Market Definition and Market Power in Payment Card Networks

Eric Emch T. Scott Thompson*

September 14, 2005

Abstract

We discuss competition among payment card networks, and in particular how antitrust practitioners might approach questions of market definition and market power in these markets. Application of the hypothetical monopolist test to define markets, and the use of traditional metrics to measure market power, may be less straightforward for card networks than for many markets. The interrelationships between network pricing to merchants, to bank issuers of payment cards, and indirectly, to final consumers, complicates the analysis. The "two-sidedness" of the market does not, however, overturn the basic logic of the hypothetical monopolist test or traditional measurements of market power. We demonstrate some practical ways to apply these antitrust principles to competition among payment card networks.

*Antitrust Division, U.S. Department of Justice, 600 E Street NW, Suite 10-000, Washington, DC 20531. The authors are Assistant Chiefs of the Competition Policy Section and the Economic Regulatory Section, respectively, of the Antitrust Division. The authors served as staff economists for the Department of Justice in U.S. v. Visa USA, *et al.* and U.S., *et al.* v. First Data Corporation and Concord EFS, Inc. The views expressed in this paper are those of the authors and are not purported to reflect the views of the U.S. Department of Justice.

1 Introduction

Payment card networks have come under the increasing scrutiny of competition authorities in the United States and around the world in recent years. The Antitrust Division of the United States Department of Justice brought two recent cases centering around competition between payment card networks. Private litigants in the United States have successfully challenged the legality of Visa and Mastercard's "Honor All Cards" rules, and unsuccessfully (so far) challenged the legality of interchange fees set by credit card associations.[1] Regulators around the world have capped multilateral interchange fees charged by the associations, and in some cases, banned association "no surcharge" rules.[2]

At the heart of many of these cases is the nature of competition between payment card networks. This is particularly true in the two Justice Department cases with which we are the most familiar. The first of these, U.S. v. Visa USA, *et al.*, challenged bylaws of the Visa and Mastercard associations that prevented members of each association from issuing cards on proprietary networks such as American Express and Discover.[3] The District Court, upheld by the Appellate court, ruled that these restrictions, by denying critical scale to ri-

[1] *See*, In re Visa Check/MasterMoney Antitrust Litigation, 297 F. Supp. 2d 503, re: "honor all cards rule;" and, National Bancard Corp. ("NaBanco") v. VISA USA, Inc., 596 F. Supp. 1231, re: interchange rates.

[2] For example, the Royal Bank of Australia capped interchange fees and banned no-surcharge rules in 2002. The decision was phased in over the course of 2003. In 2002, Visa agreed to a settlement with the European Commission that capped interchange on cross-border transactions. In Denmark, interchange fees for domestic card payments are prohibited by law. Sweden and the Netherlands have also outlawed no-surcharge rules. For a comprehensive summary of these developments, see Weiner and Wright (2005).

[3] The government also challenged the dual governance of the associations – the fact that Mastercard board banks were members of Visa and typically had substantial portfolios of Visa cards, and vice-versa. The Court ruled against the government on this count, in part because it found that in the years leading up to the decision, the board banks of each association had become increasingly dedicated to their respective associations.

vals of the card associations, hampered competition between payment card networks.[4] Since the District Court's Final Judgment became effective in late 2004, American Express has announced issuing partnerships with Visa and Mastercard member banks, and Discover has purchased the PIN debit network Pulse.[5]

In late 2003, in U.S., *et al.* v. First Data Corporation and Concord EFS, Inc., the United States sued to block the merger of two of the largest PIN debit card networks in the United States, NYCE and STAR,[6] on the grounds that the merger would substantially reduce competition among PIN debit card networks. The complaint alleged that the merger would yield higher prices to merchants for PIN debit network services, and that at least some of these higher prices would be passed on to final consumers. On the eve of trial, the parties resolved the government's concerns by agreeing to divest the NYCE network, which was bought by Metavante Corp.

An issue in these recent payment card cases has been the question of appropriate antitrust market definition for markets including payment card network services.[7] Such markets are often characterized as "two-sided" because a payment card network completes a trans-

---

[4] *See United States v. Visa USA, Inc.*, 163 F. Supp.2d 322, 335-39 (S.D.N.Y. 2001), aff'd 344 F.3d 229 (2d Cir. 2003).

[5] PIN debit payment cards are similar to (and often the same as) bank ATM cards in that a card holder authorizes a funds transfer from his or her bank account by entering a personal identification number (PIN) into a point of sale terminal. In contrast, a signature is all that is required to authorize a signature debit transaction. Signature debit transactions typically take some time to process, so funds are not immediately deducted from the card holder's bank account. In contrast, verification of available funds and transfer of funds are immediate in a PIN debit transaction. A single card can often be used for either type of debit transaction or for ATM transactions.

[6] The network merger was part of a larger merger of the parent companies, First Data Corp. and Concord EFS. *See* Verified Complaint, *United States v. First Data Corp.*, Civ. A. No. 03CV02169 (Oct 23, 2003 D.D.C. 2003), available at `http://www.usdoj.gov/atr/cases/first0.htm`.

[7] For an overview of the approaches to market definition taken in several of these cases, see Hesse and Soven (2005).

action by transferring information and funds between merchants and card issuers or their agents, typically charging a separate (possibly negative) price to each party to the transaction. As a consequence, demand for network transactions responds on two separate margins to prices charged by the network for these services. These margins are determined jointly by the responses of card holders, card issuers and merchants to the pricing and other incentives that they face.

