Langlet, Markus

Doctoral Thesis

Payment Card Network Pricing - A Theoretical Approach Analyzing the Relationship between Downstream Market Characteristics and the Merchant Usage Fee

Payment Card Network Pricing

A Theoretical Approach Analyzing the Relationship between Downstream Market Characteristics and the Merchant Usage Fee

Dissertation

Zur Erlangung des akademischen Grades

doctor rerum politicarum

eingereicht an der EBS Universität für Wirtschaft und Recht

Wiesbaden

von Dipl.-Ing. Markus Langlet

geboren am 7.2.1974 in München

Gutachter:

1. Prof. Dominique Demougin, Ph.D.

2. Prof. Dr. Jenny Kragl

Acknowledgements

The present thesis would not exist without the support of numerous people.

First of all, I owe a huge thank you to Dominique Demougin, who deserves nothing less than the German title 'doctor-father'. Thank you for your excellent supervision, wise guidance, and continuous encouragement, which was more than crucial for finishing this thesis. I am also very thankful for the constructive and fruitful cooperation of my co-authors in the second and third essay, Jens Uhlenbrock and Clemens Buchen. You have both proven to be profound in your analysis and reliable in our co-workmanship. In addition, a big 'thank you' to Jenny Kragl for your thoughtful observation and constructive feedback on my research work and Gabriella Zopcsak for your friendly support. Thank you all for adopting me at the Department of Governance and Economics at the EBS Universtät.

Along the way, numerous comments and suggestions were helpful in compiling this work, contributed by Robert Göhlich, Andre Schmidt, Ralf Fendel, Anja Schöttner, Abe Aburachis, Atul Gupta, Francesco Ciampi, Sean Phillipp, Michael Noel and the participants of the Global Conference of Business and Economics in 2008, 2009, and 2010. Thanks also to DrPam and GrammarQueen, two anonymous editors, furthermore, to Harald Stritzelberger for your up-spirit, pep-talk, and continuous encouragement.

I am deeply grateful to my parents, Hilmar & Brigitte Langlet, for raising me. Words cannot capture your portion–surely not only in my academic and professional career. In addition, I am profoundly indebted to my mother-in-law, Martha Weber, for your care, council, and unresting support–not only regarding this thesis. My American 'parents', Ron & Karen Willis, chipped in with never ending encouragement and personal examples (not only regarding the language). Thanks to Alain Etouem, for carrying your share as a friend and to Martin & Johanna Ücker for your inspiration.

Furthermore, I am most profoundly grateful to my wife Christiane, for your unfailing love, patience and encouragement, for guarding my back, and for your continuous support; furthermore, for believing in me during this entire time.

Last but not least, He is the wind beneath my wings!

Contents

List of Figures

List of Tables

1 Introduction

Over the course of the past few decades, all around the world, payment cards have developed into a major, if not *the,* payment instrument.[1] What first began as a means of "dine now, pay later" (the Diners Club Card), soon became a prominent convenience for consumers. On the other hand, for the business world, it became an efficient and safe way of payment processing. With no doubt, the rise of the internet and globalization (i.e., increased international payments and travelling) sped up this development. Even though payment cards can be viewed as the shooting stars of modern payment processing, there is a critical backside to this development. Only a few networks seem to control the worldwide payment card business, with some regional heroes here and there. In 2008, in the U.S. alone, merchants spent over $60 billion in payment network fees (Federal Reserve Bank of Kansas City, 2009, p. 1). This development has caught policy-makers' attention, as more than 50 anti-trust cases since 2005 illustrate, mainly against two global players, Visa and MasterCard (Bradford & Hayashi, 2008, p. 1).[2]

For processing card payments, the so called acquirer (i.e., the institution connecting a merchant to a payment card network) deducts the merchant usage fee[3] from the nominal transaction value, a fraction typically between 1 and 3 percent, sometimes even up to 5 percent. Hence, merchants claim that the acquirers take advantage of

---

1 In 1986, the U.S. ratio of payments conducted by card was about 3% (Evans & Schmalensee, 2005 b, p. 3). This growth has continued even over the past few years. Regarding the percentage of payment cards used for in-store purchases in the U.S., Bolt & Chakravorti (2008, p. 15) refer to an increase from 43% in 1999 to 56% in 2005 (original source: American Bankers Association and Dove Consulting, 2005).

2 For example, in Australia, the U.S., and the EU, policy-makers investigated against Visa and MasterCard. For further detail, see also Schmalensee (2005a), European Commission Competition DG (2006), and U.S. Government Accountability Office (2009).

3 Merchant service charge and merchant discount rate are common alternative terms for what I refer to as the merchant usage fee.

their seemingly powerful market position and charge excessive fees for highly automated processes which involve very low processing costs.[4] Such claims raised quite a few questions among policy-makers, especially since payment card networks are very complex. This has spawned interest among economists since the very first lawsuit of NaBanco against Visa in 1979.[5]

Payment network complexity is inherent in the so called “two-sidedness” of the payment networks.[6] In contrast to any other “one-sided” market, which still involves two market sides, this so called “two-sided” market is a platform that can only exist if it succeeds to attract two different sets of customers at the same time, i.e., men and women to come to a night club or art collectors and owners willing to sell their precious pieces in an auction house. Such a platform could not function properly if one side, or the other, stayed away.[7]

---

4 Even so, some authors suggest that (purely) cost based payment card network pricing is not equipped to achieve social efficiency (Evans & Schmalensee, 2005b; Wright 2003; Gans & King, 2003). This has to do with the fact that payment card networks are a type of two-sided market, as will be explained in the remaining. Thus, most of the costs and benefits of such a platform cannot be allocated to either one of the platform sides, i.e., in the case of a payment card network; merchants and consumers respectively.

5 NaBanco argued that the monopolistic setting of Visa’s unified interchange fee was anti-competitive. Even so, the Eleventh Circuit stressed the pro-competitive elements of the unified fee, and thus, ruled against the claim and in favor of Visa. Refer to: National Bancard Corp. (NaBanco) vs. Visa U.S.A., Inc., 596 F. Supp. 1231, 1265 (S.D. Fla. 1984), *aff’d.*, 779 F.2d 592 (11th Cir. 1986), *cert. denied*, 479 U.S. 923 (1986). For details, see Evans & Schmalensee (2005a).

6 A number of authors refer to payment card networks as a classic example of a two-sided market; Armstrong (2006), Evans & Schmalensee (2005a & 2005b), Jullien (2005), Rochet & Tirole (2002, 2006a, 2006b), Rysman, 2007–just to name a few. Other examples of two-sided markets are video game consoles, television channels, auction houses and shopping malls. Evans & Schmalensee (2005a, pp. 133ff.) offer three clusters of two-sided markets; market makers, audience makers, and demand coordinators (e.g., payment card networks).

7 As a result, payment card networks have to carefully consider their price structure, i.e., the split of prices between the two platform sides, rather than only considering the price level (total). Hence, it might be more efficient for a TV-channel to subsidize the audience by offering its programs for free and charge high prices for advertising minutes.

1.1 Motivation

The literature broadly discusses questions related to the payment networks' pricing systems and their impact on social welfare.[8] Even so, there remains little consensus on efficient payment card network pricing from a societal point of view (Bolt & Chakravorti, 2008, p. 16). Beyond that, there is also a need for a better understanding of the determination, i.e., private optimization of the merchant fee (Evans & Schmalensee, 2005b, pp. 30, 40). For instance, while investigating in a case against Visa, the European Commission Competition Directive Generale[9] (2006) found that in the EU, the average merchant usage fee spreads between 0.4% and 2.4% of the transaction value, depending on the merchant sector (Figure 2-1). In addition, the U.S. Government Accountability Office (2009) observed that the price level of payment card services depends on the "merchant category", i.e., the merchants' line of business and their competitive environment.[10] The existing literature on payment card networks does not sufficiently explain these variations in payment card fees. Hence, the purpose of the present study shall be to help fill this gap, and thus, to gain a better understanding of the determination of payment card network prices (and especially the merchant usage fee).[11]

Most of the payment card pricing literature focuses on the aspects of the so called "two-sidedness" of the payment card market. For instance, Armstrong (2006) presents three determinants of payment card networks; the level of cross-group

8 For an overview and the historical development of the literature on payment card network pricing, refer to Bolt & Chakravorti, 2008, and Evans & Schmalensee, 2005b.

9 In the remaining, it is referred to as the European Commission of Competition DG.

10 Apparently, these deviations in merchant usage fees raise questions among policymakers (see e.g., European Commission Competition DG, 2006, p. 40-42). At first glance, these variations in fees might suggest the market harming behavior of the payment networks and the banks, respectively.

11 Simultaneously, and without being aware of each others' research, Wang (2010) presented a model with a fairly similar set-up as the present study.

(network) externalities, the mix of fixed and per-transaction prices, and the level of multi-homing, i.e., how many different payment cards consumers hold and use and how many different cards a merchant accepts. In other words, these factors concern the (vertical) upstream payment card market, but do not take into consideration much of the downstream market, i.e., the market where a merchant offers and sells products to consumers.[12]

It seems that none of these upstream market determinants can sufficiently explain the above variation of the merchant usage fee; neither can the (process) costs of the network(s)[13], nor the merchants' benefits from payment card processing. The latter becomes vital when considering the fact that a hotel does not necessarily reap much greater benefits (e.g., costs savings or expanded sales) from letting consumers pay by card than a petrol station does. Still, hotels face significantly higher merchant fees than petrol stations, i.e., 2% versus 1% (EU Commission Competition DG, 2006, p. 41).

