What Do We Know About Interchange Fees and What Does it Mean for Public Policy?

Commentary on Evans and Schmalensee

Michael L. Katz

I. Introduction

In their paper, David Evans and Richard Schmalensee provide a broad overview of the current state of our theoretical and empirical knowledge regarding the efficiency effects of interchange fees.[1] They also draw broad conclusions regarding the implications of this state of knowledge for public policymaking. Evans and Schmalensee assert that:

> There is no apparent basis in today's economics—at a theoretical or empirical level—for concluding that it is generally possible to improve social welfare by a noticeable reduction in privately set interchange fees. Thus, if antitrust or other regulators had to show that such intervention would improve welfare, they could not do so.[2]

I agree with several aspects of Evans and Schmalensee's characterization of our current state of knowledge. But I disagree with their bottom line for policymaking. In part, this disagreement arises because I think they are addressing the wrong question. To my mind, the central question is not whether one can conclude it is generally possible to improve social welfare by reducing interchange rates. Instead, I think the central issue is whether the circumstances of a specific market under consideration are such that governmental intervention can be expected to improve economic welfare.[3] Even with respect to this narrower question, I think that Evans and Schmalensee are too pessimistic about our ability to understand the economics of network fee setting. The best way I know to explain our differences in this regard is to offer my own overview of the state of our theoretical and empirical knowledge, and to offer my own view of the implications for policymaking.

The remainder of this comment is organized as follows. The next two sections provide a foundation for the later analysis. Section II lays out an economic approach to policy analysis. Section III argues that the focus on interchange fees can be misleading in some circumstances. In these settings, a better approach would be to focus on network fees more broadly. Section IV provides a very brief survey of the theoretical literature regarding network fees more generally, and Section V addresses the current state of empirical knowledge. Section VI then argues that the informational demands for policymaking may not be as great as they might first appear. A brief conclusion closes.

II. A FRAMEWORK FOR POLICY ANALYSIS

There is a consensus among economists that certain public policies, such as the definition and enforcement of property rights, are fundamental to the existence of well-functioning markets. There is less of a consensus regarding the extent to which public policy should go further and intervene to limit private-sector price setting or impose other forms of economic regulation. Interestingly, there is wide agreement on the appropriate analytical framework in which to address questions of the appropriate degree of intervention. Evans and Schmalensee summarize this consensus view in terms of two questions:

- "Is the performance of the market or markets being considered substantially suboptimal?"
- "Is there a practical regulatory policy that is reasonably certain to improve market performance substantially?"[4]

One can debate some of the fine points of these two questions (for example, the second question might better be stated in terms of expected benefits and costs rather than terms such as "substantially" and "reasonably certain"), but they do capture the central points.

An economic analysis of regulation's effects weighs the potential costs and benefits of regulation. The latter depend, in part, on how well markets function in the absence of regulation. There is broad agreement among economists that market outcomes generally possess desirable characteristics, but that unregulated markets can lead to inefficient or unfair outcomes in certain circumstances. To the extent that a market outcome is inefficient or unfair, there may be scope for regulation to improve the outcome.[5]

Poor market performance alone, however, does not imply that regulation will improve matters. One must consider whether intervention can address the problem identified. One must also consider the costs of regulation,

both administrative costs and the costs of unintended consequences. The latter can be very significant and typically arise because policymakers have limited information on which to act. Consequently, the state of knowledge has far-reaching implications for policy analysis.

III. INTERCHANGE, SWITCH FEES, OR ALL OF THE ABOVE?

Interchange fees are the focus of much of the policy debate regarding payment networks. In this section, however, I will argue that this focus sometimes runs a danger of confusing the issues.

