Competition Policy and Regulation in Credit Card Markets: Insights from Single-sided Market Analysis

BY DENNIS W. CARLTON & RALPH A. WINTER

Competition Policy and Regulation in Credit Card Markets: Insights from Single-sided Market Analysis

BY DENNIS W. CARLTON & RALPH A. WINTER[1]

This paper reexamines the economics of two common features of credit card networks: the interchange fee paid by merchant banks, or acquirers, to cardholder banks, or issuers; and the restraint commonly placed on merchants against surcharging for credit card transactions. We show that the parallels with the economics of conventional one-sided markets offer insights that have been overlooked in the credit card economics literature, which stresses the two-sided nature of the market. The characterization of the optimal interchange fee is equivalent to the Dorfman-Steiner theorem from conventional price theory. The principle that the interchange fee maximizes output when an optimum exists and the possibility of interchange fee neutrality also have precise parallels in one-sided markets with promotion. Our analysis shows that the no-surcharge rule is equivalent to a retail MFN constraint. The no-surcharge rule raises prices to merchants due to a competition-suppression effect as well as a cost-externalization effect. The market condition underlying interchange neutrality (when surcharging is allowed) eliminates the impact of the no-surcharge rule in the case of a credit-card duopoly. Yet the same condition magnifies the impact in the presence of cash customers.

I.   INTRODUCTION

Regulation and competition law impose a wide range of restrictions on credit card markets around the world. Those restrictions deal primarily with interchange fees and the ability of merchants to surcharge buyers who purchase with credit cards. Regulatory ceilings are imposed on interchange fees in some jurisdictions but not in others. Regulation in some jurisdictions not only allows surcharging but prohibits credit card companies from imposing "no–surcharge rules" on merchants. Yet in other jurisdictions regulation intervenes with exactly the opposite policy, directly prohibiting surcharges.[2]

Antitrust scholars have expressed an equally wide range of views on interchange fees and regulation. On interchange fees, Frankel & Shampine[3] among others argue that positive interchange fees are unnecessary and anticompetitive. Rochet & Wright[4] and Wright[5] offer models of credit card networks in which the interchange fee, while not inherently anticompetitive, always exceeds the level that maximizes consumer surplus.

Most of the literature on the economics of credit cards, on the other hand, argues that the interchange fee is set at the level that maximizes the volume of a credit card network by balancing the impact of price changes on both sides of the credit card market (e.g., Klein, Lerner, Murphy, & Plache;[6] Emch & Thompson (2006);[7] and Evans & Schmalensee).[8] Klein et al., as well as Emch & Thompson, derive expressions for the profit-maximizing interchange fee and that fee turns out to maximize output, all else equal.[9] As a price that is set to maximize output, it would seem odd to describe the interchange fee as anticompetitive.

A third strand of credit card literature develops conditions under which the interchange fee is completely irrelevant to equilibrium in a credit card network (Carlton & Frankel and Gans & King[10]). This literature shows that in the absence of a no-surcharge rule, interchange fees have no real effects in a world without transaction costs, apart from the transaction fees set within the credit card network. The interchange fee is irrelevant in the sense that the equilibrium payoff to any party in the network is unaffected by a change in the fee.

It is hard to imagine a wider range of views on the role of prices in any market. Depending on which scholars policymakers listen to, the interchange fee is inherently anticompetitive, output maximizing, or completely irrelevant.

DEPENDING ON WHICH SCHOLARS POLICYMAKERS LISTEN TO, THE INTERCHANGE FEE IS INHERENTLY ANTICOMPETITIVE, OUTPUT MAXIMIZING, OR COMPLETELY IRRELEVANT.

On the no-surcharge rule, again, the literature contains at least three views. Some argue that the rule has a role in preventing excessive merchant surcharging (Wright).[11] Others, notably Boik & Corts,[12] conclude that the restraint suppresses competition in a way that is parallel to a retail most-favored nation ("MFN") restraint, in which retailers are constrained against charging more for a manufacturer's product than for rivals' products.[13] And one of the most prominent contributions analyzing the no-surcharge rule concludes that the welfare impact of the no-surcharge rule is ambiguous (Rochet & Tirole).[14]

With the wide range of views on interchange fees and surcharges, academic scholars nonetheless agree on one proposition: The two-sided nature of credit card markets is fundamental to any analysis of regulation of either the interchange fee or restrictions on the freedom to contract for no-surcharge rules. A credit card network must attract both cardholders and merchants to survive in the market. Neither side will join the network without the other.[15]

This paper offers a different perspective. Credit card networks are undoubtedly two-sided, but economic analysis of such markets should keep a tighter link to what we know about conventional, one-sided markets than scholars have done to this point. Failure to do so will likely lead to faulty policy analysis.

CREDIT CARD NETWORKS ARE UNDOUBTEDLY TWO-SIDED, BUT ECONOMIC ANALYSIS OF SUCH MARKETS SHOULD KEEP A TIGHTER LINK TO WHAT WE KNOW ABOUT CONVENTIONAL, ONE-SIDED MARKETS THAN SCHOLARS HAVE DONE TO THIS POINT.

We start with the interchange fee. Assuming the fee matters (i.e., is not neutral), the profit-maximizing interchange fee unambiguously maximizes the total volume of transactions, holding constant other network fees. The fee balances the marginal contributions to volume of the two sides of the market. Yet the economic fundamentals that generate this result do not depend on the two-sided nature of the market, despite suggestions that it is the two-sidedness that is responsible for this result.

The expression for the profit-maximizing interchange fee developed in the literature[16] is just a re-expression of the classic Dorfman-Steiner theorem[17] on a firm's optimal (i.e., profit maximizing) advertising,

quality, or promotion. And this has to be the case, given the structure of four-party credit card networks. The inescapable interpretation of the cash flows in a credit card network is that the credit card company (e.g., Visa) sells the right to use its card and network to the merchant (along with the right to offer the same service to consumers) for a price equal to the sum of the interchange fee and the acquirer processing service fee; the interchange fee net of the issuer network fee is used for (i.e., creates incentives for) promotion (advertising and consumer rewards) as well as other issuer services.