In the wake of the increasing antitrust activity in the payment card arena, a growing literature has examined the economics of two-sided markets, in which a platform intermediates demand between two separate groups of purchasers, and demand of each group depends on usage on the other "side" of the market.[8] One robust finding of this line of research has been that welfare-maximizing and profit-maximizing prices on each side of the market depend on cost and demand on both sides of the market. Thus, the conventional wisdom that pricing close to marginal cost is efficient does not hold when each side of a two-sided market is examined in isolation. It may well be that efficient pricing dictates a price above, or below, marginal cost on a particular side of the market.

The "two-sidedness" of a market makes traditional antitrust analysis of relevant markets and market power somewhat less straightforward. It does not, however, invalidate the exercise altogether, as has sometimes been alleged in some of the antitrust actions described above. Some commentators have explicitly or implicitly suggested, in particular, that the widely cited hypothetical monopolist paradigm for market definition, set out in the *Horizontal Merger Guidelines* (U.S. Department of Justice and Federal Trade Commission 1997, hereinafter *"Guidelines"*) and elsewhere, is not appropriate in two-sided markets. A central purpose of this paper is to show that this is not the case. The logic of market definition in the *Guidelines* can be

[8] *See*, *e.g.* Schmalensee (2002), Evans (2002), Rochet and Tirole (2003), Evans and Schmalensee (2005), and Guthrie and Wright (2005).

applied to two-sided markets generally, and to markets for payment card network services in particular. We discuss practical aspects of this application and some questions that may arise, including the question of which methods for assessing market power in two-sided markets may be appropriate.

Since the analysis is directly dependent on an understanding of monopoly pricing incentives, the next section provides an extended discussion of monopoly pricing in two-sided markets. In the course of this discussion we also consider measures of market power that naturally arise from application of the hypothetical monopolist paradigm to two-sided markets. Section 3 discusses how this theoretical framework provides guidance for implementing the hypothetical monopolist tests of the *Guidelines*. Section 4 sketches scenarios under which market power may be exercised on one side of the market or the other in the context of payment card networks. A final section summarizes our conclusions.

2 Monopoly pricing in two-sided markets

We first consider the economics of a monopoly payment card network, in order to better understand the application of the *Guidelines* hypothetical monopolist test in this framework. We assume a single firm monopolizes sales of network transactions, which we take to be a homogeneous service. It sells network services to payment card issuers and to merchants. For purposes of this paper we abstract away any services associated with the transaction provided by other intermediaries, such as a merchant acquiring bank or a merchant payment processor. Implicitly we are assuming that these services are competitively priced and procured separately from network services. Likewise, we assume that the firm is not vertically integrated into card issuance.[9]

[9] While the two largest payment card networks, Visa and Mastercard, have historically been owned by issuing and acquiring banks, Mastercard is in the process of becoming less

Demand for the network service is assumed to depend on three prices set by the monopolist. These are a pair of switch fees s_m and s_i charged to merchants and issuers respectively, and an interchange fee X that is charged to merchants and paid directly to the card issuer. Historically, but not today, interchange fees for PIN debit networks were paid in the other direction, from issuers to merchants.[10] This case can be accommodated easily by allowing X to take negative values, if necessary, but generally we will assume that all three of these prices are positive. Since interchange fees are not retained by the network, the net revenue collected per transaction by the monopolist is the sum of the two switch fees, which we denote $s \equiv s_m + s_i$. We assume that the network incurs fixed marginal cost of c for each transaction so that its variable margin on each transaction is $s - c$.

Much of the recent literature on payment card networks has focused on the interchange rate, and whether that "price" is set at the efficient level under monopoly and under particular forms of network competition.[11] Our focus here, for the purposes of analyzing market definition and market power in payment card networks, is not on interchange *per se*, but on the net price charged by the network to merchants and issuers. Since interchange fees are not retained, an independent network monopolist has little incentive to exercise market power through excessive interchange rates. Rather the total price charged by the network and retained for itself is the proper focus of both the hypothetical monopolist test and any subsequent market power analysis.

The net per-transaction prices charged by the monopolist to mer-

---

so. It has announced an IPO to be held in 2006, with 85% of voting shares available to non-bank investors.

[10]Interchange paid from issuer to merchant is still the norm for PIN debit cards in at least one country – Australia – where in recent years interchange fees ranged from 18 to 25 cents per transaction, paid by the issuer to the acquirer. (EFTPOS Industry Working Group 2002, p. 4).

[11]*See, e.g.* Schmalensee (2002), Rochet (2003), and Guthrie and Wright (2005).

chants and issuers are denoted p_m and p_i respectively. These are related to the first three prices by the relationships $p_m = s_m + X$ and $p_i = s_i - X$. It follows that the network's net revenue per transaction can also be written $p_m + p_i$ since this equals the sum of the two switch fees. Accordingly, we will sometimes refer to s as the 'total price' for a network transaction. Provided interchange rates are large relative to switch fees, the net price charged to card issuers for each transaction will be negative. We assume this is the case here, although the assumption isn't needed for our conclusions. However, to the extent that the net price to issuers is negative, an exercise of market power on the issuer side of the market will resemble monopsony rather than monopoly behavior.