This study shall help to better understand the determination of payment card network prices, also looking for the reason in the variation of the merchant usage fee between different merchant sectors. I suggest that the deviations in the merchant fees–at least to some extent–depend on the conditions of the downstream market. This claim is derived from the observation that a payment card network obtains a share of the economic rent from the merchant who accepts the network's card. The corresponding

---

[12] One exception to this might be the discussion of the surcharging behavior of merchants. For further detail, see e.g., Schmalensee & Evans (2005b). This discussion arose with the observation that the payment card networks sometimes imposed the so-called "no surcharge rule" on merchants by which a merchant was prohibited to surcharging the consumers' use of payment cards. The main concern of policy-makers and economists was that such a rule might bring about inefficiencies in the payment card markets, since the merchants were forced to hand the payment card network fee on to all consumers, cash, as well as card paying.

[13] Note that the network's variable costs of payment processing are negligibly small, especially when compared to the huge fixed costs of setting up and running the network (e.g., infrastructure costs and the costs of developing secure technologies; cards, readers, etc.).

intuition is that the factors determining the profit margin of a merchant similarly influence the merchant fee of a payment card network. Exemplary measures of the downstream market condition might be the price elasticity of consumer demand regarding the product exchanged between a merchant and a consumer, product substitutability[14], and the competitive condition of merchants (i.e., monopoly power). For example; fuel is a homogeneous product, thus, drivers are rather price sensitive. Hence, petrol stations face fairly strong competition. In contrast, restaurants offer rather inhomogeneous products, spreading from fast food to exclusive cuisine. Consequently, competition tends to be weaker among restaurants[15] than among petrol stations. As such, we can expect a restaurant selling a meal to make a bigger profit margin than a petrol station when selling fuel. Hence, the observation that restaurants, on average, pay a merchant fee of 2% and petrol stations pay 1% of the transaction value supports the above hypothesis (European Commission Competition DG, 2006, p. 41).[16]

The next subsections contain, firstly, a brief introduction to the history and the business setting of payment card networks, and secondly, a brief introduction to the general models derived in this thesis and a brief summary of the corresponding results.

---

[14] I suggest product substitutability as a downstream market determinant of the merchant usage fee, being particular intuitive. Product substitutability can be characterized on a continuum which ranges from products being (perfect) substitutes (e.g., fuel) via no substitutes at all (i.e., the monopolistic merchant) to the products being (perfect) complements (video game consoles and the respected games).

[15] Even so, we can expect much stronger competition among fast food restaurants than among gourmet restaurants.

[16] Also, due to the rather strong competition, I expect that petrol stations will tend to more quickly deny credit cards (or similarly, surcharge card payments) than gourmet restaurants. Interestingly, in the U.S., there is a trend that petrol stations have started surcharging the use of payment cards. In contrast, empirical studies in the Netherlands and in Sweden found that, in general, merchants mostly restrained from surcharging (ITM Research, 2000, and IMA Market Development AB, 2000).

1.2 Framework

Besides traditional cash, payment cards have grown into a prominent payment instrument. Examples for further payment instruments are checks, stored value cards, gift certificate cards, and automated clearing houses. All together, there are three types of payment cards; credit, debit, and charge cards. While this study focuses on the payment function of cards, we need to notice that credit cards furthermore expand liquidity (credit function), thus enabling forwarded consumption which we do not concern.

Payment card networks involve up to five players; a consumer and his/her bank, a merchant and his/her bank, as well as the payment card network providing a common framework (rules & standards, technology, process, brands).[17] Consider the case where a consumer buys a product from a merchant and pays with a card. The card issuing bank (issuer) deducts the price from the account of the consumer and transfers it to the acquirer, i.e. the institution connecting the merchant to the payment card network.[18] When transferring the money to the acquirer, the issuer deducts a processing fee from the nominal transaction value, the interchange fee. The acquirer then credits the price of the product minus the merchant usage fee to the account of

---

17 For an overview of the payment card industry, refer to Evans & Schmalensee (2005a).

18 Acquirers enable merchants to convert card payments into balances in their bank accounts. On the other hand, issuers provide payment cards to consumers.

the merchant. Note that the interchange fee seems to be a lower bound for the merchant usage fee.[19]

In general, payment card networks can be classified as unitary and multi-party systems. Examples of multi-party payment systems include the worldwide Visa and MasterCard networks, and local heroes such as the EC-card network in Europe. On the other hand, the most prominent examples of unitary networks are American Express and the Discover Card. Multi-party systems function in the way of a franchise organization. Member banks in the network serve as issuers and acquirers; the network operator provides a common brand, as well as the system framework (i.e., transaction standard, operations' rules, to some extent processing infrastructure). In contrast, unitary payment networks issue cards, execute the acquirers' function, and operate the network all from one organization (Figure 1-1). Even so, unitary payment card networks involve more players than a first glance might suggest. Besides having to maintain a customer relationship with card holders and merchants accepting the system's card, the network also has involved two more players in the payment process, i.e., the merchant's bank and the consumer's bank.

The first general-purpose payment card was invented in 1949 by Frank McNamara, the president of the New York credit company; the "Diners Club" card was designed as a unitary payment card network that at first enabled consumers to pay for their dinner at one of the fourteen club restaurants in New York without cash. By 1960, three T&E networks (mainly used for travel and entertainment–T&E) dominated the

[19] The U.S. Government Accountability Office (2009, p. 35-38) observed that the merchant usage fee consists of two components, the interchange fee and the processing costs. Merchants report, that they are able to negotiate the processing costs, but seemingly many are not successful in bargaining over the interchange fee. *„[Several] merchants told us that they generally paid the rates listed in the Visa and MasterCard networks' default interchange fee schedules. Although the ability to refuse to accept Visa and MasterCard should provide merchants with the leverage to negotiate lower interchange fees, merchants report that they could not refuse to take such cards because of customer demand."* In contrast, the networks report they are willing to negotiate with merchants and there seems to be a minority of merchants affirming this statement.

national U.S. payment card market; Diners Club, American Express and Carte Blanche, a card issued by Hilton Hotels. These three only faced significant competition by banks on a regional level, since interstate banking regulations and other hurdles made it hard for banks to expand their card business.

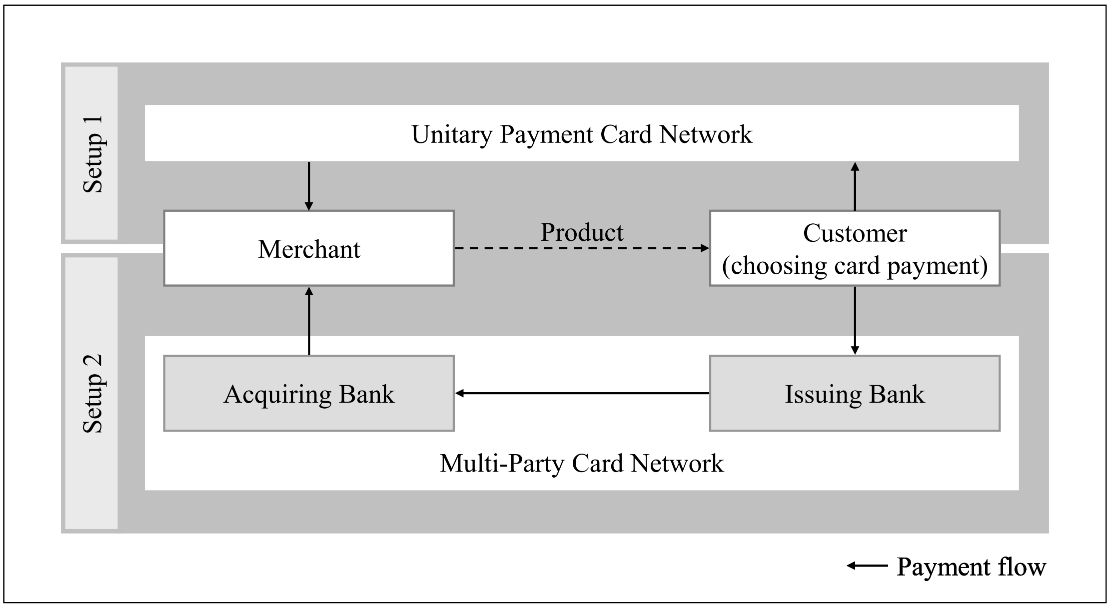

Figure 1-1: Alternative Setups of Payment Card Networks.

To overcome these hurdles, in 1966, the Bank of America decided to expand their business with a sort of franchise concept, which later became the worldwide Visa network. Hence, the first multi-party payment network had evolved. Other banks quickly adapted this strategy of forming an alliance and issuing payment cards. Even so, by the year 1968, it became apparent that two of these alliances would make the race: the BankAmericard franchise system and the Interbank cooperative system (Evans & Schmalensee, 2005, pp. 53ff).

The unitary networks, often due to their familiarity with certain consumer groups, have a rather focused consumer approach. For example, American Express, being historically very inclined with the travelling business (i.e., traveler checks), is fairly strong in the travelling segment; a successful strategy, since the travelling business turned out to be a prime market segment. In contrast, from the beginning, the multi-party payment systems had a broader consumer approach since a multi-party network enables the associated member banks to provide payment services (acquiring) and cards (issuing) to the banks' client base. This strategy quickly outgrew the unitary networks strategy.