The Dorfman-Steiner theorem, which provides an expression for the optimal portion of revenue to allocate to promotion in a conventional market, applies directly to the interchange fee. Even the characterization of the profit-maximizing interchange fee as volume maximizing has an exact parallel to single-sided markets: the Dorfman-Steiner theorem can be interpreted quite naturally as following from a volume-maximization principle. Issuer promotion of the card is undertaken in a decentralized way by issuers in a four-party credit card network rather than entirely by the credit card company. But decentralized promotion is not unusual in the economy at large and does not affect the application of the Dorfman-Steiner principle.

Turning to the normative or policy side of interchange fees, the Dorfman-Steiner parallel to interchange fees leads immediately to the insight that—in the context of profit-maximizing credit card companies—regulating the interchange fee is exactly like regulating promotion decisions of conventional, one-sided firms. Just as we tend not to want to regulate promotional activities of a single firm in our usual one-sided market, so too should we be skeptical of the advantages of regulating a credit card company's interchange fees, when those fees have effects only on the users of that credit card company, all else equal.

We next apply our perspective on the structure of credit card markets to the issue of whether surcharges should be allowed (i.e., whether the no-surcharge rule should be prohibited) or—the opposite policy—whether surcharges should be prohibited.

A no-surcharge rule is parallel to a retail MFN vertical restraint, which requires in a conventional market that a retailer not charge more for one manufacturer's product than for its rival's product. Exploiting

EXPLOITING THIS PARALLEL, WE SHOW THAT NO-SURCHARGE RULES RAISE THE RELEVANT FEES THROUGH TWO EFFECTS

this parallel, we show that no-surcharge rules raise the relevant fees (e.g., the total cost to merchants who then pass along this increased cost to consumers) through two effects: by suppressing competition between credit card companies; and by adding to the incentive for credit card companies to raise credit card fees to merchants by effectively requiring that the cost to final customers of using a credit card in a transaction be spread across consumers using all transactions methods, including cash or other non-credit payment cards such as debit cards.

From a two-sided perspective, increased fees resulting from no-surcharge rules transfer wealth from particular consumers (non-charge card consumers) to the other side of the market for credit card networks. Our perspective is that the rules raise prices to both cash and debit customers and should therefore be prohibited, notwithstanding the positive impact on profits to credit card companies and on promotion and issuer-provided consumer rewards to credit card users.

In competition law in conventional markets, agreements among competitors to set a monopoly price—or to adopt practices that elicit monopoly prices—cannot typically be successfully defended on the basis that higher prices elicit greater promotion or non-price competition.[18] The notion that the competitive market produces the wrong quality or promotion cannot typically be used to justify collective price setting among competitors.

> THE NOTION THAT THE COMPETITIVE MARKET PRODUCES THE WRONG QUALITY OR PROMOTION CANNOT TYPICALLY BE USED TO JUSTIFY COLLECTIVE PRICE SETTING AMONG COMPETITORS.

This insight from one-sided markets should carry over to two-sided markets as well. Otherwise, a simple conversion to a two-sided market structure could be adopted as a strategy to avoid liability for collective price setting. It follows that a practice such as no-surcharge rules can usually be assessed on the basis of its impact on prices rather than promotion if the rule is reached by collective agreement among competitors. For example, if competing banks form a joint venture to issue a credit card (e.g., the creation of Visa) and they adopt a no-surcharge rule ("NSR"), that rule could easily raise antitrust concerns. In the case of a single credit card company (e.g., as Visa is now configured) with market power rather than a joint venture (e.g., as Visa used to be configured), the use of a NSR could be challenged under Section 2 of the Sherman Act since the unilateral decision to adopt NSR could be characterized as a way to extend the market power in credit cards to non-credit card customers.[19]

## II.   REVIEW OF BASIC CASH FLOWS IN A FOUR-PARTY CREDIT CARD NETWORK

We focus on four-party networks in this paper in order to offer a new, or at least different, perspective on the interchange fee that sets the stage for competitive analysis of practices in this market. But the analysis of the competitive effects of no-surcharge rules apply to three-party networks as well. Four-party credit card networks actually involve five parties: the credit cardholder; the bank that issues the credit card (the "issuer"); the merchant; the merchant's bank, which acquires the merchant's accounts receivable (the "acquirer"); and the credit card company. Consider a credit card transaction for $100. After the transaction (setting aside fees for the moment) the acquirer pays the merchant $100 and then collects this amount from the issuer, who then collects payment at the end of the month from the cardholder.

The transactions fees are the central issue in the economics of credit card networks. This is one context in which we cannot simply set aside transactions costs. We illustrate in Figure 1 representative values for the fees associated with a $100 transaction in a market in which merchants are free to surcharge consumers/cardholders. As illustrated in the figure, the acquirer pays a network fee of $0.06 to the credit card company as well as an interchange fee of $1.50 to the issuer. The acquirer's total cost of $1.56 is passed on to the merchant (we assume that the market for acquisition services is competitive and for simplicity that the acquirer has no additional costs). The merchant then passes on the $1.56 to the consumer to some extent via some combination of a surcharge and perhaps a change in the retail price of its product.

> THE TRANSACTIONS FEES ARE THE CENTRAL ISSUE IN THE ECONOMICS OF CREDIT CARD NETWORKS.