An interchange fee is a monetary amount that flows between the financial institution serving a merchant at which a payment card has been used and the financial institution that issued that card. A distinguishing feature is that this fee is not retained by the payment card network. Rather, it is passed through from one participating financial institution to the other. With only one exception of which I am aware, debit and credit card systems have interchange fees that flow from the merchant acquirer to the card issuer.[6]

To see why an excessive focus on interchange fees can be misleading in some circumstances, it is helpful to introduce a few pieces of notation. Let t denote the interchange fee, measured as the flow of money from the acquirer to the issuer. Let w_A denote the (switch) fee that the payment network charges an acquirer when the network processes a transaction involving that institution. And let w_I denote the (switch) fee that the payment network charges an issuer when the network processes a transaction involving that institution. In terms of this notation, the net price paid by an acquirer for a transaction over this network is $w_A + t$, and the net price paid by an issuer for a transaction is $w_I - t$.[7] Observe that these are the prices of interest for purposes of understanding acquirers' and issuers' economic incentives and the resulting efficiency effects. Fundamental economic logic indicates that the level of transactions, the prices charged by issuers and acquirers, and network profits all can be expressed as functions of the *net* prices to acquirers and issuers, ceteris paribus.

In general institutional settings, knowing the value of t alone tells one nothing about the net prices acquirers and issuers face for using a network. The following table shows why. As illustrated by the first two rows, identical interchange fees can be associated with very different net prices. The reason is that the network's switch fees may attenuate or strengthen the flows associated with interchange. The first and third rows further illustrate this general point by showing how different values of the interchange fee can give rise to identical net acquirer and issuer fees.

w_A	t	w_I	Net price to acquirers	Net price to issuers
0.5%	1.0%	0.5%	1.5%	-0.5%
1.0%	1.0%	0.0%	2.0%	-1.0%
0.0%	1.5%	1.0%	1.5%	-0.5%

The following irrelevance result provides still another way of seeing that an overemphasis on interchange can be misleading in some circumstances. For any value of t, the network can set w_A and w_I to obtain any pair of net prices desired, and the value of t has no effect on the resulting network revenue per transaction.[8] It follows that if the network is free to choose its issuer and acquirer fees, then the level of the interchange fee is irrelevant.

Of course, in certain circumstances, an interchange fee is a well-defined concept. Suppose, for example, there is no central network authority collecting fees (for example, the Australian EFTPOS system), so that both w_A and w_I equal 0.[9] Then the value of t uniquely determines the prices faced by acquirers and issuers. Similarly, if a centralized network prices its services to break even and sets w_A and w_I at "low" levels, one can focus on t and largely ignore w_A and w_I without doing too much violence to the welfare and efficiency analysis.[10] But it should be kept in mind that when one is considering a network for which w_A and w_I are subject to choice and are set at non-negligible levels, they can no longer safely be ignored.

The point made in this section is more than one of terminology. First, it shows that by strategically setting its transactions fees, a network might evade regulations that focused solely on interchange fees. Second, efficiency issues concerning issuer and acquirer fees, both in absolute terms and relative to one another, can arise even in the absence of interchange fees (for example, with a network, such as American Express, that does not have formal interchange fees even when it has issuer partners). Third, because some payment networks do not have formal interchange fees, regulation that focuses on interchange fees may apply to some networks but not all.

The effects of asymmetric regulation merit particular attention when that regulation concerns the price *structure* (for example, the ratio of the acquirer and issuer prices) rather than absolute levels. Consider a "standard" market in which a regulatory body imposes a binding ceiling on the prices of only one of several competing suppliers. When the regulated supplier is forced to lower its prices, competing suppliers may be forced to match

in order to retain customers. In this case, the asymmetrically applied regulation may come close to having symmetric effects on all of the suppliers. In contrast, when regulation imposes a price structure, competing suppliers may not be forced to match. Instead, they might well continue to offer price structures that more successfully attract usage of their networks. Hence, the market outcome resulting from asymmetric regulation of price structures may be itself asymmetric, and there may be greater potential for distortions in competition and a resulting loss of efficiency.[11] Of course, it does not follow that all asymmetric regulation is inefficient or otherwise undesirable. There may be sound reasons for treating different networks differently (for instance, significant differences in the degree of market power), but it is useful to keep in mind the possibility of distorting network competition.

IV. A SHORT REVIEW OF A LONG LIST OF THEORETICAL PAPERS

With this background, now consider what economic theory has to say about the setting of interchange fees, paying particular attention to the comparison of privately and socially optimal rates. In recent years, there has been an explosion of work on these issues. In this section, I will offer a highly selective overview that is just comprehensive enough to identify the points central to the policy debate. I will then draw out initial implications of these findings.

