 (2003), pp. 990-1029.

Rochet, J.-C and Tirole, J. "Two-Sided Markets: A Progress Report," *RAND Journal of Economics*, Vol. 37 (2006), pp. 645-667.

Rochet, J.-C. and Tirole, J. "Must-Take Cards: Merchant Discounts and Avoided Costs," *Journal of the European Economic Association*, Vol. 9 (2011), pp. 462-495.

Rochet, J.-C. and Wright, J. "Credit Card Interchange Fees," *Journal of Banking and Finance*, Vol. 34 (2010), pp. 1788-1797.

Rysman, M. "An Empirical Analysis of Payment Card Usages," *Journal of Industrial Economics*, Vol. 55 (2007), pp. 1-36.

Rysman, M. “The Economics of Two-Sided Markets,” *Journal of Economic Perspectives*, Vol. 23 (2009), pp. 125-144.

Schmalensee, R. “Payment Systems and Interchange Fees,” *Journal of Industrial Economics*, Vol. 50 (2002), pp. 103-122.

Schwartz, M. and Vincent, D. “The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network,” *Review of Network Economics*, Vol. 5 (2006), pp. 72-102.

Verdier, M. “Interchange Fees in Payment Card Systems: A Survey of the Literature,” *Journal of Economic Surveys*, Vol. 35 (2010), pp. 273-297.

Vickers, J. “Public Policy and the Invisible Price: Competition Law, Regulation and the Interchange Fee,” *Competition Law Journal*, Vol. 4 (2005), pp. 5-16.

Wang, Z. “Market Structure and Payment Card Pricing: What Drives the Interchange?” *International Journal of Industrial Organization*, Vol. 28 (2010), pp. 86-98.

Weyl, G. “A Price Theory of Multi-Sided Platforms,” *American Economic Review*, Vol. 100 (2010), pp. 1642-1672.

Wright, J. “Optimal Card Payment Systems,” *European Economic Review*, Vol. 47 (2003), pp. 587-612.

Wright, J. “Determinants of Optimal Interchange Fees,” *Journal of Industrial Economics*, Vol. 52 (2004), pp. 1-26.

Wright, J. “Why do Merchants Accept Payment Cards?” *Review of Network Economics*, Vol. 9 (2010), Article 1.