Table 1: The Flow of Funds in a Credit Card Transaction with a Surcharge Fee
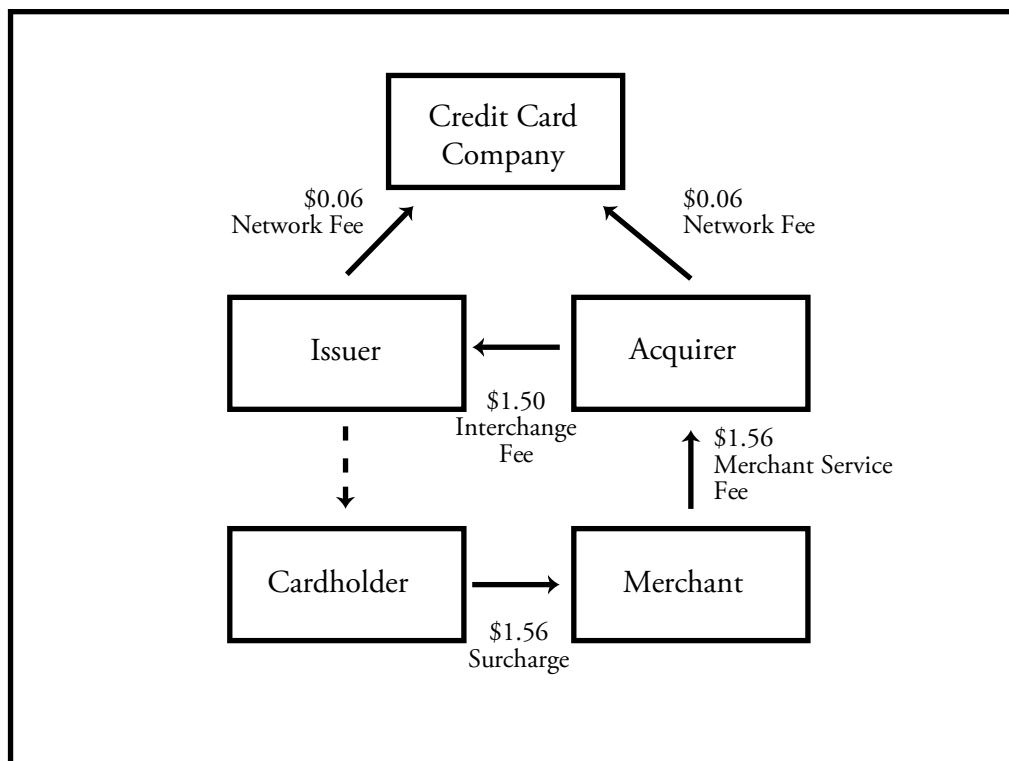
In our example, the merchant passes on the full amount of the $1.56 as a surcharge, although in reality the merchant may surcharge more or less than its cost depending on the relative demand elasticities of those who buy with the card and those who use other transactions methods such as cash, holding all else equal., The issuer receives the interchange fee, pays the issuer network fee, uses some of the funds to cover the costs of its issuing services, uses some to cover the costs of promotion and consumer rewards, and retains the balance as profits.

III.   THE INTERCHANGE FEE

A.   *Profit-Maximizing Interchange Fee: The Two-Sided Market Perspective*

The credit card network is a two-sided market in the sense that both cardholders/consumers and merchants

THE INTERCHANGE FEE CHOSEN TO MAXIMIZE PROFITS BALANCES THE TWO SIDES OF THE MARKET SO THAT THE MARGINAL IMPACTS OF A CHANGE IN THE INTERCHANGE FEE ARE OFFSETTING ON EITHER SIDE.

must be attracted to the network. Neither side will join without sufficient numbers of agents on the other side of the market. The interchange fee is not itself a source of revenue to the credit card company in Figure 1 but rather represents a transfer from one side of the market to the other—from the acquirer/merchant side of the market to the issuer/cardholder side. The interchange fee chosen to maximize profits balances the two sides of the market so that the marginal impacts of a change in the interchange fee are offsetting on either side.[22] This ensures that transaction volume is maximized—all else equal—and volume maximization is the same as profit

maximization because profits equal the product of transaction volume and the sum of network fees.

To be more specific, the price that acquirers pay per dollar of transactions completed on the network is the acquirer network fee plus the interchange fee. We denote this price by $p_a = f_a + I$. The price that issuers pay (receive, if negative) is $p_i = f_i - I$.[23] The total dollar volume or quantity of transactions, **Q**, flowing through the network depends on prices on each side of the market, . An increase in the price on the merchants' side of the market will deter merchants from accepting the cards; this will have a feedback effect on the other side of the market through deterrence of cardholders from taking out the card. There is a similar feedback effect for an increase in the price on the cardholder side of the market.

If we denote the elasticities of transactions volume with respect to the prices on the two sides of the market as and , then straightforward profit maximization by the network shows that the profit-maximizing interchange fee (i.e., the volume-maximizing fee) is characterized by the following expression.

$$\frac{\varepsilon_a}{p_a} = \frac{\varepsilon_i}{p_i}$$

(1)

Only when (1) is satisfied are the marginal impacts of a change in the interchange fee on the two sides of the market offsetting and only then can volume, and profit, be maximized.[24]

B. *Profit-Maximizing Interchange: The One-Sided Market Perspective*

Suppose that the interchange fee, instead of flowing directly from the acquirer to the issuer, spent one millisecond in the accounts of the credit card company. Then the cash flows to and from the credit card company would look quite conventional. The credit card company would be collecting a price from the merchant, p_a, via the competitive acquirer intermediary, and it would be spending some of the price on issuing activities such as promotion via the net payment p_i to the issuer. The remaining funds would cover the credit card company's operating costs, costs of direct advertising, and profits. Like any firm, the credit card company would simply collect revenue and spend some portion of the revenue on promotion. These cash flows are illustrated in Figure 2.

> SUPPOSE THAT THE INTERCHANGE FEE, INSTEAD OF FLOWING DIRECTLY FROM THE ACQUIRER TO THE ISSUER, SPENT ONE MILLISECOND IN THE ACCOUNTS OF THE CREDIT CARD COMPANY.