We assume that the network faces a single demand that depends separately on each of the two net prices p_m and p_i. That is, the quantity of transactions processed is given by

$$Q = Q(p_m, p_i). \tag{1}$$

The two-sided nature of the market is reflected in this assumption about demand. Because demand depends on two prices separately, the monopolist faces competition from alternative products and services on two separate margins, and profit maximizing behavior will require the monopolist to account for the marginal revenues from both margins when choosing prices to maximize profits. We assume demand is downward sloping with respect to each net price. Note, however, that when p_i is negative, as we assume here, then this means that demand increases as p_i becomes more negative, corresponding to an increase in the net payment made to issuers.

Network demand responds to the merchant price because merchants can respond to higher or lower prices for network services in one or more ways. At the most basic level, a merchant must decide whether or not to accept cards associated with a network at all. Also, PIN debit cards can be associated with more than one network. When this occurs, merchants may be able to choose to route the transaction over one network rather than the others associated with the card, based on prices charged to the merchant by the networks. A third possibility is for merchants to steer customers from using one form of payment to another, either through explicit surcharges or discounts, or through other incentives or persuasion.[12]

But the responsiveness of network demand to merchant prices does not arise entirely from merchant behavior. To the extent that high prices to merchants cause fewer merchants to accept a card, or make it costly to consumers to use it, the perceived value of the card to consumers will fall. This could in turn lead to fewer presentations of the card for payment even at merchants who continue to accept the card. Finally, effects like these could influence behavior of issuers as they try to compensate for the reduction in perceived quality and reduced usage by card holders.

Similarly, the responsiveness of demand to issuer prices does not arise entirely from issuer behavior. Issuers might be more inclined to issue a given card or to promote particular kinds of usage of a card to the extent that they receive direct payments from card transactions. These kinds of responses could in turn influence card holder perceptions of card value, card holder behavior and, ultimately, the decisions of merchants about whether to accept a given card, or whether to surcharge for its use.

The details of all of these effects on the price responses of network demand are interesting and important for many purposes. However, for purposes of evaluating the pricing problem of a payment card network monopolist, all that matters is the overall effect of each price on total demand for transactions. The network monopolist does not especially care *why* a transaction disappears following a price increase,

[12]Even if no surcharges are used, the merchant price represents a cost of doing business for the merchant, which could influence the price of goods, and thus consumer purchasing and payment patterns. Such effects are likely to be considerably smaller than those that would obtain from explicit surcharges for use of specific payment cards.

but only *whether* it disappears. As we will demonstrate, the magnitudes of the marginal demands associated with small changes in either price are critically important for monopoly pricing. But it is not important to identify exactly which direct or indirect customers of the network (*e.g.* issuers, card holders or merchants) are responding to any given price change in order to characterize the monopolist's problem. Neither is it directly important for understanding monopoly pricing of network services that both margins are influenced by the same general classes of direct and indirect customers.

Because demand depends ultimately on just two net prices that are each linear combinations of the three prices set by the network, the choice of prices by the monopolist is indeterminate without some normalization. We find it convenient to hold s_i fixed, often at zero, thus equating s with s_m and p_i with $-X$. There is no loss in generality from this approach, provided these normalizations are kept in mind when interpreting prices and price elasticities, and provided the network can freely set the remaining prices.[13]

Assuming profit maximizing behavior by the monopolist, prices must solve

$$\max_{p_m, p_i} (p_m + p_i - c)Q(p_m, p_i) \tag{2}$$

We find it more revealing to rewrite this problem in the equivalent form:

$$\max_{X, s_m} (s_m + s_i - c)Q(s_m + X, s_i - X) \tag{3}$$

Assuming an interior solution with respect to prices on both sides of

[13] We note in passing that in our model attempts to regulate a network's interchange fee directly are unlikely to have any effect on outcomes, provided the network can freely set switch fees at any level, including negative levels, in response.

the market, the necessary first-order conditions for a solution are

$$\frac{\partial Q}{\partial p_m} - \frac{\partial Q}{\partial p_i} = 0 \tag{4}$$

and

$$Q + (s - c)\frac{\partial Q}{\partial p_m} = 0. \tag{5}$$

The first of these necessary conditions demonstrates that the hypothetical monopolist will choose an interchange rate that equalizes the slopes of demand with respect to prices on both sides of the market. This balancing is exactly the condition needed to maximize the number of transactions, given any particular values for the switch fees. The monopolist's choice of interchange rates therefore is conditionally efficient in that it maximizes the value of output given the total price charged.

The first condition can be re-arranged as

$$\frac{p_m}{p_i} = \frac{\epsilon_m}{\epsilon_i} \tag{6}$$

where ϵ_m and ϵ_i denote the two elasticities of demand with respect to the prices on both sides of the market. (Note that ϵ_i is positive if p_i is negative.) Thus relative prices are directly related to relative price elasticities in a monopoly equilibrium.

The tendency of a network monopolist to choose interchange rates that balance the marginal effects on demand of prices on the two sides of the market is an immediate consequence of the fact that the network does not retain any of the interchange fee for itself. Its incentives, therefore, are essentially competitive with respect to this price, assuming that it is otherwise free to set other fees. The monopolist is interested in interchange fees only to the extent that this price influences the total amount of business done on the network. This important insight has been previously noted and discussed by, *e.g.*, Schmalensee (2002).

The remaining first order condition (5) can be written

$$\frac{s-c}{p_m} = -\frac{1}{\epsilon_m}, \tag{7}$$

which has the general form of the Lerner condition for monopoly pricing in a conventional one-sided market. This can be clarified by noting that $s - c = p_m - c_m$ where $c_m \equiv c + X$ can be interpreted as the marginal cost to the network of completing a transaction on behalf of a merchant customer, given the cost of obtaining participation in the transaction on the issuer side, which requires a net payment to the issuer of X. The fact that the interchange rate is chosen by the monopolist to maximize profits does not upset this fact. For any given level of interchange, the monopolist will choose a total price to merchants exactly as if the market were a one-sided market to merchants alone. This is a consequence of the envelope theorem.