In the 1960s and 1970s, multi-party payment card networks were the only ones succeeding in the credit card market. In fact, in the late 1960s, American Express had tried to go into the credit card business with its Uni-Card, but due to a lack of success, decided to get out and did not try again for another twenty years (Evans, & Schmalensee, 2005, pp. 65). Apparently, the multi-party approach of banks issuing cards to their clients helps to better provide the necessary information to control the risk of consumer illiquidity.

Card Issuer	**U.S. market share [%] by outstanding credit card balances**
JP Morgan Chase	21%
Bank of America	19%
Citi	12%
American Express	10%
Capital One	7%
Discover	6%
Wells Fargo	4%
HSBC	3%
U.S. Bank	2%
USAA Savings	2%
Total	**88%**

Table 1-1: Ten Largest U.S. Credit Card Issuers as of Year-End 2008
(Source: U.S. Government Accountability Office, 2009, Table 1, p. 6)

Even though the payment process of a unitary payment network at a first glance might appear less complex than that of the multi-party systems, this impression is wrong, as the payment process of the unitary network also involves the (house) banks of the merchant and the consumer. In addition, the concept of issuing cards, alongside other banking services, to existing customers might involve synergies (lower customer and account management costs, etc.) which the unitary networks generally can not generate. But even though in the U.S. alone, for example, more than 6,000 institutions issue credit cards, the issuing market has been highly consolidated over the past decade. In recent years, a number of large merchants (Target, Kohl's, etc.) have sold their card business to the big card issuers. Table 1-1 shows the ten largest U.S. issuers who account for 88% of the market, including American Express and Discover (U.S. Government Accountability Office, 2009, p. 6). Another interesting trend is that, increasingly, merchants seem to approach their customers by issuing cards. In doing so, merchants apparently recapture some of the paid merchant fees, and beyond that utilize their existing customer relationships to generate further interchange fee revenues. When issuing reward cards, merchants can create synergies with their core line of business; such as e.g., airlines issuing credit cards, giving away 'miles' for card purchase volume, thus free flights, etc. This trend reflects the present competition for consumers who, not surprisingly, play a key role in the payment card markets.

In the past, the multi-party payment card networks were organized as non-profits associations. In 2005, MasterCard became a publicly traded for-profits organization, in 2008, Visa followed. Even so, apparently there seems to be no change of strategy of the networks. Industry observers assume Visa and MasterCard underwent the organizational change in order to avoid potential conflicts with anti-trust laws (Wang, 2010, p. 92).[20]

---

[20] Note, that being an independent public firm may help the networks with regard to anti-trust claims, since monopolistic pricing in itself is not considered an anti-trust issue.

With the present study's aim at gaining a better understanding of the relationship of the merchant usage fee and the downstream market characteristics, three models, each illustrating a payment market, are constructed that incorporate certain downstream market characteristics. The primary purpose of the present approach is that the models view the payment industry as a vertical to the so called downstream market (where merchants and consumers interact) with monopolistic networks on top; even so, oligopolistic settings can easily be derived from this. In detail, this study takes four "downstream determinants" of the merchant usage fee into account; the consumer price elasticity of the good exchanged, product homogeneity, the competitive condition of merchants, and the relative frequency of card usage indicating the consumers' preference to pay by card or with cash. Similar to Wang (2010), the below models do not rely on the distribution of consumer card benefits, but much rather incorporate the benefit of the cards in the interaction of consumers and merchants.

This approach is in contrast to much of the existing literature, which relies on the distribution of merchant and consumer card benefits as well as strategic competition of the merchants. Therefore, the present literature brings about results which *"…are less conclusive in terms of ... explaining stylized facts (Katz, 2001; Hunt 2003; Rochet, 2003a; Rochet and Tirole, 2006b)."* (Wang, 2010, p. 87.)

1.3 Results

In the present thesis, a mature payment card market is analyzed using a general model adapted and analyzed for three different settings. A mature payment card market shall be a market with equilibrium prices and no strategic considerations of the respected card networks and banks. In detail, the players of the general model are the consumers, the merchants, the card network (or card scheme) and the banks, in the case of a so called multi-party payment card network. While the acquirer is the bank connecting the merchant to the card network, the issuer is the bank providing cards to consumers. For every card payment, I consider the acquirer to charge the merchant a transaction fee which shall be a percentage of the transaction volume; the merchant usage fee.

The timing of the considered game is as follows. Firstly, acquirers and merchants bargain over the merchant usage fee. In the last setting of a multi-party payment network, the network defines the interchange fee before the acquirers go into negotiations over the merchant usage fee. Secondly, merchants determine their payment policy. Thirdly, in the case of quantity competition, the merchants decide how many products to supply to the market, or in the case of price competition, the merchants define the product price. Lastly, consumers choose whether to acquire the good at the specific market price depending on their demand function. In the remaining, the game is solved by backward induction. Results are then derived via comparative statics. Over the course of this study, the results will be affirmed for three settings; unitary and multi-party payment networks, merchants practicing quantity and price competition.

The main result of this study is that the merchant usage fee depends on the downstream market conditions. Thus, I find reason for the variation of the merchant usage fee depending on the merchant sector.[21] Again, the payment market is analyzed within three settings, each of which is a self-contained study (Sections 2, 3, and 4).[22] Firstly, I model a monopolistic unitary payment card network with merchants under Cournot quantity competition and one homogeneous product being

[21] This is an observation of the European Commission Competition DG (2006, pp. 40-42).

[22] Repetitious elements in the different sections of this thesis are unavoidable, since it consists of three self-contained papers.

sold in the downstream market (Langlet, 2009).[23] The homogeneous product allows for the assumption of isoelastic demand, thus providing an excellent setting to analyze the impact of the price elasticity of demand onto the merchant usage fee. I find the merchant usage fee increasing in the price elasticity of consumer demand and decreasing in the relative frequency of card usage. In addition, we determine the maximum merchant usage fee, a threshold merchant usage fee for which merchants are at the edge of denying cards.

Secondly, in a joint study with Jens Uhlenbrock, we extend the first study by considering merchants under price competition (Langlet & Uhlenbrock, 2010). Here, the linear demand function with duopoly merchants under the Bertrand price competition[24] of Singh and Vives (1984) allows for an analysis of inhomogeneous products. Besides the affirmation of the determinants of the price elasticity of consumer demand and the relative frequency of card usage, product homogeneity is clearly shown to impact the merchant fee.

Thirdly, in a joint study with Clemens Buchen, we expand upon the first model, the case of a multi-party payment card network, by assuming quantity competition among merchants (Langlet & Buchen, 2010). In this study, we model the population

---

23 In the case of Cournot quantity competition, a merchant maximizes profits over supply quantity. In doing so, the merchant anticipates the behavior of the other merchants, i.e., the product quantities offered by the other merchants. By doing so, this merchant decides to offer that quantity which from his/her point of view is optimal; such that, the chosen quantity is the "best response" to the anticipated quantity offered by the others. In this study Cournot quantity is assumed competition, since one focus of the determinant is the price elasticity of consumer demand. Thus, it is especially appealing to consider isoelastic demand which Cournot quantity competition allows for.

24 In the case of Bertrand price competition, a merchant maximizes profits over product price. In doing so, the merchant anticipates the behavior of the other merchants, i.e., the other merchants' price. Compared to Cournot quantity competition, Bertrand price competition is known to achieve more societal efficient results. In this model, the assumption of Bertrand price competition enables the analysis of heterogeneous products, which allows for an analysis of product substitutability as, what we found, a downstream market determinant of the merchant usage fee. This broadens the aim of this study, thus providing a broader generality with regard to the results. In the case of Bertrand price competition, a merchant maximizes profits over price.

of network banks and find two general types of banks, all-purpose banks and specialist banks. An all-purpose bank is a bank that serves both sides of the payment card platforms, i.e., merchants and consumers, to an equal extent. Thus, the bank's balance of paid and received interchange fees evens out. Still, even in the case of an all-purpose bank, the level of the interchange fee is relevant for the bank's negotiations with the merchants over the merchant fee, since the interchange fee serves as a lower bound to the merchant fee. Consequently, an all-purpose bank is interested in a rather high interchange fee, thus transferring profits from the acquiring to the issuing side of the network.[25] In contrast, specialist banks mainly serve either as a card issuer or as an acquirer; thus, the level of the interchange fee determines the revenue and the profits of the specialist banks.[26] In reality, we find that all-purpose banks and specialist banks are both examples of actual business models in the industry. Among the eight issuers associated with multi-party payment card networks in Table 1-1 (excluding the unitary networks American Express and Discover), there are a mere two examples of the same company issuing and acquiring to a similar extent, i.e., being all-purpose banks; JP Morgan Chase and Bank of America. On the other hand, there are firms who specialize in either acquiring or issuing. According to the net interchange fee balance, all banks associated with a multi-party payment card network could be plotted on a continuum reaching from perfect issuers, via an all-purpose bank to a specialist acquirer.

---

[25] This finding might serve at least as a partial explanation of the observation that issuing is far more profitable than acquiring (European Commission Competition DG, 2006, pp. 62-77).

[26] Note that the balance of paid and received interchange fees within the whole network is zero, as every transaction involves one acquirer paying and another issuer receiving the interchange fee. In reality, only the rather little processing costs of the network will be deducted from the interchange fee; these processing costs include the network operator's Transaction Processing Fee. Wang (2010, p. 89) refers to an unreferenced Visa-source stating that the networks charges to the member banks consists of a cost-based Transaction Processing Fee as well as member's contribution fee depending on the number of issued cards, the total transaction, and the sales volume. Besides, the networks are known to collect an entry fee from newly associated member banks.