Figure 2: Equivalent Flow of Funds in a Credit Card Transaction with a Surcharge Fee
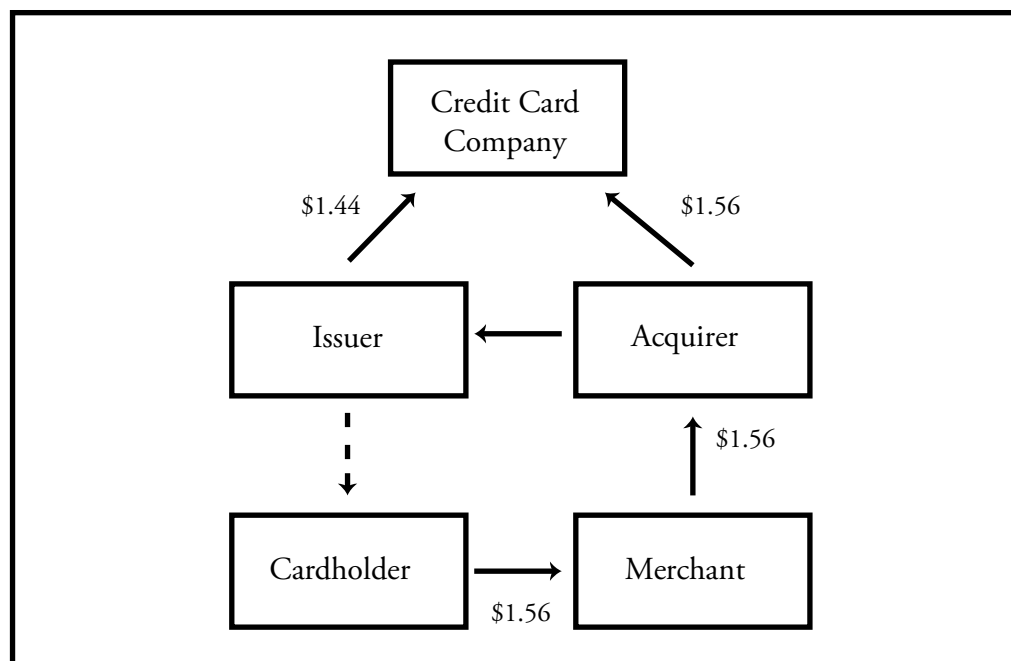
The one somewhat unusual aspect of these cash flows is that the credit card company decentralizes promotion, relying on competition among issuers to elicit promotional activities including consumer rewards.[25] This decentralization of promotion is an elegant aspect of credit card network economics, but does not change the basic characterization of cash flows under our "one millisecond" hypothesis: Revenue is collected from the acquirer and some portion of this revenue is allocated to promotional and quality enhancing activities ("promotion").

Figure 2 depicts our interpretation of interchange activities as promotion in a credit card network in that the credit card company receives $1.56 per $100 transaction and spends $1.44 of this on issuer activities, which increase demand. Like any firm, the credit card company receives revenue from sales of its product or service and allocates a portion of these revenues to promotion (defined as any demand-enhancing activity). The credit card company collects from acquirers revenue per unit—that is, a price—given by $p_a = I + f_a$, and then allocates a portion, $p_i = I - f_i$, of this revenue to promotion. The company sets a price $I + f_a$, and spends a total amount $A = (I - f_i)Q$ on promotion to sell a given volume Q of transactions on its network.

1. The Optimal Interchange Fee and the Dorfman-Steiner Theorem

The conventional theory of optimal advertising or promotion in a one-sided market applies. The Dorfman-Steiner theorem provides the profit-maximizing allocation of funds to advertising.[26] Let p and Q be the price set and quantity sold by a firm and A be the firm's dollar expenditure on advertising; let $q(p,A)$ be the firm's demand and let the elasticities of demand with respect to price and

THE CONVENTIONAL THEORY OF OPTIMAL ADVERTISING OR PROMOTION IN A ONE-SIDED MARKET APPLIES.

advertising be and , respectively. Dorfman-Steiner showed that the following expression is necessary for profit maximization:

$$\frac{A}{pQ} = \frac{\varepsilon_a}{\varepsilon_p}$$

(2)[27]

The Dorfman-Steiner theorem necessarily applies to our hypothetical credit card company that is allocating some portion of revenues to promotion like any other company.

It is straightforward to show that (1) and (2) are equivalent if we interpret A in (2) as the net amount of funds allocated to issuers, **p** as the fee paid by the acquirer per dollar transacted on the network, and **Q** as the total volume transacted on the network.[28]

The equivalence of profit-maximizing promotion and profit-maximizing interchange fees must follow as a matter of simple economics, not just algebra. The funds directed towards issuers are the funds allocated to promotion and issuer quality enhancement (or at least to the issuer-controlled dimensions of these variables). Our hypothesis that funds spend a millisecond in the accounts of the credit card company cannot possibly matter, since the credit card company controls the value of the interchange fee with or without this hypothesis. A credit card company devotes pi dollars to issuer activities per dollar transacted, of the total of $p_a Q$ raised from acquirers. The Dorfman-Steiner theorem and the profit-maximizing interchange fee describe the identical optimization problem, so the solutions must be equivalent.

2.   The Output Maximization Principle

The output-maximization property of the profit-maximizing interchange fee also does not depend on the two-sided market nature of the credit card market. Consider a firm in a conventional market making a decision on the following variables: advertising expenditure per unit, e; allocation per unit to the sum of operating expenses per unit and profit per unit, which allocation we denote as **x** (**x = c +** π, where these are per-unit variables); and price, **p**. We have **p = e + x** as an accounting identity. The demand can be expressed as **q(p,e)**. The firm's profit-maximizing decision can be expressed as the choice of any two elements in {**e,x,p**}, for example **x** and **e**. Conditional upon **x**, the profit-maximizing choice of **e** will maximize volume since profit = **(x-c) q(x+e, e)**. (At a given **x**, **p**, and **a** move together one-for-one so the choice of either **p** or **a** maximizes volume. Solving this output maximization problem yields again the Dorfman-Steiner theorem.)

Moving to the credit card context we find a special case of this general output-maximizing principle. In the credit card context we have $p = I + f_a$, $e = I - f_i$ and $x = f_a + f_i$. The general principle that maximizing profit with respect to e, given x, also maximizes output implies directly that the profit-maximizing interchange fee maximizes output, at given network fees. This result has nothing to do with the two-sided nature of markets

and instead is a straightforward result of the Dorfman–Steiner model applied to conventional markets.