A. *Key findings*

The following five findings from the theoretical literature lie at the heart of the policy debate.[12]

> *1. Changes in the net prices to acquirers and issuers will be passed through to consumers in ways that generally affect transactions volumes.*

At first glance, an interchange fee is solely a payment that goes between a card-issuing institution and a merchant-acquiring one. However, holding the network switch fees constant, the level of interchange affects acquirers' costs and thus can be expected to affect merchant service fees. Merchant service fees, in turn, may affect merchants' card-acceptance decisions and the prices merchants charge their customers. Moreover, again holding other network fees constant, the interchange fee affects an issuer's incentives for card issuance, including incentives with regard to both pricing and promotional strategies aimed at cardholders. Economic theory thus suggests that the level of an interchange fee (or the levels of network fees more generally) can have significant effects on card use and resulting levels of economic welfare and efficiency.[13]

There are sometimes attempts to assert that changes in interchange fees will not predictably affect the pricing and promotional incentives of issuers.[14] Such claims, however, run counter to standard economic theory (as well as raise questions of why networks have interchange fees flowing to issuers).[15] A change in incremental costs will affect a supplier's profit-maximizing output level, regardless of whether the firm is a perfect competitor, monopolist, or something in between.[16]

There is a theoretical result about the effects of interchange fees on other prices that has attracted considerable attention because it leads to a second interchange irrelevance result: If merchants can and do costlessly surcharge and issuers can and do costlessly adjust their transactions prices (including rebates) to cardholders, then the level of the interchange fee is irrelevant.[17] Intuitively, each side of the market simply passes the effects of any interchange fee through to the cardholder/customer; an increase in the interchange fee raises the rebate for card use, which encourages card use, but it also increases the surcharge for card use, which exactly offsets the increased rebate.

2. The network prices that maximize total surplus depend on demand conditions.

Consumption of payment services generally involves two sides of the transaction—a consumer and a merchant—each of whom takes actions, receives benefits, and bears costs. Moreover, the decisions made by one side of the transaction generally affect the economic well-being of the other. The fact that each side of a transaction needs to take actions in order for the transaction to take place raises the possibility each side will make privately optimal, but socially inefficient, decisions because each will fail to take into account effects on the economic welfare of the other party to the transaction. The prices that acquirers and issuers charge for card holding and transactions play a role in determining the incentives for consumers to hold and use cards and for merchants to accept them. Network pricing matters in all of this because it affects the resulting prices charged by acquirers and issuers.

Both commercially and socially optimal network pricing entail making tradeoffs in terms of the effects that various price structures will have on consumer and merchant holding, use, and acceptance decisions. An important lesson of this analysis is that, in general, making the socially optimal tradeoff depends on the benefits realized by each side of the transaction (also known as the demand conditions), as well as the costs of transactions. Consequently, purely cost-based rules (for instance, "set the price charged to the cost-causer at marginal cost") are unlikely to attain full efficiency.

The dependence of efficient prices on demand conditions is not as unusual as some people appear to believe. It also arises, for example, when marginal cost is not constant. Specifically, the rule that price should equal marginal cost does not determine uniquely the price if marginal cost changes with the volume of production. In this case, demand conditions are relevant to efficient pricing because demand will influence the scale of production. Similarly, the theory of multiproduct Ramsey pricing—which characterizes efficient pricing subject to the constraint that suppliers earn non-negative profits—finds that efficient pricing depends on demand conditions when average costs are greater than marginal costs, so that marginal-cost pricing would lead to supplier losses.

3. Even with constant unit costs, the prices that maximize total surplus generally are below cost.

In a non-two-sided market, a rule that sets price equal to marginal cost does define the unique efficient price when there are constant unit costs (and certain other conditions are met as well).[18] In a two-sided market, however, the prices that maximize total surplus generally are below cost even when there are constant unit costs.[19] Intuitively, each side's decision to engage in a transaction generates benefits for the other side, but a self-interested decision maker will ignore these effects when he or she decides whether the benefits of a transaction are greater than the price he or she faces.[20] Thus, each side should face prices below marginal cost in order to help internalize these effects.