Specialist issuers are interested in high interchange fees. In contrast, specialist acquirers are interested in low interchange fees, if not a zero interchange fee. Since there is symmetry regarding the paid and received interchange fee revenue volume of issuing and acquiring, and the majority of the associated banks (i.e., 'net' issuers and all-purpose-banks) are interested in high interchange fees, the outcome will be rather high interchange fees. In addition, the network operator (i.e., Visa or MasterCard) is interested in maximizing the interchange fee revenues, since it receives a part of the interchange fee, the so called Transaction Processing Fee. For this reason, Wang (2010) assumes privately optimized interchange fee revenues for the networks.

To conclude, fee regulation is found to be particularly useful, since it reduces merchant usage fees, especially for less competitive markets where consumers already face inefficiently high prices for the products and services they purchase. Furthermore, regulation of the interchange fee appears useful, since the multi-party payment card networks rather tend to bring about market inefficiencies due to double marginalization of acquirers and issuers. In addition, multi-party networks are likely to transfer profits away from the acquiring to the issuing side of the network, thus weakening the bargaining power of the merchants. However, fee regulation requires a thorough and ongoing observation of the payment card markets, especially since it lacks adaptability and self-regulatory features. Furthermore, fee regulation might render unintended effects, such as encouraging card over-usage.

2 Unitary Network with Quantity Competition among Merchants[27]

Even though there has been an enormous increase in the amount of literature on two-sided markets over the last decade, there remains a significant need for a deeper understanding of the determination of (i) privately optimized and (ii) socially optimized payment card network pricing (Evans & Schmalensee, 2005b, pp. 30, 40). In this section, the behavior of payment card network players is analyzed with the particular aim of understanding how payment card network fees are determined. The focus is on three determinants of the merchant fee: the consumer price elasticity of the good exchanged on the downstream market, the relative frequency of card usage, and the competitive condition of merchants (i.e., monopoly power).

2.1 Introduction

Economic issues inherent in payment card systems such as Visa, MasterCard and American Express have motivated growing interest in research in the area of two-sided markets. In fact, a number of authors have referred to payment card networks as a classic example of a two-sided market[28]. Recent anti-trust cases, among others in Australia, the U.S., and the EU, have added momentum to the ongoing discussions. Questions related to the card networks' pricing system and its impact on social welfare have been widely discussed. Even so, how payment card fees are determined and many questions regarding the regulation of such networks remain unsolved (Evans & Schmalensee, 2005b, p. 4).

27 This section has been published in a slightly altered version; Langlet (2009).

28 For further details see Armstrong (2006), Evans & Schmalensee (2005a & 2005b), Jullien (2005), Rochet & Tirole (2002), and Rysman (2007).

In this section, I introduce the model of a monopolistic unitary payment card network. My investigation aims at understanding the factors that influence the network's choice of the merchant fee. In the model, the competitive condition of merchants is implemented as the number of merchant players, which is indicated by N. When merchants make supernormal profits (i.e., for all $N<N^*$), the merchant fee appears to be independent of the merchants' competitive condition.

Evans and Schmalensee (2005b) present a valuable overview of the historical development of payment card systems in a broad discussion of the challenge of payment card regulation and, in particular, the regulation of interchange fees. In general, Baxter (1983) was the first to approach payment systems with a distinct focus on two-sidedness. Baxter therefore introduces a model that incorporated platform costs and benefits, thereby explaining how the corresponding per-transaction fees are determined. With a simple numerical example, Baxter shows that an interchange fee might be necessary to help reallocate the costs and benefits of the card system.

The challenge of efficiently regulating payment card pricing is rooted in the two-sidedness of payment card networks. Tirole & Rochet (2006a) define a two-sided market as a market for which the fee level and fee structure are relevant with regard to the price equilibria. Tirole and Rochet go on to define the fee structure as the allocation of fees between participants on either side of a platform (in the case of a payment card network, these participants are consumers and merchants). With such a complex fee structure, the common principles of cost-based regulation cannot generally be applied. One robust conclusion of the economic literature is that a pure cost-based regulation in regard to the interchange fee is not socially optimal (Evans & Schmalensee, 2005b, pp. 32, 38).

Tirole and Rochet (2006a) consider a one-sided market as a market for which the fee structure is irrelevant with regard to the market equilibria, as long as the (total) fee level is kept constant. In fact, the participants in a one-sided market will negotiate away the structure of fees imposed on them. A common example of such a one-sided market would be the value-added tax (VAT). The VAT can be viewed as a platform, and it does not matter which side, consumers or merchants, is charged the VAT. According to this definition, ATM networks would also be one-sided markets, since surcharging is common practice.[29]

Carlton and Frankel (1995a, 1995b) find that, in a perfectly competitive system and with no costs of price discrimination in regard to the payment method, card usage will be efficient. Under these assumptions, the payment card fee structure will be irrelevant to the transaction volume (Carlton & Frankel, 1995a and 1995b; Evans & Schmalensee, 1995), which would make payment card networks one-sided markets. It turns out that the assumption of a perfectly competitive system is not even necessary to achieve card usage efficiency. Even with costless surcharging alone, the fee structure will be irrelevant since it will be negotiated away between merchants and consumers. Note that the assumption of costless surcharging includes the absence of a “no discrimination rule” (Gans, King (2003), pp. 12, 13).

Even so, payment card networks are considered two-sided markets, since merchants rarely surcharge (or give cash discounts), a behavior that seems, for the most part, to be independent of an existing “no discrimination clause” that a card network might impose on merchants. Surcharging apparently is not costless; in fact, merchants fear consumers’ preference for patronizing stores that do not surcharge (Evans & Schmalensee, 2005b, pp. 26-27; Chakravorti, 2003, p. 55; ITM Research, 2000, pp. 7-10; IMA Market Development AB, 2000, p. 18).

Even though there has been an enormous increase in the amount of literature on two-sided markets over the last decade, there remains a significant need for a deeper

---

[29] Rochet & Tirole (2006a, pp. 648ff.) offer a formal definition and further discussion of one- and two-sided markets.

understanding of the determination of (i) privately optimized and (ii) socially optimized payment card network pricing (Evans & Schmalensee, 2005b, pp. 30, 40). Rochet and Tirole (2003a) introduce a model in which the profit-maximizing players never set an interchange fee below the social optimum; a condition that depends on the assumption of identical merchants. Wright (2004) shows that relaxing this assumption leads to unpredictability in the deviation of the interchange fee from the social optimum. In fact, the interchange fee might be higher or lower than the social optimum requires (Wright, 2004).

Analyzing the European market of payment card systems, the European Commission Competition DG (2006) observes that network fees vary significantly between countries and between certain merchant sectors (Figure 2-1). These results remain mostly stable, even when analyzed for different networks (European Commission Competition DG, 2006, p. 41). The Commission's report makes no attempt to explain these differences in the merchant fee, but the impression of a market failure remains.

Armstrong (2006) finds three general determinants of equilibrium prices of two-sided markets: the level of cross-group (network) externalities, the mix of fixed and per-transaction prices, and the level of multi-homing of platform participants. Even though these determinants promise to hold true for payment card networks, as one example of a two-sided market, they cannot sufficiently explain the findings of the European Commission Competition DG. In general, there appears to be a gap in the literature that can explain the observations of the EU Commission Competition DG. This section is meant to help fill this gap.

In this section, I analyze payment card networks with the hypothesis that consumer demand elasticity is a determinant of the merchant fee of such networks. This research generally aims at gaining a deeper understanding of privately optimized payment card fees. The findings shall (i) help untangle the behavior of payment card systems and (ii) support policy-makers, managers and economists in gaining a deeper understanding of the payment card industry.

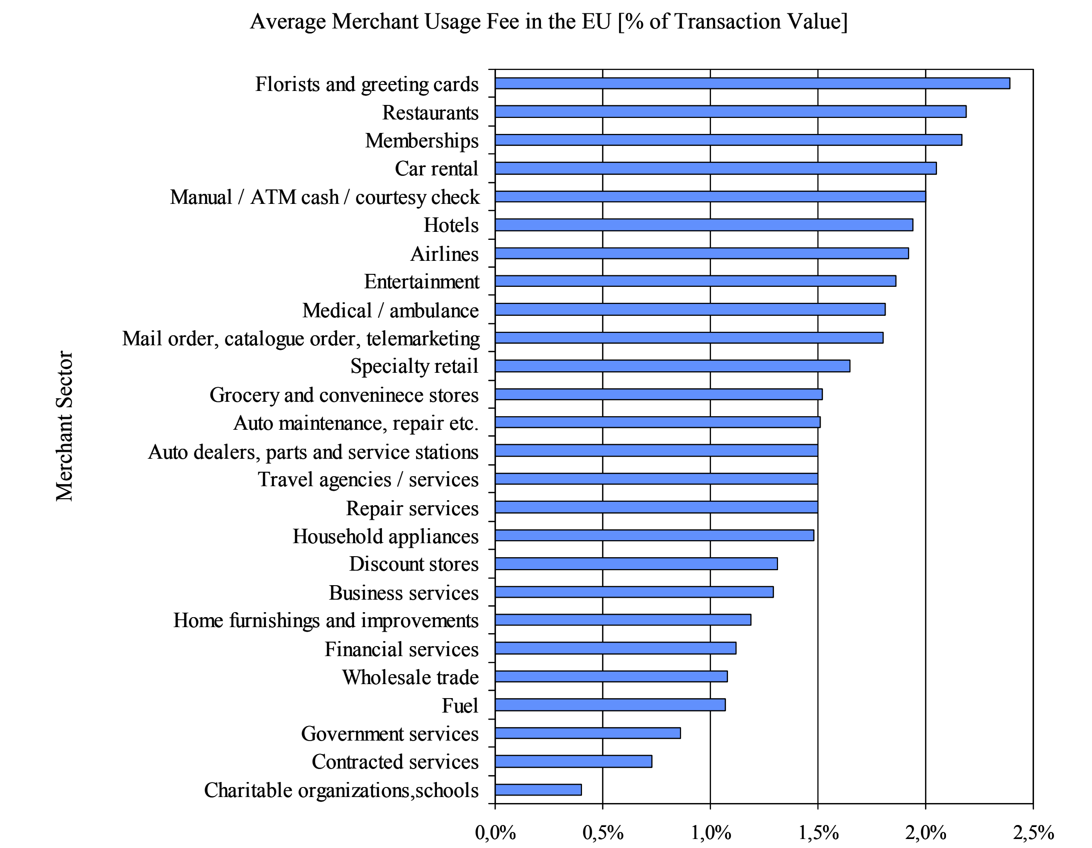

Figure 2-1: Weighted Average Merchant Usage Fee per Merchant Sector across the European Union in 2004 (Source: European Commission Competition DG, 2006, p. 41, Graph 18).