Note that we cannot draw any inferences in the credit card setting about market power or pricing efficiency from the output-maximizing property of the interchange fee. Any firm with any degree of market power chooses price to maximize output, holding **x** constant.

Drawing the connections between issuer activities in a credit card network and promotion by a firm in a single-sided market sets the stage for the analysis below of the competitive impact of the no-surcharge rule.

First, however, we must complete our characterization of the profit-maximizing interchange fee by discussing the theoretical conditions under which changes in the interchange fee are neutral or not. If the interchange fee were irrelevant, then the discussion above would be irrelevant because there would be no profit-maximizing interchange fee.

C.   *The Neutrality—or Not—of the Interchange Fee*

Issuers promote their cards and provide consumer rewards in a number of dimensions. They advertise, set interest rates, set terms of payment, provide reward points, and in some cases offer consumers a percentage refund on their monthly payments.

We consider here the consequences of a simple set of assumptions, which we label the assumption of a "perfect credit card network." A perfect credit card network is one with rational agents, and is free of any transactions costs other than the explicit fees that we have specified. In particular, merchants in such a market can set precise issuer-specific surcharges; and issuers can offer precise rebates to consumers who use the issuer's credit card, with consumers making credit card transaction decisions on the basis of surcharges net of rebates. (The label "perfect credit card network" parallels economists' use of the term "perfect markets".) Suppose that merchants can set a surcharge fee precisely (down to a single basis point) and that consumers make credit card purchase decisions based on the opportunity cost represented by the surcharge net of any rebates offered on purchases by the issuer.

Under these assumptions, the level of the interchange fee is completely irrelevant. Any change in the interchange fee, holding constant the other network fees, is offset by prices along the network that leave all agents with the same payoff and taking the same actions.

To prove this result in a simple way, we make liberal use of a basic proposition in the economics of public finance: The side of a market on which the tax is imposed is irrelevant to the distribution of the tax burdens. Price will adjust in the market so that the incidence on buyers and sellers is the same regardless of which set of economic agents pays the tax.[29]

Consider a perfect credit card network "in equilibrium": That is, the issuer is choosing the profit-maximizing level of promotion and consumer rewards in each dimension, competing for cardholders. Merchants are setting prices to maximize profits. And cardholders are purchasing quantities given the merchants' prices and the issuer's promotion and rebates on credit card payments.

Suppose that the credit card company raises the interchange fee by one percentage point. The increased "tax" of one percent imposed on the acquirer could equivalently be imposed on the merchant, since it is a tax on each dollar transacted between the acquirer and the merchant. But a tax on the merchant per dollar unit of transactions is equivalent to a tax on the consumer/cardholder on the same transaction. And a tax on the cardholder is equivalent to a tax on the issuer because the cardholder and the issuer are engaged in a contract that involves a payment to the consumer per dollar transacted, the rebate to the consumer. To shift the tax incidence from the consumer to the issuer when the consumer pays the tax instead of the issuer, a one-percent additional rebate is offered.

In short, the basic tax-incident-irrelevance theorem tells us that a one-percent increase in the interchange fee is equivalent to the sum of the one-percent additional benefit on each dollar transacted that the issuer receives directly from the acquirer plus a one-percent cost on each dollar transacted that is effectively transferred—with offsetting price adjustments—around the circle of the network. The price adjustments are the one-percent higher merchant fee, the one-percent higher surcharge, and the one-percent higher rebate on credit card payments. At these new prices, and with the new interchange fee, the consumer purchase decisions will obviously remain unchanged and the issuer's marginal costs of promoting in each dimension also remain unchanged. The change in the interchange fee is irrelevant.[30]

How did we get a theory and formula for the profit-maximizing interchange fee in the previous discussion, when we have irrelevance of the interchange fee in a perfect credit card network? The answer is in the mathematical assumption in the previous discussion that a profit-maximizing interchange fee existed. Specifically, our characterization of the profit-maximizing interchange fee followed from the first-order conditions for the volume-maximizing interchange fee. Using the first-order conditions to characterize the profit-maximizing interchange fee involves an assumption that the volume of transactions is a strictly concave function of the interchange fee, whereas under the assumption of a perfect credit card network this assumption fails.

HOW DID WE GET A THEORY AND FORMULA FOR THE PROFIT-MAXIMIZING INTERCHANGE FEE IN THE PREVIOUS DISCUSSION, WHEN WE HAVE IRRELEVANCE OF THE INTERCHANGE FEE IN A PERFECT CREDIT CARD NETWORK?

The interchange fee in reality seems not to be irrelevant. Regulatory constraints on interchange are contentious and have some bite, which they would not if interchange fees were irrelevant. Departures from the world of a perfect credit card network can explain this. Consumers may react differently to a discount than to a surcharge. Importantly, surcharges are often prohibited by the credit card companies (the no-surcharge rule), or constrained by regulation. Moreover, there can be costs in transacting with a different surcharge on each payment—and differential surcharges and rebates are necessary for interchange irrelevance in the face of

interchange fees that differ across various credit cards.

IV. COMPETITIVE IMPACT ANALYSIS OF NO-SURCHARGE RULES

A. No-Surcharge Rules as Retail MFNs

We now assess the competitive impact of NSRs by examining their effect on equilibrium prices to acquirers and merchants. We set aside any incentives for increased promotion via a change in interchange fees, but return to this issue below. The no-surcharge rule is an example of a retail MFN restriction, which is a restraint imposed by a manufacturer that a retailer not charge more for that manufacturer's product or service than for the products or services of its rival producers.[31] In our context, the service being provided by the upstream credit card company is the right to transact with its credit card. The price charged to the acquirer/merchant for this service is the acquirer service fee plus the interchange fee.