Because merchants and cardholders interact with a payment network through acquirers and issuers, a second pricing consideration arises. When acquirers and/or issuers are imperfectly competitive and individual suppliers have some degree of market power, they will tend to set their prices above their private marginal costs. Fully efficient network prices would account for these markups. Suppose, for example, that acquirers are perfectly competitive but that issuers exercise market power and set prices above their marginal costs.[21] Then, in theory, it can be optimal to lower the price charged to issuers and raise the price charged to acquirers to offset the differential markups.[22] In addition to considering acquirer and issuer markups, one also could consider the degree to which merchants mark their prices above their costs.

It is worth observing that this situation, while unusual, is far from unique. First, consider economic consequences that cut across individuals. The socially optimal price of vaccinations, for example, is less than marginal cost because getting a vaccination confers benefits on other, unvaccinated people,

who are then less likely to be exposed to the disease. Next, consider the effects of imperfect competition. Similar effects arise whenever one is considering the price of input provided to a producer with significant market power or one is considering a public policy of subsidizing such producers.

> *4. The network prices that maximize network profits generally diverge from total-surplus-maximizing network prices, but may do so in a variety of ways.*

Now consider the choice of network fees by a profit-maximizing network. It is useful to decompose potential distortions into two components: the network's margin and the balance of its charges between acquirers and issuers.[23] A first observation is that a profit-maximizing monopoly network sets its margin higher than is efficient. The reason for this distortion is essentially the same as the standard monopoly output restriction. Second, taking its margin as given, a profit-maximizing monopoly network sets the issuer and acquirer charges to maximize transactions volume. This behavior can be a second source of distortion because the network owner does not take into account the fact that different transactions can have different net social benefits associated with them. In theory, the resulting distortions in the relative prices charged to issuers and acquirers can run in either direction.

Equilibrium under network competition can also entail pricing distortions. Specifically, suppliers may respond to one side of the market (for example, lower the net price to that side of the network to attract that side's patronage) to an inefficient degree. This is similar to a situation that can arise under Ramsey pricing: Competing suppliers set their prices in response to firm-specific elasticities, but socially optimal prices may be functions of market elasticities.[24]

Although there are theoretical arguments running in both directions, there is a powerful intuition suggesting that relative network prices will be distorted toward high acquirer fees because of the nature of merchant incentives. There are two elements. First, so-called "merchant resistance" to high merchant service fees will be too low due to externalities across merchants.[25] Specifically, by accepting particular payment cards, a merchant may increase its sales but do so at the expense of rival merchants. Thus, the collective benefits of a merchant's accepting payment cards may be much lower than the merchant's individual benefits. Consequently, there are situations in which a set of merchants will find it individually rational to accept a payment card even though the merchants would collectively be better off if none did.

The lack of sufficient merchant resistance is reinforced to the extent that merchants tend to accept the cards of multiple payment networks and afford their customers the choice of which payment mechanism to use without taking significant efforts to steer consumer choice. In this setting, a network will have incentives to set fees that are relatively attractive to issuers in order to induce them to take actions that promote consumers' use of that network's cards.

> *5. The network prices that maximize a weighted average of acquirer and issuer profits generally diverge from the total-surplus-maximizing prices.*

MasterCard and Visa manifestly do not set their interchange and network fees to maximize network profits. Several papers model the card associations as maximizing a weighted combination of issuer and acquirer profits. These models focus on interchange fees and implicitly assume that w_A and w_I are fixed at negligible levels.