A corollary is that $-1/\epsilon_m$ – the inverse elasticity of demand with respect to the merchant price[14] – is an appropriate measure of the ability of the network to profitably price above marginal cost, provided marginal cost is interpreted to include the cost of obtaining participation in the transaction on the issuer side. As such, the price elasticity of demand with respect to the merchant price is informative about the market power of a network monopolist on the merchant side of the market.

Given equation (4) and its corollary equation (6), equation (5) can also be written in two other equivalent forms that offer additional insights. The first of these is

$$\frac{s-c}{X} = \frac{1}{\epsilon_i}. \tag{8}$$

[14] Note that the elasticity of demand with respect to the merchant price is not necessarily the same as the elasticity of demand of merchants due to the indirect effects of a price change on card holder and issuer behavior that we noted previously.

This is a Lerner condition for monopsony pricing on the issuer side of the market – a condition that is clarified by noting that $s - c = v - X$ where $v \equiv p_m - c$ can be thought of as the value of a marginal unit of output before netting out the cost of procuring issuer participation. This equation demonstrates that the inverse elasticity of demand with respect to the issuer price is a measure of the network's ability to profitably mark down interchange below the value otherwise created by a marginal transaction. As such, it is a suitable measure of monopsony power against the issuer side of the market in this context.

A second variant on equation (5) is

$$\frac{s-c}{s} = -\frac{1}{\epsilon_s} \tag{9}$$

where $\epsilon_s = s\epsilon_m/p_m = s\epsilon_i/p_i$ is the elasticity of demand with respect to total switch fees when prices are varied subject to the restrictions of equation (4). Put differently, let

$$Q^*(s) = \max_X Q(s+X, -X). \tag{10}$$

Then ϵ_s is the elasticity of the concentrated demand function Q^* with respect to the total price s retained by the network, assuming that interchange is optimally set to maximize the number of efficient trades at any given total network price. So equation (9) is a standard Lerner condition for monopoly pricing in a one-sided market that is equivalent to our two-sided market if relative prices on the two sides of the market are always chosen to balance marginal demand effects. It follows that the inverse elasticity of the concentrated demand function Q^* is informative about the monopolist's ability to set the total price above the network's own marginal cost.

3 Relevant markets for two-sided products

We turn now to the question of defining relevant markets for purposes of antitrust enforcement. The two-sidedness of demand for payment card network services does not present any new problems of geographic market definition, so we restrict attention to product market issues here.

Our framework is that of the *Horizontal Merger Guidelines* issued jointly by the U.S. Department of Justice and the Federal Trade Commission ("the Agencies"). The *Guidelines* set out the policy of the Agencies for review of mergers between horizontal competitors. The framework of the *Guidelines* has been widely recognized by courts in the United States and elsewhere.[15] The same framework can be extended to non-merger cases to the extent that they present the same kinds of economic issues as horizontal mergers. We believe that to a significant degree, consideration of the role of two-sidedness in horizontal merger analysis will capture the essential concepts that are generally applicable to any kind of antitrust analysis involving two-sided products.

3.1 Purposes of merger review

> "The unifying theme of the Guidelines is that mergers should not be permitted to create or enhance market power or to facilitate its exercise. Market power to a seller is the ability profitably to maintain prices above competitive levels for a significant period of time. . . . In any case, the result of the exercise of market power is a transfer of wealth from buyers to sellers or a misallocation of resources."[16]

As suggested by this quote, *Guidelines* methodology is intended to

---

[15] *See* Hesse and Soven (2005).

[16] *Guidelines*, section 0.1. The *Guidelines* define market power for buyers in analogous fashion.

identify mergers that are likely to create new market power or facilitate its use. This, in turn, is motivated by a presumption that exercise of market power is harmful to consumers and inefficient.

3.2 The Hypothetical Monopolist Test

If the ultimate purpose of merger review is to determine "whether the merger is likely to create or enhance market power or to facilitate its exercise,"[17] the purpose of market definition is to facilitate an initial screen of proposed mergers based on measures of market concentration. The idea is to delineate a market within which market shares and concentration can be evaluated for purposes of this initial screen. If concentration is sufficient to create a presumption of possible anticompetitive effects from the merger then the proposed merger is subjected to further scrutiny.[18]

Market share analysis is unlikely to be informative unless the market is defined in such a way that creation of market power within the market is at least a theoretical possibility. If not even extreme concentration, *e.g.* monopoly, would result in an exercise of market power then the concentration analysis has little probative value. The Guidelines test a candidate market for antitrust 'relevance' by asking, in essence, "Is this market worth monopolizing?" If not, then subsequent market share analysis within the candidate market is unlikely to provide a meaningful screen, and a larger candidate market must be considered.

The Guidelines' market definition test itself is conceptually straightforward. For a candidate market, one must decide whether a profit-maximizing, unregulated monopolist would profitably impose at least a small but significant and non-transitory increase in price—a 'SSNIP'

[17] *Guidelines*, section 0.2.