> THE NO-SURCHARGE RULE IS AN EXAMPLE OF A RETAIL MFN RESTRICTION, WHICH IS A RESTRAINT IMPOSED BY A MANUFACTURER THAT A RETAILER NOT CHARGE MORE FOR THAT MANUFACTURER'S PRODUCT OR SERVICE THAN FOR THE PRODUCTS OR SERVICES OF ITS RIVAL PRODUCERS

To illustrate the effect of the NSR on the equilibrium involving credit cards, we consider first a duopoly (which sheds light on the impact of NSRs on competition between credit card companies)[32] and then a monopoly credit card firm facing competition from the consumers' alternative to transact in cash.

1. Duopoly

In a duopoly, in which two symmetric firms sell through the same retailers downstream, a retail MFN raises prices through two effects. The first we can label the "competition-suppression effect." This effect operates by removing the incentive to cut prices. Suppose for simplicity that the demands for the manufacturers' products are symmetric and that retailers downstream are competitive. Consider the incentive for either manufacturer to cut its wholesale price to the retailer if both manufacturers are currently setting the joint profit-maximizing prices. This incentive is zero. If one manufacturer cut its wholesale price and that leads a retailer to cut its retail price, the manufacturer knows that its rival's retail price will follow its own retail price cut, dollar for dollar. The retail MFN eliminates the sales gain of stealing sales from a rival by undercutting the rival's price, which is the essential competitive mechanism. Once both manufacturers adopt the retail MFN, there is no incentive at all to price below the collective monopoly price(s).

But the anticompetitive impact of the restraint does not end here. The second effect of a MFN is to create an incentive to raise prices above the jointly profit-maximizing prices. We label this effect the "cost externalization effect." Suppose, starting again from the position of both firms setting the joint profit-maximizing prices, that one firm considers raising its wholesale price by one dollar. If its own customers at the retail outlet bore the full brunt of this price increase via a one-dollar increase in the retail price, then the increase beyond the joint monopoly prices would not be profitable.

Its own retail consumers, however, bear only *half* the consequence of the price increase: The competing retailers downstream charge a common retail price based on the average wholesale price and therefore raise the price of each product, in response to the one dollar wholesale price increase in one product, by about 50 cents. The joint monopoly price is not sustainable as a Nash equilibrium because each upstream manufacturer has the incentive to increase the wholesale price, due to the negative externality imposed on the rival manufacturer. The combination of the two effects of the MFN mean that the equilibrium price after the duopolists have adopted a MFN is, in the simplest theoretical model, greater than the monopoly price. Effectively the restraint changes the two substitute products into complements, since an increase in the price of one lowers the demand for the other once the restraint is adopted. For complements, the non-cooperative price always exceeds the joint profit-maximizing price.

The application of these effects of a MFN to the use of a NSR in credit cards is direct. Think of two competing credit card companies and, for simplicity, ignore cash and ignore promotion including rebates to credit card customers. In the absence of the NSR, merchants would compete with each other by differentially surcharging retail transactions on each card depending on the particular card's fees to the merchant. But with a NSR, that is not possible and the consequence is that competition between the two card companies gets distorted in the same way as the competition between manufacturers gets distorted in our previous example. Both the competition-suppression effect and the cost externalization effect are at work. The Canadian Competition Tribunal, in the 2010 Canadian case involving Visa and MasterCard, discussed both effects extensively in its assessment of the overall competitive impact of the NSR.[34]

This analysis of the impact of the no-surcharge rule by analogy to the MFN in a conventional market treats only one side of the credit card market: the price to acquirers. What is the consequence of the suppression of competition for the issuer side? An answer to this question again draws on price theory of conventional markets. Stigler[35] pointed out that when firms maintained monopoly prices, non-price competition between the firms is magnified, which eats into firm profits. (Stigler's analysis was for the case of cartel pricing, but the same principle holds for price competition suppressed through the adoption by individual firms of practices that suppress price competition.) High prices lead to greater promotion but, unless promotion is a perfect substitute for prices, rents will not be completely dissipated through the increased intensity of the non-price competition. Here, interchange fees will rise, with greater promotion by issuers, but not enough to offset the higher prices—unless cash rebates as a component of promotion perfectly offset surcharges.

2. NSR and the Competition From Cash

The effect of a NSR is evident even when there are no competing card companies and the only alternative means of payment is cash. To focus on the cash alternative, suppose that there is only one card company. The imposition of a NSR here is assumed to mean that transactions through

the credit card and through cash must be at the same price—and therefore any increase in price charged by the credit card service provider is spread over all transactions. The extraction of this transfer from cash customers creates an incentive for even a monopoly credit card company to raise its fees above the fee that would otherwise be profit maximizing.

THIS SITUATION CREATES THE INTERESTING POSSIBILITY THAT CASH CUSTOMERS COULD SUE THE CREDIT CARD COMPANY FOR IMPOSING THE NSR SINCE THE NSR ALLOWS THE MONOPOLY CREDIT CARD COMPANY TO EXTEND ITS MONOPOLY FROM THE CREDIT CARD MARKET TO THE PREVIOUSLY COMPETITIVE MARKET FOR THE USE OF CASH TO TRANSACT.

However, in this case—unlike the pure duopoly case—the monopoly credit card firm is better off raising its price (e.g., the f_a fee it charges acquirers)[37] above the monopoly level because of the cost externalization effect. One way to think of this is that the NSR enables the credit card company to exercise market power over cash customers and collect a "tax" on them equal to the elevation in the retail price that occurs as a result of the NSR. This situation creates the interesting possibility that cash customers could sue the credit card company for imposing the NSR since the NSR allows the monopoly credit card company to extend its monopoly from the credit card market to the previously competitive market for the use of cash to transact.[38]

In both the duopoly credit card model and the monopoly credit card/cash model, for simplicity we have set aside a detailed analysis of decisions on promotion (whether through rebates or other means) or interchange fees. But there is one condition under which it is essential to incorporate interchange decisions in analyzing the impact of the NSR. Consider the "perfect credit card network" condition in our earlier analysis, in which changes in the final retail price to consumers could be offset perfectly by opposite changes in rebates on credit card bills. In the case of a credit card duopoly with no cash customers, if the perfect credit card market condition holds then the NSR has no impact at all; the competition between the two credit card companies will be reflected in higher rebates on credit card bills, which offset perfectly the suppressed competition from the NSR.[39]

In contrast, in the case of a monopoly credit card firm and cash customers, the presence of neutrality (in the absence of the NSR) does not undercut the effect of the NSR that cash customers are harmed. In the duopoly case all retail customers (i.e., the customers of both credit card firms) can receive rebates from their firms that undo the impact of the restraints. In the case with cash customers, the availability of cash rebates actually magnifies the incentive to raise the price of credit card services; the price increase for the credit card company's own customers can be offset with the rebates allowing a complete externalization of the impact of price increase on cash customers.