A cross-subsidy between the platform sides is a common strategy for two-sided markets. Platforms such as payment card networks will often try to attract one market side at low prices, if not for-free-services, in order to stimulate participation on the second side (Jullien, 2005, p. 257). I do not assume a certain population of card-holders here, since the study does not aim to analyze the network externalities of payment card systems. Thus, for parsimony, I assume no fixed platform fees of consumers, such as an annual membership fee. On the other hand, charging consumers no per-transaction fee is common practice in the payment card industry; even though there are exceptions[30], this model assumes a no per-transaction consumer fee. Furthermore, the model assumes only per-transaction fees and zero fixed fees for merchants; this assumption also fits with the fact that merchants typically pay only a small fixed fee beyond the merchant fee (Rysman, 2007, p. 5).

The network's variable costs related to card transactions are small; in fact, there are estimations that the payment system's variable costs per card transaction adds up to less than one Eurocent (Scheerer, 2007, December 17, p. 21). This amount appears particularly low when compared with merchant fees, which typically range between one and three percent of the transaction volume (Chakravorti, 2003, p. 53 and European Commission Competition DG, 2006, pp. 34-51). Even so, the fixed costs of establishing a viable payment card system are large since, in order to achieve an efficient payment process at the point-of-sale, large investments in information and communication technology are necessary (European Commission Competition DG, 2006, p. 7).[31]

Card payment systems generally have one of two organizational forms: unitary networks or multi-party networks.[32] For example, among the four major credit cards, there are two unitary systems (American Express and Discover) and two multi-party networks (Visa, MasterCard). A unitary system, besides operating the network, also performs the functions of issuing payment cards and acquiring merchants to join the network. In contrast, a multi-party network typically functions as a kind of franchise

---

30 For example, this was an observation of the European Commission Competition DG (2006, pp. 52-58).

31 For a discussion of this issue see Evans (2002).

32 In the literature, unitary networks are also referred to as closed or three-party systems, while, multi-party networks are sometimes called open or four-party systems.

organization by defining a common framework of technology and business principles. Multi-party networks leave card-issuing and merchant acquisition up to the associated partner banks (members). While, for parsimony, consider only the less complex unitary payment card systems, this study still promises to be a stepping stone toward similar research regarding multi-party networks. Finally, the model observes a price function with constant price elasticity.

The next subsection contains an analysis of the considerations of homogenous oligopoly merchants and of the payment card system. This analysis includes an observation of the behavior of merchants and card-holders (consumers). In the last subsection, the findings from the model are discussed and the final conclusions are offered.

2.2 Model

Consider a market of N identical merchants indexed by n, who are selling a homogenous product, with Q being the total sales quantity of the market. Let there be high price transparency, i.e., one market price p of the respective product, such that:

$$p_n = p_m = p \, , \tag{2.1}$$

where m and n are two different merchants.

2.2.1 Consumers

Merchants may offer two payment methods to consumers; card and cash. For parsimony, I assume a fixed rate γ of consumers choosing to pay by card, but who back off from the purchase if card payments are not an option. Consequently, $(1-\gamma)$ consumers choose to pay with cash. Furthermore, consider an effective no-discrimination clause that prohibits merchants to practice price differentiation with regard to the payment method, that is, merchants can neither surcharge card payments nor grant a cash discount. Let merchants incur only a per-transaction fee *a* as a percentage of the transaction volume and no fixed fee (Figure 2-2). In addition, let there be no per-transaction card-holder fee.

Since one of the objectives of this study is an observation of the impact of price elasticity on the merchant fee, I consider a price function *p(q)*, that characterizes (i) a constant inverse price elasticity ε and (ii) an inelastic price, but not perfect inelasticity.[33] Thus, from the perspective of merchant *n,* his/her inverse demand curve is:

$$p = Q^{-\varepsilon} = (q_n + q_{-n})^{-\varepsilon} \text{ with } 0 < \varepsilon < 1, \quad (2.2)$$

where q_n is the sales quantity of merchant *n* and q_{-n} is the quantity of products sold by all the other merchants (i.e., except for *n*). The next section provides an analysis of the merchants' behavior under the assumption that the market outcome is determined using the Cournot-Nash equilibrium.

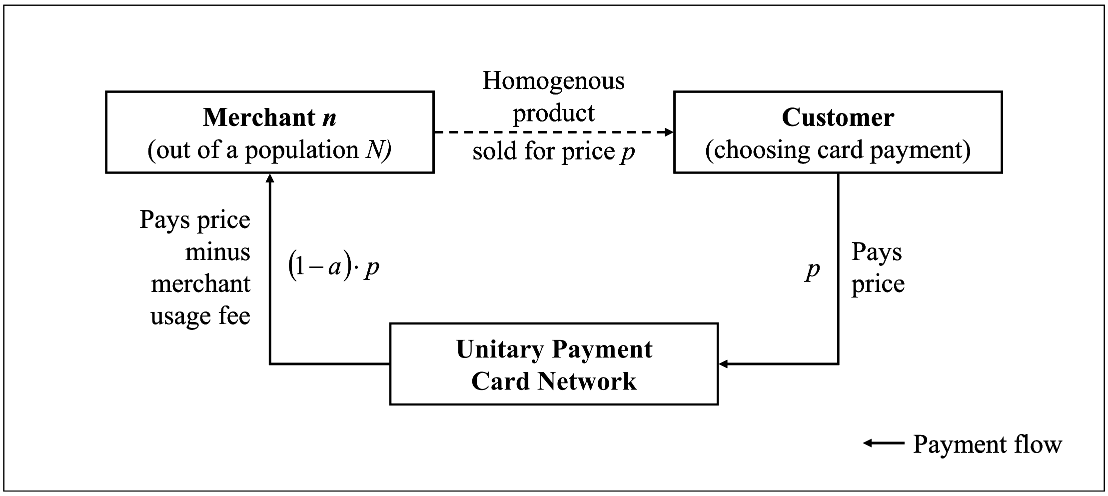

Figure 2-2: Payment Flow in a Unitary Payment Card Network (Cf. Gans and King, 2003, p. 4).

33 Note that in contrast to Section 3, ε indicates the inverse price elasticity of consumer demand.

2.2.2 Merchants

Not all merchants will necessarily participate in the payment card network. There will certainly be one $\bar{a}$ such that for any $a>\bar{a}$, merchants will refrain from accepting the respective cards (while still accepting the cards if $a=\bar{a}$). Therefore, we will assume that all merchants do accept the respective payment card(s), and will relax this assumption later in the analysis. In general, the sales revenue of merchant *n* will be the sales quantity multiplied by the price:

$$r_n = q_n \cdot p . \tag{2.3}$$

Let there be constant procurement costs *c* of merchants. In addition, merchant *n* will have to consider the payment card fee, which is the product of the per-transaction fee *a* and the relative frequency γ of card usage.[34] Consequently, the total costs of merchant n will be the total of the production costs and the total of the merchant fee, such that,

$$c_n = q_n \cdot c + \gamma \cdot q_n \cdot p \cdot a . \tag{2.4}$$

As for any other case, the profits of merchant *n* will be revenue minus costs:

$$\pi_n = r_n - c_n = q_n \cdot p \cdot (1 - \gamma \cdot a) - q_n \cdot c . \tag{2.5}$$

Thus, using Eq. (2.2), the profits of merchant *n* will be:

$$\pi_n = q_n \cdot (q_n + q_{-n})^{-\varepsilon} \cdot (1 - \gamma \cdot a) - q_n \cdot c . \tag{2.6}$$

34 Note that this behavior depends on the assumption of the merchant not price-discriminating based on the payment method; in other words, the merchant does not surcharge card payments.

Assuming merchants behave according to the Cournot-Nash equilibrium concept, merchant *n* will anticipate the expected sales quantity of all the other merchants and will maximize his/her profits by choosing q_n to solve:

$$\underset{q_n}{Max}\ \pi_n . \tag{2.7}$$

With $c > 0$ and $0 < \gamma < 1$, as well as $0 < a < 1$ and $0 < \varepsilon < 1$.