[18] The details of the concentration analysis are not presented here, but can be found in sections 1.3, 1.4 and 1.5 of the Guidelines.

in the jargon of antitrust practitioners. Therefore, an understanding of monopoly pricing in the candidate market—our analysis in section 2, for example—typically is needed to answer this hypothetical question. The Guidelines specify that a relevant product market is the smallest group of products that satisfy the SSNIP test. Further discussion of the Guidelines methodology for determining the smallest market can be found in Werden (2002).

3.3 Applying the SSNIP test to two-sided markets

Here we consider application of the SSNIP test to the simple model of a two-sided payment network services monopoly described in section 2. To the extent that such a market fails the test, additional products must be added to the market, at which point the hypothetical monopolist is assumed to jointly determine the terms of sale for all of the products. Evaluating monopoly behavior for the multi-product firm is inherently more difficult than evaluating monopoly pricing for the single product firm. However, the complications arising from joint pricing of multiple products are conceptually no different for two-sided products than for products with a single price. So our essential points can be made considering only the single, two-sided product monopolist of the previous section.

In order to apply a SSNIP test to a two-sided candidate product market one must decide which price or prices to use. Since multiple prices are chosen simultaneously by the hypothetical monopolist of a two-sided product, an appropriate answer to this question is not immediately obvious. To make a reasonable choice, it is useful to recall that the ultimate purpose of the test is to facilitate detection of mergers that possibly could create substantial market power. So we should focus on prices for which a significant increase imposed by an independent monopolist indicates an undesirable exercise of market power.

Under this reasoning, it is not appropriate to apply the SSNIP test directly to the interchange rate X. As we have seen, since the network does not retain interchange fees, an independent, profit-maximizing monopolist will set interchange to balance demand on the two sides of the market in order to maximize output, all else equal. Assuming no distortions elsewhere in the economy, this is efficient behavior. But in any case, there is no sense in which this balancing corresponds to an exercise of market power. For a seller of network services, the Guidelines define market power as "the ability profitably to maintain prices above competitive levels for a significant period of time."[19] There does not appear to be consensus among economists on how to define a competitive level for interchange rates. Furthermore, since interchange is simply passed through from customers of network services on one side of the market to customers on the other side, an interchange rate that appears to be above a given level when viewed from one side of the market necessarily appears to be below it when viewed from the other side.

Neither is it appropriate to apply the SSNIP test separately to the net prices p_m or p_i charged separately to the two sides of the market.[20] An increase in either one of these prices could be interpreted as an exercise of market power, but only to the extent that we know something about what is happening on the other side of the market. The theory of monopoly pricing suggests that any change in price on one side of the market cannot be presumed to occur in isolation from the other side. Generally prices will change simultaneously on both sides of the market, if at all. An increase in p_m, for example, could reflect an increase in interchange or an increase in merchant switch

[19] *Guidelines* section 0.1.

[20] An exception to this general rule would be if a hypothetical monopolist would not change prices on one side of the market or the other, in which case a focus on either p_m or p_i could be appropriate. Examples of this possibility are sketched in section 4.

fees, or some combination of both. If it results primarily from an increase in interchange, then this is effectively the same as a decrease in price on the other side of the market. We are thus faced with the same ambiguities of interpretation as for interchange fees if we attempt to use net price on either side of the market in isolation as the basis for a SSNIP test.

Furthermore, if, for some reason, we were to apply the SSNIP test to the net price on each side of the market separately, then we would be faced with a logical conundrum. According to the Guidelines, a relevant product market is a collection of products, not of prices, so a product either is in a relevant market or it is not. It is not permitted to be in on one side and out on the other. Nor is a single product permitted to be both in and out of a relevant market depending on which of several equally legitimate prices we use for the analyis. For these reasons, the Guidelines logically force us to choose a single price for purposes of the SSNIP test.

In our view, the only reasonable candidate for application of a SSNIP test in our model is the total price s—the sum $p_m + p_i$ of the two prices charged to the two sides of the market. This is the most direct analog to the single price charged by a hypothetical monopolist in a conventional one-sided market. It represents the total net revenue collected by the two-sided monopolist per unit sold. In fact, for purposes of market definition, conventional monopoly pricing theory for one-sided markets can be applied without modification to a two-sided market, provided only that the concentrated demand function Q^* defined in equation (10) is used to evaluate the market demand. This is a direct consequence of equation (9), which is the classical Lerner equation for monopoly pricing of a single product in a one-sided market with demand function Q^*.

There are two additional reasons why the choice of s is appropriate for application of a SSNIP test. Consider first that any increase in s raises the monopolist's variable margin on each unit of network services sold. This is the classic motivation for the exercise of market power inherent in monopoly pricing. It is not accidental that the margin of total price over marginal cost $s-c$ appears on the left hand side of the three Lerner equations (7), (8) and (9) that we derived from the two-sided monopolist's pricing problem.

Second, given reasonable assumptions about the correspondence between output and welfare, we believe that raising the total price towards monopoly levels should be presumed to lower welfare relative to what would be achieved by setting the total price closer to marginal cost, absent evidence to the contrary. Assuming that the monopolist always sets interchange rates at a level to maximize profits, changes in s will produce movements along the concentrated demand curve Q^*. In particular, since Q^* is easily seen to be downward-sloping in s, an increase in s necessarily reduces output. To the extent that welfare is monotone in total output, at least for total prices above marginal cost, any exercise of market power with respect to the total price necessarily lowers welfare. Whether this assumption about the relationship between output and welfare is correct or not depends on what is going on elsewhere in the economy to generate the two-sided demand faced by the monopolist. But the same can be said about the relationship between output and welfare in one-sided markets. So to the extent that there is a presumption of harm from an exercise of market power in one-sided markets, an equivalent presumption should apply to two-sided markets when the total price is the vehicle for the exercise of market power.