These analyses find that the privately set interchange rates depend, *inter alia*, on: (a) the relative weights assigned to issuers and acquirer and (b) the relative degrees to which issuers and acquirers pass through changes in interchange to cardholders and merchants, respectively. For example, if issuers retain some of the interchange fee so that their profits increase with the fee (holding fixed the level of transactions) but acquirers are sufficiently competitive that they earn zero economic profits regardless of the interchange fee, then a network that puts any weight on issuer profits will set the interchange rate to maximize issuer profits. Analysis of the general case demonstrates that privately set interchange fees can, in theory, be above or below the social optimum, depending on the specific weights and degrees of pass-through.[26]

The distortions due to the nature of network competition and merchant competition discussed under Finding 4 above can also arise in the setting under consideration here. There is also a consumer externality that can come into play in addition to the merchant externality identified above. In the model used by Rochet and Tirole (2002), competition between merchants results in their equilibrium profits being independent of the interchange fee level. Nonetheless, high interchange rates can give rise to welfare losses by inducing excessive payment card use. These adverse welfare effects arise because of a pecuniary externality among consumers. Anticipating that some consumers will use expensive payment instruments, merchants will raise their prices, harming all consumers, both card users

and non-card users. However, an individual consumer making a card-use decision does not take these effects into account. Hence, high interchange fees can result in a transfer of surplus from consumers to issuers through an inefficient mechanism: excessive card use.

B. How far does theory get us?

Findings 3, 4, and 5 indicate that the market outcome is almost certainly inefficient. Finding 1 indicates that these inefficiencies will flow through to the prices faced by merchants and consumers. But Finding 2 and the analyses that underlie it indicate that determining the total-surplus-maximizing prices can be a complex and difficult task.[27] Moreover, Finding 3 on the efficiency of pricing below cost indicates that policymakers may have to operate in a second-best world in which they will likely not be able to induce the first-best outcome because network suppliers might otherwise cease operation.

Where theory leaves us is with concern about market performance, but also with concern about the possibility of unintended, adverse consequences from otherwise well-intentioned regulation. Theory alone is not going to answer the question of whether public policy intervention in general, or a specific form of intervention in particular, will improve welfare. We need to look at the facts.

This conclusion should not be surprising. As Evans and Schmalensee point out, there are many theoretical models of imperfect competition in which suppliers' equilibrium choices may deviate from the welfare-optimal ones and may do so in ambiguous ways. Examples include entry in the presence of economies of scale, bundling, price discrimination, and research and development.[28] Indeed, I sometimes joke that the only general theorem in the economics of industrial organization is that there is no other general theorem. Stated another way, people sometimes object to oligopoly theory on the grounds that anything can happen. This claim is an overstatement, but it contains a grain of truth. But it also does not mean that the situation is hopeless.

Among the most important questions for public policy are whether all theories are equally plausible and whether empirical analysis can be used to resolve the ambiguities of various models. Consider for the moment an example unrelated to payment networks. Specifically, consider the public policy treatment of a price-fixing cartel comprising a number of chemical producers. One can write down theoretical models in which the cartel outcome is more efficient than the noncollusive outcome.[29] But we do not

allow this possibility to paralyze public policy toward price fixing. Indeed, the efficiency-enhancing effects are considered so unlikely that price fixing of this sort is per se illegal in the United States.

Turning to an example closer to home, Evans and Schmalensee are correct that one can construct a logically coherent theory under which a reduction in the rate at which consumers earn rewards for using their credit cards will encourage greater card use (this is known in other contexts as a backward-bending supply curve).[30] However, I would argue that until someone provides empirical evidence of such an effect (and I am not aware of any), it is reasonable to assume that consumers respond to decreased rewards by reducing their use of credit cards.

V. THE BROAD STATE OF EMPIRICAL KNOWLEDGE

Facts clearly are needed. Ideally, various theories of network fee setting would give rise to empirically testable hypotheses, and those hypotheses would be tested against data. And, ideally, we would have a series of comprehensive, rigorous studies of cost and demand conditions for alternative payment mechanisms that would be based on generally accepted principles and methodologies.

At present, reality falls short of this ideal. Although various theories have been put forth about how interchange rates (or the prices to the two sides of various payment systems, more generally) are privately set, these theories have not been subjected to rigorous empirical tests. Similarly, although there are many different studies of the costs of various payment instruments, there is a lack of consistency in terms of the definitions and methodologies used. Turning to demand studies, data collection efforts are made difficult by the personalized nature of prices that people face. Some consumers enjoy interest-free periods on their credit card accounts, while others (for example, those carrying a balance) do not. Rewards programs can have pricing elements that are sensitive to purchase volumes in a nonlinear way. Similarly, debit card accounts may allow a certain number of free transactions per month before a transaction fee takes effect.