We have outlined some of the effects of the NSR in a credit card market, making some simplifying assumptions to highlight our points. A more precise analysis requires a full model of competing networks in which the decisions of cardholders, issuers, credit card firms, and merchants are explicit; in which the concept of a competitive retail sector is spelled out; and in which the impact of the NSR is set out. We offer the fuller analysis in our companion paper.

V.   CONCLUSION

Two-sided markets, and especially credit card markets, have received much attention with an emphasis on understanding the special features that two-sidedness creates. Although we agree that two-sidedness presents the necessary framework to understand credit card markets and other markets, we have offered a perspective that the essential economic forces at work in the basic credit card network, and in the impact of no-surcharge rules, have little to do with two-sidedness. Failure to understand that insights from one-sided markets also apply to two-sided markets obscures rather than clarifies the analysis of how to reach sound policy decisions in credit card markets.▲

WE HAVE OFFERED A PERSPECTIVE THAT THE ESSENTIAL ECONOMIC FORCES AT WORK IN THE BASIC CREDIT CARD NETWORK, AND IN THE IMPACT OF NO-SURCHARGE RULES, HAVE LITTLE TO DO WITH TWO-SIDEDNESS.

[1]   Dennis Carlton is David McDaniel Keller Professor of Economics at Booth School of Business, University of Chicago and NBER; Ralph Winter is Professor, Strategy and Economics Division, Sauder School of Business, UBC. We thank David Evans, Alan Frankel, Gregory Pelnar, Allan Shampine, Craig Wildfang, and Yulai Zhao for helpful comments. Both authors have been involved in litigation adverse to Visa, MasterCard, and other credit card companies. In addition, Carlton has worked for Discover Card. This paper outlines themes developed more formally in a companion paper: Dennis W. Carlton & Ralph A. Winter, *Vertical Restraints in Two-Sided Markets: Credit Card Merchant Rules,* in preparation.

[2]   Hayashi & Maniff offer an overview of government regulation in payment card markets, Fumiko Hayashi & Jess Leigh Maniff, *Public Authority Involvement in Payment Card Markets: Various Countries: August 2014 Update,* Federal Reserve Bank of Kansas City (2014).

[3]   Alan S. Frankel & Allan L. Shampine, *The Economic Effects of Interchange Fees* 73 Antitrust L. J. 627 (2006).

[4]   Jean-Charles Rochet & Julian Wright. *Credit Card Interchange Fees*, 34 J. Banking & Finance 1788 (August 2010). *See also* Jean-Charles Rochet, *The Theory of Interchange Fees: A Synthesis of Recent Contributions*, 2 Rev. Network Economics 1 (June 2003); Jean-Charles Rochet & Jean Tirole, *An Economic Analysis of the Determination of Interchange Fees in Payment Card Systems*, 2 Rev. Network Econ. 1 (June 2003).

[5]   Julian Wright, *Why Payment Card Fees Are Biased Against Retailers,* 43 RAND J. Econ. 761 (December 2012).

[6]   Benjamin Klein, Andres V. Lerner, Kevin M. Murphy, & Lacey L. Plache, *Competition in Two-Sided Markets: The Antitrust Economics of Payment Card Interchange Fees,* 73 Antitrust L.J. 571 (2006).

[7]   Eric Emch & T. Scott Thompson, *Market Definition and Market Power in Payment Card Networks*, 5 Rev. Network Econ. 45 (2006).

[8]   David S. Evans & Richard Schmalensee, *The Economics of Interchange Fees and Their Regulation,* Competition Pol'y Int'l 1-33 (2007).

[9]   Weyl develops a model in which heterogeneity of parties on each side of a two-sided model is essential; E. Glen Weyl, *A Price Theory of Multi-sided Platforms,* 100 Amer. Econ. Rev. 1642 (2010).

[10]   Dennis W. Carlton & Alan S. Frankel, *The Antitrust Economics of Credit Card Networks,* 63 Antitrust L.J. 643 (1995); Joshua Gans & S.P. King, *The Neutrality of Interchange Fees in Payment Systems,"* 3(1) Topics

in Econ. Analysis & Pol'y, 3 (2003).

[11]   Julian Wright, *Optimal Card Payment Systems,* 47 Eur. Econ. Rev. 587 (2003).

[12]   Andrew Boik & Kenneth S. Corts, T*he Effects of Platform MFNs on Competition and Entry, Working Paper,* University of Toronto (June 2013).

[13]   Winter analyzed the no-surcharge rule as equivalent to a retail MFN in his testimony in *The Commissioner of Competition v. Visa Canada Corporation and MasterCard International Incorporated et al,* CT-2010-010.

[14]   Jean-Charles Rochet & Jean Tirole, *Cooperation Among Competitors: Some Economics of Payment Card Associations,* 33 RAND J. Econ. 549 (Winter 2002). *See* also Hélène Bourguignon, Renato Gomes, & Jean Tirole, *Card Surcharges and Cash Discounts: Simple Economics and Regulatory Lessons,* 10(2) Competition Pol'y Int'l (Fall, 2014). The latter paper focuses on the lack of consumer information about surcharging. In this paper, when we discuss surcharging, we assume consumers are fully informed about all fees.