The first step in solving the model is to solve the first-order condition of Eq. (2.6). This solution requires the assumption that all merchants maximize profits and, therefore, choose identical sales quantities; thus,

$$q_n = \frac{q_{-n}}{N-1} = \frac{Q}{N} . \tag{2.8}$$

The transformation of yield in the second-order condition proves that this solution maximizes the profits of merchant *n*, bringing about the equilibrium Cournot sales quantity q^*.

Proofs are delegated to the Appendix.

Lemma 2.1 *The equilibrium Cournot sales quantity of merchant n will be:*

$$q^* = \frac{1}{N} \cdot \left(\frac{\left(1 - \frac{\varepsilon}{N}\right) \cdot (1 - \gamma \cdot a)}{c} \right)^{\frac{1}{\varepsilon}} . \tag{2.9}$$

Combining Eq. (2.9) with Eq. (2.8) the equilibrium sales quantity of the whole industry will be:

$$Q^* = \sum_{n=1}^{N} q_n = \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-\gamma\cdot a)}{c} \right)^{\frac{1}{\varepsilon}} \tag{2.10}$$

and the equilibrium Cournot price of the respective product will be:

$$p^* = \frac{c}{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-\gamma\cdot a)}. \tag{2.11}$$

In these steps, I assumed that the merchants were all participating in the payment card network, although whether they do so depends on the fee level and the fee structure imposed on merchants by the card system. The rational merchant will compare the expected profits while participating in the card network with the expected profits while not participating in the card network. I use the assumptions of γ consumers paying card and (1-γ) consumers paying with cash, and with $a=\bar{a}$ merchants are indifferent to participating in the payment card network. Assume that merchants deny cards for $a>\bar{a}$, while for $a=\bar{a}$ merchant n makes zero profits from sales paid by card[35]; thus,

$$\pi_n^{card} = \frac{Q}{N}\cdot Q^{-\varepsilon}\cdot(1-\bar{a}) - \frac{Q}{N}\cdot c \overset{!}{=} 0. \tag{2.12}$$

[35] This condition brings about an approximation for the cut-off usage fee. For parsimony, this condition implies the assumptions of cash and card paying consumers to be equally distributed along the continuum of the sales quantity q, thus the cash and card paid segments can be viewed like separate markets. More importantly, let the revenue of cash paid sales remain constant whether the merchant accepts cards or not. Even though latter is a simplification of reality, this assumption does not conflict with the generality of the model. In contrast to this approximation, in Section 2 the maximum merchant usage fee is thoroughly, even so implicitly derived.

Solving Eq. (2.12) while including the equilibrium Cournot sales volume of the industry (Eq. 2.10), brings about the cut-off merchant fee $\bar{a}$.

Lemma 2.2 *There is a maximum merchant fee $\bar{a}$, such that for any $a > \bar{a}$, merchants will refrain from accepting the payment card(s). Consequently in order to still attract merchants to participate in the payment card network, the payment card system(s) will have to define the merchant fee such that $a \leq \bar{a}$, where:*

$$\bar{a} = \frac{\varepsilon}{N - N \cdot \gamma + \varepsilon \cdot \gamma}. \tag{2.13}$$

With the elasticity ε and the relative frequency of card usage γ as constants, there will be one N^* such that for any $N \geq N^*$, merchants will make only normal profits from sales paid by card, for simplicity consider zero profit. Furthermore, for all $N < N^*$, merchants will achieve supernormal profits even from sales paid by card. Consequently, for all $N \geq N^*$ the network will set the equilibrium merchant fee at $a^* = \bar{a}$.

Discussion

When $a = \bar{a}$, an increase of the Elasticity of consumer demand by $\Delta\varepsilon$ under otherwise constant circumstances, will allow the network to also increase the merchant fee $\bar{a}$, because:

$$\frac{\partial \bar{a}}{\partial \varepsilon} = \frac{(1-\gamma) \cdot N}{(N - N \cdot \gamma + \varepsilon \cdot \gamma)^2} > 0 \text{ for all } N \geq N^*, \tag{2.14}$$

since $1 - \gamma > 0$ with $0 < \gamma < 1$ as well as since $N > 0$ and $(N - N \cdot \gamma + \varepsilon \cdot \gamma)^2 > 0$.

When there are normal merchants' profits, that is, when $N \geq N^*$, an increase in the relative frequency of card usage by $\Delta\gamma$ under otherwise constant circumstances, will also allow the network to increase the merchant fee $\bar{a}$, because:

$$\frac{\partial \overline{a}}{\partial \gamma} = \frac{\varepsilon \cdot (N - \varepsilon)}{(N - N \cdot \gamma + \varepsilon \cdot \gamma)^2} > 0 \text{ for all } N \geq N^*, \tag{2.15}$$

since $\varepsilon \cdot (N - \varepsilon) > 0$ with $0 < \varepsilon < 1 \leq N$ and $(N - N \cdot \gamma + \varepsilon \cdot \gamma)^2 > 0$.

Finally, from:

$$\frac{\partial \overline{a}}{\partial N} = \frac{-\varepsilon \cdot (1 - \gamma)}{(N - N \cdot \gamma + \varepsilon \cdot \gamma)^2} < 0 \text{, since } \varepsilon > 0 \text{ and } 1 - \gamma > 0, \tag{2.16}$$

we learn that, when there are normal merchants' profits, an increase in the number of merchants by ΔN will bring about a decrease in the merchant fee $\overline{a}$.

In the next section, we will observe the behavior of the unitary payment card network and determine N^*.

2.2.3 The Payment Card Network

Consider a monopolistic payment card system where the network incurs only fixed costs C (i.e., zero variable costs). Clearly fixed costs will not influence the pricing decisions of the card system, as long as the fixed costs do not consume all profits. In general, the card network's profits Π will be the accumulated fee revenue from the respective transaction minus fixed costs C, such that:

$$\Pi = p \cdot Q \cdot \gamma \cdot a - C = Q^{1-\varepsilon} \cdot \gamma \cdot a - C = \left(\frac{\left(1 - \frac{\varepsilon}{N}\right) \cdot (1 - \gamma \cdot a)}{c} \right)^{\frac{1-\varepsilon}{\varepsilon}} \cdot \gamma \cdot a - C. \tag{2.17}$$

First, consider the case where the payment network makes a take-it-or-leave-it offer, and later we will discuss a Nash bargaining case. In general the network will maximize profits Π over merchant fee:

$$\underset{a}{Max}\ \Pi\,, \tag{2.18}$$

under the side condition that overall costs coverage is assured. From solving this condition, the equilibrium merchant fee when merchants earn supernormal profits can be derived. The intention of the proof is solving the first-order condition of Eq. (2.18). When we assume that merchants participate in the card network, the transformation of yield in the second-order condition provides the final proof that the payment card system maximizes profits by defining the merchant fee such that $a=\hat{a}$.

Lemma 2.3 *Since, for $N<N^*$, merchants do participate in the card network, the card network will maximize profits by defining the merchant fee according to:*

$$\hat{a} = \frac{\varepsilon}{\gamma}\,. \tag{2.19}$$

From the perspective of the network the equilibrium merchant fee $\hat{a}$ (Eq. 2.19) is the optimum fee, thus maximizing the network's profit. Even so, we have learned that merchants will not accept any merchant fee $a>\bar{a}$. These results lead to the first Proposition.

Proposition 2.1. *The card network will define the merchant fee according to the optimum merchant fee $\hat{a}$ (Eq. 2.19) but with the constraint of not exceeding the maximum merchant fee $\bar{a}$ (Eq. 2.13), that is,*

$$a^* = \min\left\{\frac{\varepsilon}{\gamma};\frac{\varepsilon}{N - N\cdot\gamma + \varepsilon\cdot\gamma}\right\}. \tag{2.20}$$

In fact *N*=*N** is the competitive condition of merchants where: (i) merchants make zero profits (only normal profit) from sales paid by card, that is, $a=\bar{a}$ and (ii) the network is able to achieve the maximum profits by acquiring the merchant fee $\hat{a}$ according to Eq. (2.19). The solution of these two conditions, by substituting $\hat{a}$ from terms (2.19) in Eq. (2.13), brings about the second Proposition (See Figure 2-3 and details of the proof in the Appendix).