4 Network Competition

Payment card networks, in our basic framework, compete for the business of issuing banks and merchants. Though we do not model it explicitly, they can also be said to compete indirectly for the business of cardholders. Demand on each side of the market depends on the decisions of cardholders about which cards to hold and which cards to use. For some payment card networks, such as the American Express and Discover credit card networks, the relationship between network and cardholder is direct. The network is also a card issuer and sets the prices that the cardholder faces. In the MasterCard and Visa credit card networks, as well as all debit card networks, the network interacts with the cardholder indirectly, on one side of the market through the issuing bank and on the other side of the market through the merchant or merchant acquiring bank.

One question of interest to antitrust enforcers is how a reduction in competition between payment card networks might manifest itself in pricing to one or both sides of the market. An extreme version of this question forms the basis for the hypothetical monopolist test described earlier. More generally, though, we are interested in how a reduction of competition among networks, caused for example by a merger among networks, might manifest itself in pricing.

Guthrie and Wright (2005) present a model of competition among payment networks, as do Rochet and Tirole (2003), within a general two-sided market setting. In the former case, the focus is on the effect of competition on interchange fees, and the model assumes that the total network price is equal to cost, though the ratio of the two sides might change via changes in interchange. Here we are interested in methods by which network pricing might be raised above cost. The model of Rochet and Tirole (2003) provides some robust results in a more general two-sided market setting, and the accompanying public policy insights correspond to results we discuss here. In particular, their intuition about captive customers and multihoming on one side of the market resulting in a price structure more favorable to the other side of the market are illustrated here in the course of determining which side of the market is more vulnerable to a price rise following an increase in concentration.

Here, in the context of our simple model of payment card networks, we describe some of the factors favoring exercise of market power on one side of the market or the other following an increase in concentration. We sketch two simple scenarios that illustrate conditions under which market power will be exercised on the issuer side and on the merchant side, and describe the real-world market characteristics that favor each scenario. In our third scenario, increasing concentration raises price on one side of the market but lowers it on the other. We discuss the antitrust implications of such a situation.

Scenario 1: Network competition benefits issuers but not merchants

In our first scenario we assume that each payment card carries a single network brand and that consumers each carry a single card. Consumers carrying a payment card also have an alternative method of payment (*e.g.* cash or checks). A merchant accepts a card if acceptance lowers his transaction costs relative to alternative payment methods on a set of transactions.[21] Assume a particular transaction price $\widehat{p}_m$ charged by a network equalizes costs for the merchant between the payment card and alternative forms of payment. Thus, merchants will accept a card if $p_m \leq \widehat{p}_m$, but will not accept the card if $p_m > \widehat{p}_m$.

Suppose there are n payment card brands. Demand for network j, Q_j, is a function of prices of all n firms. Let $\mathbf{p}_m$ and $\mathbf{p}_i$ represent the price vectors $\{p_m^1, \ldots, p_m^n\}$ and $\{p_i^1, \ldots, p_i^n\}$ respectively, so that

$$Q_j = Q_j(\mathbf{p}_m, \mathbf{p}_i). \tag{11}$$

Because of the structure of demand on the merchant side, however, merchant acceptance of network j does not depend on the other net-

[21] Assume too that there are no strategic effects from merchants' acceptance of cards. As has been pointed out in the literature, if a merchant's card acceptance affects its downstream market share, externalities between merchants can yield higher prices to merchants than if merchant's acceptance decisions are non-strategic. See, *e.g.*, Guthrie and Wright (2005), and Rochet and Tirole (2002).

works' pricing. An implication is that any network that is in business in equilibrium must charge a merchant price equal to $\widehat{p}_m$. Let $\widehat{\mathbf{p}}_m$ denote an n-vector whose elements all equal this threshold price for merchant acceptance. Then

$$Q_j(\mathbf{p}_m, \mathbf{p}_i) = Q_j(p_m^j, \mathbf{p}_i) = \begin{cases} Q_j(\widehat{\mathbf{p}}_m, \mathbf{p}_i) & \text{if} \quad p_m^j \leq \widehat{p}_m \\ 0 & \text{if} \quad p_m^j > \widehat{p}_m \end{cases} \tag{12}$$

where p_m^j and p_i^j represent network j's prices to merchants and issuers, respectively.

In this scenario, effectively networks do not compete with each other for the business of merchants, since one network's pricing has no effect on another network's merchant acceptance. Networks still compete, however, for the business of issuers, who choose the network brand under which to issue their card.

Assume

$$\frac{\partial Q_j}{\partial p_i^j} < 0 \text{ and } \frac{\partial Q_j}{\partial p_i^k} > 0 \text{ for } k \neq j \tag{13}$$

and that there exists a symmetric equilibrium in any symmetric n-firm network competition. Define $\varepsilon_n(p_i)$ to be the own-price elasticity of demand faced by each firm on the issuer side when all firms charge prices $\widehat{p}_m$ to merchants and p_i to issuers. If $\varepsilon_n(p_i) > \varepsilon_{n-1}(p_i)$ for all values of p_i and n then the symmetric equilibrium price to issuers decreases (i.e. the net payment to issuers rises) as the number of networks increases, while p_m remains constant at $\widehat{p}_m$. This simple, one card per customer, one network per card framework yields market power exercised on the issuer side, but not on the merchant side of the market. In this case, an exclusive focus on issuer side harm could be appropriate.