Even though the state of empirical knowledge is far from ideal, it would be a mistake to conclude that we do not know anything about payment instruments. In fact, there is a lot of information available about costs, demands, and fee setting. It would be desirable to get more information, but a major task before public authorities is to pull together what exists today and to make use of it through triangulation, approximation, and a healthy dose of caution.

In some cases, a few facts may be enough to resolve theoretical ambiguities. Models that address the consequences of banning no-surcharge rules provide one example. Several of these models have focused on the limiting case of frictionless surcharging and the resulting irrelevance of interchange fees. Understanding this limiting case is useful in understanding the economic forces at work more generally. But there is little reason to believe that, in practice, interchange fees will become meaningless if merchant surcharging is allowed. This is so in part because, although there is debate regarding the exact extent of surcharging, the actual degree of surcharging has been limited in those countries where it is allowed.[31] Here, one does not need a deep investigation of the facts to rule out an extreme theoretical possibility.

Consider a second example. As discussed in the previous section, economic theory indicates that interchange fee setting by a card association will depend on the weights the association's objective function assigns to issuer and acquirer profits, as well as the relative degrees of issuer and acquirer pass-through. Industry wisdom is that most of the weight is put on issuer profits, at least in the United States, and that acquirers generally pass through a higher percentage of fee changes to their customers than do issuers. Both of these factors (in addition to others discussed above) suggest that a privately set interchange rate will be above the efficient level, not below.

VI. Decentralization and economizing on information

Almost any public policy decision must be made under conditions of uncertainty and limited information. Policymakers almost never have all of the information that might be relevant to their decisions because collecting complete information generally is too costly and takes too long. In response to the lack of complete information, policymakers implicitly or explicitly must form beliefs about the probabilities of various possible outcomes, which then are combined to develop projections of average likely net benefits.

More information would be useful, but one should not overstate the needs. One of the great virtues of the market economy is the way in which it economizes on information costs. In some circumstances, regulation can make use of market forces to economize on information as well.

The choice between private automobiles and public transportation illustrates how the market economy economizes on information and how public policy can take advantage of this virtue. Suppose that policymakers wish to promote efficient use of public transportation by pricing it appropriate-

ly. It might seem that policymakers would need to know a lot about consumer demands for various modes of transportation. But as long as consumers face relative prices that reflect underlying social costs, the choice among transportation modes can efficiently be decentralized. A rational consumer will choose the transportation mode that maximizes the surplus of the consumer's benefits over the costs that he or she faces. When the consumer faces prices that reflect social costs, he or she will thus choose the transportation mode that maximizes the surplus of benefits over social costs.

It may be possible to apply a similar approach to the choice among payment mechanisms. There is, however, a potentially important difference between choice among payment mechanisms and the automobile/public transportation decision. Payment networks are examples of platforms or two-sided markets. That is, they operate as intermediaries. As discussed above, that means that—in theory, at least—one needs to consider demand on both sides of the network at once.

There is an argument made by some analysts, however, that implies that "mature" payment networks might reasonably be treated as one-sided platforms at the margin. In this context, "mature" refers to a situation in which the payment network is so well established that most merchants feel the need to accept it and would be highly insensitive to small changes in merchant service fees. In this case, interchange fees have the greatest impacts through their effects on consumer behavior. For example, Rochet has posited that "[when] the choice of payment instrument is ultimately a decision of the buyer, that impacts the net costs of the seller."[32] Thus, to a first-order approximation, one might be able to treat mature payments systems as being one-sided markets, at least for certain changes. When a consumer faces prices that reflect the underlying social costs of the alternative payment mechanisms, the consumer can be expected to choose the payment mechanism that maximizes the surplus of benefits over costs.[33] There is no need for policymakers to have detailed knowledge of the consumer's benefits and resulting demand functions.