[15]   *See* Rochet & Tirole (2003), *supra* note 4; Jean-Charles Rochet & Jean Tirole, *Two-sided Markets: A Progress Report*, 37 RAND J. Econ. 645 (2006); Jean-Charles Rochet & Jean Tirole, *Competition Policy in Two-sided Markets, with a Special Emphasis on Payment Cards, Handbook of Antitrust Economics,* 543-582 (Paulo Buccirossi, ed. 2008).

[16]   E.g., Emch & Thompson, *supra* note 7; Klein, Lerner, Murphy, & Plache, *supra* note 6.

[17]   Robert Dorfman & Peter O. Steiner, *Optimal Advertising and Optimal Quality,* 44 Amer. Econ. Rev. 826 (1954).

[18]   We are not suggesting that cooperation on pricing or restrictions on competitive pricing are always per se illegal. Firms in a joint venture can defend a restriction on the basis that without the restriction on competition some product could not be provided and therefore the rule of reason should apply. This is called the problem of "characterization" in the legal literature. The line between a "new" product and a product with different promotional characteristics is a thin one for an economist to draw, but we note that courts have been relatively clear that "new" products are the exception, not the rule, in determining the legality of collaboration among competitors who collectively have market power.

[19]   We are not aware of any successful case brought on this theory, indicating that there might be legal hurdles to overcome. Notice that the harm is occurring in the non-credit card market, not the credit card market.

[20]   The dashed arrow in Figure 1 indicates that an issuer may or may not provide cash rebates (on the monthly cardholder bill) for purchases with the credit card.

[21]   Some regulations that allow surcharging also constrain the amount of the surcharge.

[22]   This assumes that a profit-maximizing interchange fee exists (i.e., that changes in the fee are not neutral). We explore the conditions for this below.

[23]   With three variables and only two prices, there is a normalization required. For simplicity, set $\mathbf{f_i=0}$.

[24]   Emch & Thompson, *supra* note 7.

[25]   Such decentralization occurs in industries other than credit cards. For example, in many industries firms rely on independent salesmen who receive a commission for each sale.

[26]   We use the term "promotion" to include all issuer activities related to promotion such as advertising, quality improvement, or consumer rebates. To keep the terminology here the same as Dorfman-Steiner, we

refer only to advertising.

[27]   The Dorfman-Steiner theorem follows from the first-order conditions in the maximization of profit, **(p – c)q(p,A) – A**, with respect to **p** and **A**. We take the convention of expressing the price elasticity as a positive number.

[28]   The elasticities in (2) refer to the elasticities of the demand function **q(p,A)** with respect to **p** and **A**; and the elasticities in (1) are of the demand function $\tilde{q}(p_a, p_i)$. We remind the reader that **A** is total advertising while **pa = (I+fa)** is price to the acquirer and $\mathbf{p_i} = \mathbf{I} - \mathbf{f_i}$. We have $Q = q(p, A); Q = \tilde{q}(p_a, p_i); A = p_i Q; P = P_a$ and **pi = I-f_i.** It follows directly that

$$\tilde{q}(p_a, p_i) = q(p_a, p_i \tilde{q}(p_a, p_i)) \tag{3}$$

Totally differentiating (3) (in logs) with respect to $\mathbf{p_a}$, $\mathbf{p_i}$ yields

$$\varepsilon_i / \varepsilon_a = \varepsilon_A / \varepsilon_p \tag{4}$$

Rewrite (1) as

$$p_i / p_a = \varepsilon_i / \varepsilon_a \tag{5}$$

Multiplying the denominator and the numerator of the LHS of (5) by Q and substituting (4) yields (2). Thus the expressions (1) and (2) are equivalent.

[29]   This proposition holds regardless of whether the market is competitive or one in which firms have market power.

[30]   The concept of interchange fee irrelevance is not to be confused with another form of irrelevance. In the triplet of fees in the credit card network **(I, f_i, f_a)** consisting of the interchange fee and the two network fees, there are only two degrees of freedom. If the interchange fee is raised by one basis point simultaneously with a one basis point increase in the issuer network fee and a one basis point decrease in the acquirer network fee, the net cash flows to each party in the network are unchanged. This is an accounting identity, not economic neutrality that depends on equilibrium price adjustments. The economic neutrality discussed in the text is about the irrelevance of the interchange fee holding network fees constant, unlike the accounting identity.

[31]   We analyze the impact on pricing of the adoption of MFN's by both firms in a duopoly (i.e., the "pricing subgame" following adoption of MFN's by both parties). In our companion paper we discuss the full game in which the adoption of MFN by each firm is endogenous. We also ignore here the extensive margin on which the retailers can drop a product.

[32]   We recognize that MFNs can sometimes serve pro-competitive functions such as when they prevent free riding. We abstract from those effects here.

[33]   Note that for these effects only one manufacturer need have a MFN that requires that all retail prices be equal (assuming symmetry of the duopolists) regardless of the method of transacting as long as both manufactures sell through the same retailers.

[34]   *The Commissioner of Competition v. Visa Canada Corporation and MasterCard International Incorporated et al,* CT-2010-010.

[35]   George J. Stigler, *Price and Non-Price Competition,* 76 J. Political Econ. 149 (1968).

[36]   The economics literature on these types of practices typically focuses on the impact on prices, and

ignores the impact of the practices on non-price competition.

[37]   Recall that the credit card company charges issuers $\mathbf{f_i}$ and acquirers $\mathbf{f_a}$.

[38]   Although cash may be an alternative to credit cards for some to transact, we assume—as many courts have found—that there is a sufficient convenience to using credit cards so that cash is not in the same antitrust market as credit cards. As a result of the NSR, the cost of transacting in cash rises from 0. It is possible that if there were several credit card firms competing with each other that the effect of the NSR would be to leave profits of credit card companies unchanged at the competitive level but credit card customers could benefit from increased rewards funded by the "tax" on cash customers.

[39]   This is analogous to the impossibility of sustaining monopoly profits in a conventional duopoly through a cartel agreement or retail MFN when there is a dimension of promotion that substitutes perfectly for lower prices.