Scenario 2: Network competition benefits merchants but not issuers

Imagine alternatively that payment cards carry more than one network brand. Many PIN debit cards historically have had this feature in the United States, and have identified multiple brands, or "bugs," on the back of the card. If a card carries several network brands, the question of which network carries the transaction when a card is presented at a merchant arises. If the merchant decides the routing, assuming no quality differences in the networks, it will route over the network that offers it the lowest price of all of the networks on the card. The Department of Justice's complaint in the the merger of First Data and Concord identified merchants' power to "least-cost" route transactions over their favored networks as an important factor in the determination of network prices to merchants.[22]

Multiple bugging of cards in the United States is, in part, an artifact of the origins of PIN debit networks as regional alliances of banks. Since issuance and acceptance networks for a particular network tended to be local, in order to ensure broad coverage outside of a particular region, cards needed the ability to be run over multiple networks. In a simple model of multi-branding and merchant routing, increased market power on the network side of the market may manifest itself on the merchant side but not the issuer side.

In this scenario, assume that all cards must carry all brands of networks to be viable in cardholders' eyes. An issuer will accept any network for which the net per-transaction fee it receives, $-p_i$, at least covers its marginal cost of processing transactions. Therefore it will not issue a card if any network price is above a threshold value $\widehat{p}_i$. In what follows, we take as given that competition between issuers causes all networks to choose the threshold price $\widehat{p}_i$ in equilibrium.

Assume there is a set of "competitive" merchants that accept mul-

[22]See Verified Complaint, supra note 6.

tiple networks and route transactions over the network with the lowest merchant price. There is also a set of "noncompetitive" merchants, who each choose a single network and only accept cards enabled on that network. A network's demand is made up of two components: its share of competitive merchant transactions that it obtains via favorable routing decisions against competing networks, and its share of noncompetitive merchant transactions for which cash and checks are the only competing alternatives. Assuming networks cannot price discriminate, a network j's demand is given by

$$(14) \qquad \underbrace{L_j(p_m^j, \widehat{p}_i)}_{\text{"noncompetitive" demand}} + \underbrace{Q_j(\mathbf{p}_m, \widehat{p}_i)}_{\text{"competitive" demand}} .$$

Let total demand for network transactions be fixed as long as all networks price to merchants at or below $\widehat{p}_m$.[23] As the number of networks increases, we assume that some merchants respond by switching from the noncompetitive to the competitive group. Therefore the competitive portion of this total market increases relative to the noncompetitive portion when n increases. Let $C(n)$ be the competitive portion of the market as a function of the number of networks, and $L(n)$ be the noncompetitive portion of the market. By assumption, then, $L(n)/[C(n) + L(n)]$ is decreasing in n.

We assume that total demand from the competitive segment of the market distributes into equal shares for each network posting the best competitive price that merchants will accept. Formally, demand for network j from competitive merchants equals:

$$(15) \qquad Q_j(\mathbf{p}_m, \mathbf{p}_i) = \begin{cases} C(n)/s & \text{if} \quad p_m^j = \min\{p_m^1, \ldots, p_m^n\} \leq \widehat{p}_m \\ 0 & \text{otherwise} \end{cases}$$

where s is the number of networks pricing at $\min\{p_m^1, \ldots, p_m^n\}$. Like-

[23]We assume that if a network prices above $\widehat{p}_m$, then its share of the noncompetitive merchant transactions switch to another form of payment.

wise, each network receives an equal share of the noncompetitive demand, provided that it posts a competitive merchant price. Therefore, demand for network j from noncompetitive merchants is

$$L_j(p_m^j, \mathbf{p}_i) = \begin{cases} L(n)/n & \text{if} \quad p_m^j \leq \widehat{p}_m \\ 0 & \text{otherwise} \end{cases} . \tag{16}$$

Given an n firm symmetric equilibrium, the equilibrium merchant price for network j, p_m^{j*}, will occur when each network is indifferent between pricing $\widehat{p}_m$ to its noncompetitive merchants alone, and setting a lower price to capture also a share of the competitive merchant business. This requires

$$(\widehat{p}_m + \widehat{p}_i - c)\frac{L(n)}{n} = (p_m^{j*} + \widehat{p}_i - c)\left(\frac{C(n)}{n} + \frac{L(n)}{n}\right) \tag{17}$$

or

$$p_m^{j*} = c - \widehat{p}_i + \frac{L(n)}{C(n) + L(n)}(\widehat{p}_m + \widehat{p}_i - c). \tag{18}$$

As the number of firms n decreases, the equilibrium network price to merchants p_m^{j*} increases, while the issuer price p_i remains the same. Networks compete for a subset of merchant business through routing decisions, but do not compete directly for issuers. In this way, network market power manifests itself on the merchant side of the market but not on the issuing side of the market.

Scenario 3: Network competition helps one side of market, but hurts the other

Suppose one side of the market actually benefits from a reduction in network competition, at the expense of the other side of the market. This scenario was discussed in Section 3 in the context of the hypothetical monopolist test. If a hypothetical monopolist network is unable to profit from price increases relative to a more competitive benchmark, then the services provided by that network would not be considered a market for antitrust purposes. This observation leaves open the question of the welfare consequences of a reduction in network competition, however. Assessing the welfare effects of network pricing is beyond the scope of the simple models we present here. The model below simply sketches a scenario under which a reduction in network competition does not profit networks, or the merchant side of the market, but reduces prices on the issuer side of the market through an increase in interchange rates.