VII. CONCLUSION

The title of the conference asks, "What role for public authorities?" The paper by Evans and Schmalensee and the discussion above identify a very helpful answer: Continue collecting information, including data on the costs of alternative payment mechanisms, the nature of merchant acceptance decisions, and the market structures for issuing and acquiring.

Author's note: The author would like to thank Joseph Farrell for helpful comments on an earlier draft.

ENDNOTES

[1] Evans and Schmalensee (2005).

[2] Evans and Schmalensee (2005, p. 79).

[3] Indeed, as discussed below, Evans and Schmalensee themselves state the issue in similar terms.

[4] Evans and Schmalensee (2005, p. 97).

[5] A common response by economists is to call for a division of labor in policies under which economy-wide tax schemes are used to promote fairness (that is to say, redistribute income in accordance with the objectives of the relevant policymakers) and market-specific regulatory intervention is limited to correcting efficiency problems (that is to say, market failures) or ameliorating their effects. In practice, however, many market-specific policies are implemented in part with the objective of redistributing income.

[6] The one exception is the Australian PIN debit (EFTPOS) network, in which the interchange payment flows from the card-issuing institution to the merchant-acquiring one.

[7] Interchange fees often are levied as percentages of the transaction value.

[8] That is, given interchange level t and desired net prices α and β, the network can always solve for $w_A+t=\alpha$ and $w_I-t=\beta$ earn revenues of $\alpha+\beta$ per transaction.

[9] Alternatively, w_A and w_I can be interpreted as the acquirers' and issuers' respective costs of interconnecting and communicating with one another. For many purposes, these costs can be taken as exogenously given.

[10] This is, in effect, what Rochet and Tirole (2002), among others, do.

[11] A similar situation arises when nominal prices are regulated, but quality levels are not. In this case, a regulated firm may be forced to pursue a low-price, low-quality strategy while unregulated rivals are free to compete using a high-price, high-quality strategy if that better attracts profitable patronage.

[12] Summaries of significant papers in the earlier literature are provided in Katz (2001) and Rochet (2003).

[13] For similar reasons, economic theory suggests that the levels of various network fees can raise fairness issues. For the sake of brevity, I will not address these issues further.

[14] For example, Evans and Schmalensee (2005, p. 103) focus on pricing and write that: "It would appear to be generally the case that the interchange fee is a highly imprecise instrument for affecting the volume of transactions on cards and thus for correcting any perceived market distortion. That is because there is only a loose connection between interchange fees and transactions prices to cardholders."

[15] For a brief discussion of empirical findings in this area, see Weiner and Wright (2005).

[16] One exception is certain theories of oligopoly that predict rigid prices. Under such a theory, a supplier might not change its price in response to a cost change out of concern that doing so would destabilize the current equilibrium in some undesirable way (for example, trigger a price war).

[17] A general version of this result is proved by Gans and King (2003). Two additional points are worth observing. First, surcharges can themselves internalize effects across merchants and consumers and, thus, play a role ascribed to interchange fees. (See Katz (2001) and references therein.) Second, monetary payments directly between issuers and merchants can have effects similar to those of interchange, again highlighting the value of taking a broad view of fees when conducting efficiency analyses.

[18] With constant unit costs, marginal cost is equal to average cost and both have a unique value. Thus, setting price equal to marginal cost determines a value of price that does not depend on the level of production and marginal-cost pricing will cover average costs. The additional necessary conditions include an absence of externalities and a lack of distortions in the rest of the economy.

[19] A more technical argument is needed to establish the precise result. See Evans and Schmalensee (2005) and the references therein. See also Hermalin and Katz (2004), which establishes this result in a different institutional context entailing direct sales to the network's end users.

[20] Rochet (2003) refers to this effect as the "fundamental externality" of payment networks.

[21] Rochet and Tirole (2002) explore this case in the context of interchange fee setting.

[22] Expanding on Katz (2001), Vickers (2005, p. 239) argues that these considerations provide an "unappealing" basis for setting interchange rates, and that policies aimed at promoting issuer (or acquirer or merchant) competition would be preferable.

[23] Hermalin and Katz (2004) provide an analysis along these lines.