Proposition 2.2. *With a merchant population of N=N*, the maximum merchant fee* $\bar{a}$ *(Eq. 2.13) equals* $\hat{a}$ *(Eq. 2.19) which, from the perspective of the network, is the optimal merchants fee, i.e., the fee maximizing the network's profits:*

$$N^* = \frac{1-\varepsilon}{\frac{1}{\gamma}-1}. \tag{2.21}$$

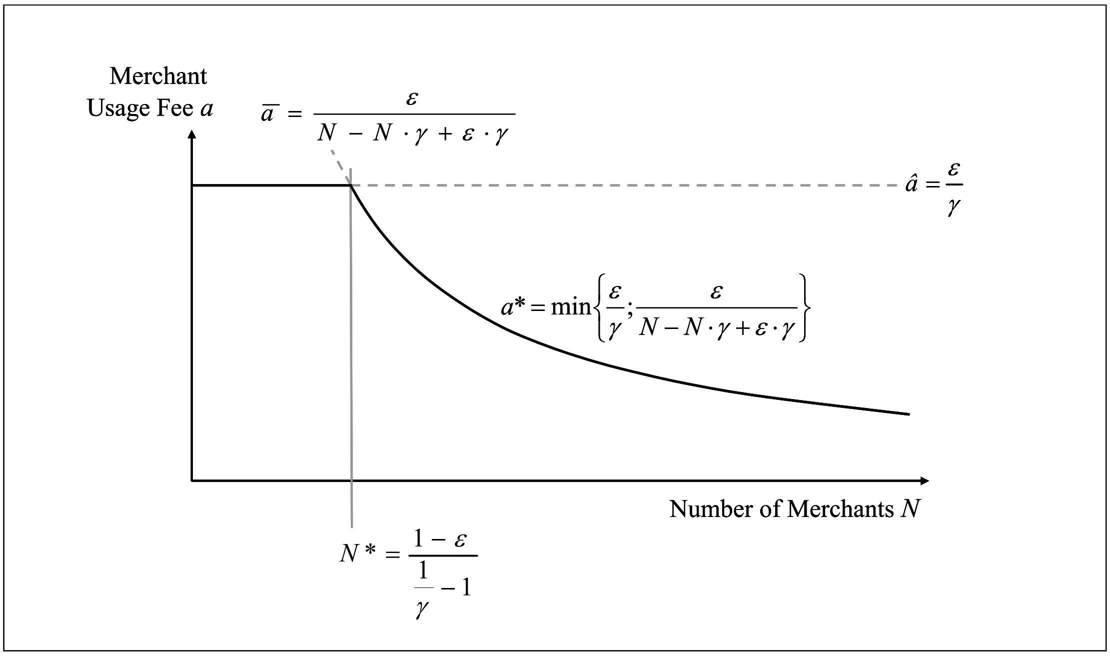

Figure 2-3: Unitary Network's Fee with Merchants under Quantity Competition.

Discussion

When there are supernormal merchant profits (as long as $N<N^*$), the competitive condition of merchants apparently does not at all impact the merchant fee. Thus, when there are supernormal merchant profits ($N<N^*$), an increase in the elasticity of demand by $\Delta\varepsilon$, with otherwise constant circumstances, will cause the payment card network(s) to increase the merchant fee a^* at an even higher rate, since:

$$\frac{\partial a^*}{\partial \varepsilon} = \frac{1}{\gamma} > 1 \text{ for all } N < N^* \text{ since } 0 < \gamma < 1. \qquad (2.22)$$

Furthermore, with supernormal merchant profits ($N<N^*$) and an increasing relative frequency of card usage by $\Delta\gamma$ under otherwise constant circumstances, the payment network(s) will decrease the merchant fee a^*, since:

$$\frac{\partial a^*}{\partial \gamma} = -\frac{\varepsilon}{\gamma^2} < 0 \text{ for all } N < N^* \text{ with } 0 < \varepsilon < 1 \text{ and } 0 < \gamma < 1. \qquad (2.23)$$

As long as merchants make supernormal profits ($N<N^*$), the profits of the payment card network depends on the competitive condition of merchants N, the production costs of merchants c and the fixed costs of the network C, as well as the elasticity of consumer demand. Interestingly, in this case, the networks profits appear to be independent of the relative frequency of card use, which can derived from the substitution of Eq. (2.19) into Eq. (2.17):

$$\Pi = \left(\frac{\left(1 - \frac{\varepsilon}{N}\right) \cdot (1 - \varepsilon)}{c} \right)^{\frac{1-\varepsilon}{\varepsilon}} \cdot \varepsilon - C \text{ with } N < N^*. \qquad (2.24)$$

By intuition, it becomes clear that the relationship shown in Eq. (2.24) will be true only in the case of a take-it-or-leave-it offer. Moreover, as long as merchants make supernormal profits, the relative frequency will not impact the market equilibria at all, but will only lead to an adjustment in the merchant fee, such that (i) the network's profits and (ii) the prices of the respective products remain constant. We learn this from combining Equations (2.11) and (2.19), thus:

$$p^* = \frac{c}{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-\varepsilon)}. \tag{2.25}$$

An increase in consumer demand elasticity by $\Delta\varepsilon$ or a decrease in the relative frequency of card usage by $\Delta\gamma$, under otherwise constant circumstances, will lead to a decrease of N^* since:

$$\frac{\partial N^*}{\partial \varepsilon} = -\frac{1}{\frac{1}{\gamma}-1} < 0 \text{ with } \frac{1}{\gamma}-1 > 0 \text{ and } 0 < \gamma < 1 \tag{2.26}$$

and since:

$$\frac{\partial N^*}{\partial \gamma} = (1-\varepsilon)\cdot\frac{-1}{\left(\frac{1}{\gamma}-1\right)^2}\cdot\left(-\frac{1}{\gamma^2}\right) = \frac{(1-\varepsilon)}{(1-\gamma)^2} > 0, \text{with } 1-\varepsilon > 0 \text{ and } (1-\gamma)^2 > 0. \tag{2.27}$$

2.3 Discussion and Conclusions

Armstrong (2006) analyzes competition in two sided markets. Examples of such markets include dating agencies, credit cards, TV channels, and shopping malls. In that paper, the author derives some key determinants of equilibrium prices of such two-sided markets; the level of cross-group (network) externalities, the mix of fixed and per-transaction prices, and the level of multi-homing. In the current section, I focus on a single market with specific characteristics allowing for more detailed results. In particular, I analyze unitary payment card systems, focusing on three additional determinants of usage fees; the consumer price elasticity, the relative frequency of card usage and the underlying competitive condition of the specific market in which the exchanged good is purchased. In the next paragraphs I will briefly discuss each of the determinants.

Elasticity of consumer demand

The retail and service industries are the downstream markets of payment card networks. With respect to merchants, payment card networks are in a strong market position, which enables them to extract part of merchants' revenues. Not surprisingly, the elasticity of demand has a significant impact on card usage fees and suggests a positive correlation between the merchant fee and factors that also have a positive correlation with the elasticity of demand, such as geographical clusters (e.g., countries, regions), merchant sectors (for instance the market for fuel, grocery stores, restaurants,, etc.) and consumer clusters (e.g., business and private). A useful extension of the current study would be an in-depth empirical analysis of the relationship of consumer price elasticity and the merchant fee.

Consumers' preference to pay by card (relative frequency of card usage)

Merchants and consumers who pay by card share the benefits of card payments (e.g., convenience, efficiency of the payment process, enhanced liquidity). On the other hand, the costs associated with card payments are typically borne by all consumers, card-paying and cash-paying alike, since merchants usually do not surcharge card payments. Therefore, policy-makers must continually reconsider the regulation of payment card systems. Without surcharging, merchants have to include the payment card fees in their pricing and profits calculations by adding the average usage fee multiplied by the probability of card payments γ. The relative frequency of card usage γ strongly influences the merchant fee. In contrast, when merchants' profits are supernormal (i.e., $N<N^*$), product prices and, therefore, merchants' and payment networks' profits, appear to be independent of the relative frequency of card usage γ (Proposition 2.1).

Competitive condition of merchants, indicated by the number of merchants

The competitive condition of the specific market in which the exchanged good is purchased, indicated by the number of merchants N, determines whether merchants will be able to make supernormal or normal profits. Thus, in the intersection of the maximum merchant usage fee and the optimal merchant fee (from the perspective of the network), i.e. $N=N^*$, depends on the elasticity of demand and the relative frequency of card usage (Proposition 2.2).

The analysis suggests a number of extensions and policy conclusions. With respect to the latter, imagine a policy-maker does limit payment card usage fees. Consider Figure 2-3: if the maximal usage fee allowed by law is above $\hat{a}$, there is no impact. However, if the maximal usage fee is below $\hat{a}$, it reduces the usage fee for less competitive markets (N small). This finding is particularly useful since, in less competitive environments, consumers already face an inefficiently high price. In the this model the reaction vector of the network is limited. Even so, when having to reduce the merchant fee due to regulator's enactment, such a limitation will entail a feedback loop, as the payment card network will react with different instruments trying to regain the loss of profits. One such instrument might be the installation of a bonus system to manipulate γ.

The purpose of this study is to disentangle how privately optimized payment card network fees are determined; how socially optimal payment card fees are determined requires further theoretical and, in particular, empirical cost/benefit analysis (Evans & Schmalensee, 2005b, pp. 30-32).

In the remaining, I discuss five possible extensions to this work. First, consider the case where N is very large. In that case, the foregoing model yields $a^*=0$ (Eq. 2.22), which would suggest that card networks do not enter such a market. However, merchants may see additional benefits to card payments which have not been modeled in the foregoing analysis. For instance, merchants may find it advantageous to lower their cash holdings; this kind of externality may render advantageous for merchants to accept a fee sufficiently large to induce card networks to enter the market, even with a large *N*. This may provide an explanation as to why discount stores have moved to accept card payments, e.g. in many European countries.

Second, we could extend the foregoing model by giving some bargaining power to merchants. Instead of assuming the above, that merchants are imposed a take-it-or-leave-it offer by the network, suppose we allow for bargaining. Bigger merchants, in particular, are often in a position where they can achieve lower merchant fees (European Commission Competition DG, 2006, pp. 35-40). Thus, consider Nash bargaining, and denote with α_i the bargaining power of the network. In that case, the fees that the parties agree on will be:

$$a^N = \alpha_i \cdot a^*, \tag{2.28}$$

where a^* is given by Eq. (2.20). In this case, N^* proves to be constant over the case of Nash bargaining or the take-it-or-leave-it offer (Appendix). In general, other than the fee level, the Nash bargaining case will not affect the above equilibria. In this respect, the findings of this section prove to be generalizable.