Consider a market in which consumers carry multiple cards, each carrying a single network brand so that merchants have no power to route transactions to the network with the lowest merchant price. If there were only one network, merchants presented with the corresponding card would be willing to pay up to $\widehat{p}_m$ for card services. The threshold price $\widehat{p}_m$ is a function of both the cost of alternative forms of payment and the chance that a consumer will leave the store if her payment card is not accepted. If instead multiple networks exist and consumers carry multiple cards, merchants are less willing to pay for any particular network once it has chosen to accept the others, because a rejected card may now lead to use of a substitute card rather than result in a consumer leaving the store. Assume merchants are now willing to pay $\widehat{p}_m - f(k)$ for each network, where k is the number of networks accepted. By assumption, $f(k)$ increases with k and $f(1) = 0$.

On the issuer side of the market, networks are undifferentiated, and issuers thus choose to issue on the network with the lowest (or most negative) issuer price. We assume that total demand is evenly shared among all networks matching this price. With these assumptions, demand for network j, Q_j, is given by

$$Q_j(\mathbf{p}_m, \mathbf{p}_i) = \begin{cases} A/k & \text{if} \quad p_m^j \leq \widehat{p}_m - f(k) \text{ and} \\ & \qquad p_i^j = \min\{p_i^1, \ldots, p_i^n\} \\ 0 & \qquad \text{otherwise} \end{cases} \tag{19}$$

where k is the number of firms with $p_m^j \leq \widehat{p}_m - f(k)$ and $p_i^j = \min\{p_i^1, \ldots, p_i^n\}$. Here A is a constant representing total market demand for network transactions. Assuming Bertrand competition between networks, the total price charged by network j, $p_m^j + p_i^j$ must converge to marginal cost c. The merchant price is constrained only by the merchants' marginal willingness to pay, and thus rises to $\widehat{p}_m - f(k)$ if k networks are accepted in the market. If all n networks participate in a symmetric market equilibrium, then equilibrium prices p_m^{j*} and p_i^{j*} for each network will be $\widehat{p}_m - f(n)$ and $c - \widehat{p}_m + f(n)$ respectively.

When the number of networks decreases from n to $n-1$, merchant prices increase by $f(n) - f(n-1)$ and net prices to issuers decrease by exactly the same amount. Put differently, interchange rates rise, but the total network price is unchanged, and the networks do not profit from the decrease in competition. In this case, networks effectively provide a forum for issuers to compete for merchant business. Merchants benefit from this competition, in that interchange fees and merchant prices are lower with more network competition. The opposite is true for issuers, who receive a lower net payment per transaction when there is more network competition.

If we assume that a hypothetical monopolist of network services would have some power over issuers and be able to raise net prices a significant amount above cost, then this model yields a network services market but no market power exercised in that market by any reduction in competition short of monopoly. Instead, a reduction in network services competition transfers money from merchants to issuers. The welfare consequences of such a transfer are beyond the scope of this simple model. However, to the extent there is welfare harm from such an increase in concentration, the avenue of harm is less straightforward than in many merger cases.

5 Conclusion

We have reached a number of main conclusions. First, while the two-sided nature of payment card network service products complicates life for antitrust analysis, these complications can be overcome. For example, the inverse elasticity of demand with respect to the total price is a useful measure of market power, assuming individual prices are set optimally. Second, the hypothetical monopolist test for market definition of the *Horizontal Merger Guidelines* reasonably can be applied to two-sided markets to determine if a market is large enough that there is some potential to exercise market power. The key is to apply the SSNIP test to the total price charged by the network monopolist while letting relative prices on the two sides to the market adjust optimally via the interchange rate. Usually it is not appropriate to simply look at the ability of the monopolist to raise price on one side of the market or the other. Third, through some examples we have shown that increasing concentration may produce increased prices on either side of the market with little effect on the other, or even with a prices moving in opposite directions on the two sides of the market. In light of these possibilities, antitrust practitioners must consider the structural features of each side of the market in order to determine how market power is likely to manifest itself.

References

EFTPOS Industry Working Group (2002). "Options for EFTPOS Interchange Fee Reform." Discussion paper, Australian EFTPOS Industry Working Group.

Evans, D. and R. Schmalensee (2005). "The Economics of Interchange Fees and Their Regulation: An Overview." MIT Sloan Working Paper 4548-05, MIT Sloan School of Management.

Evans, D. S. (2002). "The Antitrust Economics of Two-sided

Markets." Related Publication 02-13, AEI-Brookings Joint Center for Regulatory Studies.

Guthrie, G. and J. Wright (2005). "Competing Payment Schemes."

Hesse, R. B. and J. H. Soven (2005). "Defining Relevant Markets in Electronic Payment Network Cases." *Antitrust Law Journal*. (forthcoming).

Rochet, J.-C. (2003). "The Theory of Interchange Fees: A Synthesis of Recent Contributions." *Review of Network Economics*, 2: 97–124.

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

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

Schmalensee, R. (2002). "Payment Systems and Interchange Fees." *Journal of Industrial Economics*, 50: 103–122.

U.S. Department of Justice and Federal Trade Commission (1997). *Horizontal Merger Guidelines*.

Weiner, S. E. and J. Wright (2005). "Interchange Fees in Various Countries: Developments and Determinants."

Werden, G. J. (2002). "Market Delineation Algorithms Based on the Hypothetical Monopolist Paradigm." Economic Analysis Group Discussion paper EAG 02-8, U.S. Department of Justice, Antitrust Division.