[24] See, for example, Rochet and Tirole (2003).

[25] Rochet and Tirole (2002) provide an elegant formal model of this phenomenon.

[26] See Wright (2004) and references therein.

[27] Finding 2 indicates that the efficiency justification for cost-based pricing regulation cannot be that it is sufficient to achieve a fully optimal outcome. Instead, it must be that either markets can be treated as if they are one-sided at the margin (see the discussion below) or cost-based pricing is taken as an approximation that serves as the basis for making a directional improvement. As long as the resulting cost-based prices are tested against what is known about demand conditions, this type of cost-based pricing can be a rational policy response to information limitations.

[28] Evans and Schmalensee (2005, p. 96).

[29] For example, if there are economies of scale, it could (in theory) be more efficient to have a cartel allocate production to just one supplier and have the firms share the resulting monopoly profits. Alternatively, if there are other distortions in the economy, the cartel's output restriction might (in theory) offset those distortions in a beneficial way.

[30] Evans and Schmalensee (2005, p. 112).

[31] Of course, even if there is a low level of surcharging, one should not conclude that the possibility of surcharging is irrelevant. First, there is heterogeneity across merchant sectors, so that surcharging may be important in some sectors even if it is used by a small percentage of merchants overall. Second, the *possibility* of surcharging may affect equilibrium network behavior (there are some indications that this may be happening in Australia with respect to American Express and Diners Club). An analogy may be useful. One should not measure the success of traffic fines by the amount of money collected when someone runs a red light. A fine that completely deterred the running of red lights might be optimal even though it collected no revenue. Similarly, the effects of possible surcharges should not be measured solely by the degree of actual surcharging.

[32] Rochet (2003, p. 98).

[33] If merchants try to influence consumers through various financial incentives, those incentives should be factored into the calculation of the prices that consumers face.

REFERENCES

Evans, David S., and Richard Schmalensee. 2005. "The Economics of Interchange Fees and Their Regulation: An Overview," in proceedings of a conference on *Interchange Fees in Credit and Debit Card Industries: What role for Public Authorities?* (Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6, 2005).

Gans, Joshua, and Stephen P. King. 2003. "The Neutrality of Interchange Fees in Payment Systems," *Topics in Economic Analysis and Policy*, vol. 3, no. 1.

Hermalin, Benjamin E., and Michael L. Katz. 2004. "Sender or Reciever: Who Should Pay to Exchange an Electronic Message?" *RAND Journal of Economics*, vol. 35, no. 3, pp. 427-77.

Katz, Michael. 2001. "Reform of Credit Card Schemes in Australia II, Commissioned Report: Network Effects, Interchange Fees, and No-surcharge Rules in the Australian Credit and Charge Card Industry," Reserve Bank of Australia, August.

Rochet, Jean-Charles. 2003. "The Theory of Interchange Fees: A Synthesis of Recent Contributions," *Review of Network Economics*, vol. 2, no. 2 (June), pp. 97-124.

Rochet, Jean-Charles and Jean Tirole. 2002. "Cooperation among Competitors: Some Economics of Payment Card Associations," *RAND Journal of Economics*, vol. 1, no. 4, pp. 990-1029.

Rochet, Jean-Charles and Jean Tirole. (2003) "Platform Competition in Two-sided Markets," *Journal of the European Economic Association*, vol. 1, no. 4 (June), pp. 990-1029.

Vickers, John. 2005. "Public Policy and the Invisible Price: Competition Law, Regulation, and Interchange Fee," in proceedings of a conference on *Interchange Fees in the Credit and Debit Card Industries: What Role for Public Authorities?* (Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6, 2005).

Weiner, Stuart. E., and Julian Wright. 2005. "Interchange Fees in Various Countries: Developments and Determinants," in proceedings of a conference on *Interchange Fees in the Credit and Debit Card Industries: What Role for Public Authorities?* (Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6, 2005).

Wright, Julian. 2004. "The Determinants of Optimal Interchange Fees in Payment Systems," *Journal of Industrial Economics*, vol. 52. no. 1 (March), pp.1-26.