A third natural extension considers multi-party payment card networks, such as Visa or MasterCard. Consider two extreme cases, with the reality somewhere in between these two scenarios. In the first scenario, the card network consists of all-purpose banks which operate as card issuers to the same extent as they function as acquirers. For the second scenario, let there be specialist banks in the sense that these banks exclusively either operate as card issuers or as acquirers. Whereas the results of the first scenario will be similar to those of the unitary payment card networks, one finding for the second scenario will be the double marginalization of the two types of banks. This setting will be modeled in Section 4.

A fourth extension would consider several competing payment card networks, may that be unitary and or multi-party payment card networks. Among others, such a study promises findings in regard to the role of the payment card network fee regulation and, in particular, the current practice of regulating the merchant usage fee or the interchange fee of payment card networks.

Finally, another extension of the foregoing study would also consider non-homogeneous products. Singh and Vives (1984) offer a setting of duopoly merchants whose products characterize different levels of product substitutability. This level of substitutability is a continuum which ranges from the products being perfect substitutes to zero substitutability (the monopolist) or the products being complements. Thus, Section 3 offers a model extending the above study by adding a measure of substitutability based on the linear demand function of Singh and Vives (1984); this model considers merchants under Bertrand price competition.

3 Unitary Network with Price Competition among Merchants[36]

with Jens Uhlenbrock

This section investigates the determination of the merchant usage fee of a monopolistic unitary payment card network based on the characteristics of the downstream market. In this case, merchants engage in Bertrand price competition. The Bertrand price competition, in contrast to the existing literature[37], allows for an observation of heterogeneous products. We find that the payment card network extract a part of the economic rent that merchants obtain. The higher this rent, the higher the corresponding merchant usage fee. The rent, and consequently, the merchant usage fee, is increasing in the downstream market size, but decreasing in the price elasticity of consumer demand, as well as in the substitutability of products, and, interestingly, also decreasing in the fraction of consumers preferring card payments.

---

36 For full reference, refer to Langlet & Uhlenbrock (2010).

37 For approaches, similar to the present essay, refer to the other essays accumulated in this thesis, Langlet (2009) and Buchen & Langlet (2010), as well as Wang (2010).

3.1 Introduction

Over the past few decades, the increasing use of payment cards has spawned the interest of researchers and governmental regulatory agencies.[38] Payment cards are an example of a two-sided market, a market that needs to attract two different groups of customers in order to function properly. As two-sided markets exhibit the peculiar features contradicting standard microeconomic theory, they are an interesting research topic.

With regard to payment cards, there are two distinct customers: merchants accepting card payments and consumers wishing to pay by card. Any card network can only operate if it attracts a sufficient amount of both customer groups. As a result, the two-sidedness of the market requires unique pricing strategies. Payment card networks do not only have to consider price level (how much to charge in total) but also the price structure (the split of charges between the two market sides). It is a frequent feature of two-sided markets that one side is attracted toward very low prices, while the other side is charged substantially.[39] In the case of payment cards, cardholders typically do not pay anything for card usage, while merchants pay a per-transaction charge called the merchant usage fee. However, as the European Commission Competition DG (2006) has observed, as can be seen in Figure 3-1, there are vast differences in merchant usage fees across merchant sectors.

---

[38] According to the sources of Evans & Schmalensee 2005b, p. 3 (HSN Consultants 1987, 1991, 1992, 2001a, and 2001b, and U.S. Department of Commerce 2003), card usage in the U.S. has grown from a ratio of payments conducted by cards of about 3% in 1986 to 25% relative payment volume over cash in 2000. Bolt & Chakravorti (2008, p. 15), regarding the percentage of payment cards used for in-store purchases in the U.S., refer to an increase from 43% in 1999 to 56% in 2005 (original source: American Bankers Association and Dove Consulting, 2005). More than 50 anti-trust cases have been filed since 2005 by merchants contesting interchange fees.. (Bradford & Hayashi, 2008, p. 1).

[39] For a discussion of the so called “divide and conquer” strategy, refer to Caillaud & Jullien (2003).

In this section, we attempt to understand the reasons for these sectoral differences. We strive to enlighten the pricing decision of payment card networks dependent on the characteristics of a downstream market (the market the merchants face) where merchants have some market power. Our study can be seen as an extension of Langlet (2009), who investigates the merchant fee determination of a unitary payment card network with merchants under Cournot quantity competition. In contrast, we consider a unitary network and merchants under Bertrand price competition. Langlet finds that certain downstream market characteristics determine the merchant usage fee for the case of Cournot quantity competition among merchants selling a homogeneous product, e.g., the price elasticity of demand, the relative frequency of card usage and the competitive position of the merchants (i.e., monopoly power). We first illustrate a broader generality of Langlet's primary results and gain useful new insights, particularly based on the heterogeneity of products that Bertrand competition allows for.

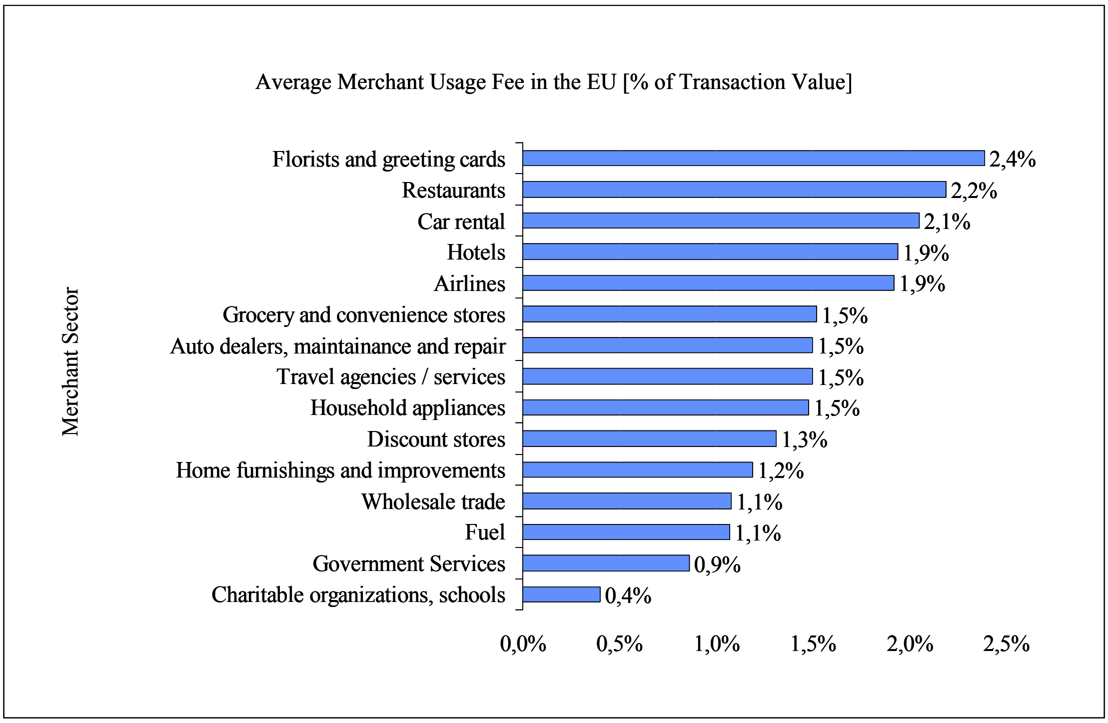

Figure 3-1: Weighted Average Merchant Usage Fee per Merchant Sector across EU, 2004 (Source: European Commission Competition DG, 2006, p. 41, Graph 18).

3.2 Literature overview

Payment card networks usually involve interactions of up to five players: a consumer, the consumer's bank (or the issuer), a merchant, the merchant's bank (or the acquirer), and the payment network providing the framework, which enables the transfer of payments among the agents (i.e., infrastructure, processes, transfer standards). When a consumer buys a product from a merchant and pays by card, the issuer charges the consumer's account and transfers the money to the acquirer who then credits the merchant's account. However, in order to execute these transactions, both the issuer and the acquirer bear costs for setting up and maintaining the needed infrastructure. The issuer typically does not charge the consumer for card usage, but covers the costs through an interchange fee deducted from the nominal amount it transfers to the acquirer. In turn, the acquirer does not credit the full nominal amount to the merchant's account, but instead deducts a so called merchant usage fee. The merchant usage fee has to be high enough to cover the interchange fee the issuer receives, as well as all other costs the acquirer incurs. All in all, through the merchant usage fee, the merchant is charged for the full costs of the card transaction and—to some extent—passes it on to consumers via higher prices.

Much of the existing literature has focused on the question of whether this price structure apparent in payment card networks is beneficial from the point of view of the networks or from a societal perspective. The interchange fee has especially attracted great scrutiny ever since it was the focus of a lawsuit NaBanco filed against Visa in the early 1980s (Chang and Evans, 2000; Evans and Schmalensee, 2007). Visa is a large multi-party payment platform which banks can join to become either acquirers and/or card issuers. To facilitate transactions among its many member banks, Visa had set a unified interchange fee that all member banks had to adhere to when conducting transactions under the Visa brand. NaBanco argued that the monopolistic price setting was anti-competitive and individual banks should be allowed to negotiate terms among them. However, the Eleventh Circuit ruled in favor of Visa, stressing the pro-competitive elements of the unified fee.

Baxter (1983), who was the first to look at payment networks in a scientific article, supplied some of the arguments adopted by the Court for applying a unified interchange fee. However, in reviewing the history of non-cash payment systems, he stressed that regulatory authorities had come to the opposite conclusion with different payment systems, essentially pushing for a zero interchange fee for checks through a clearance system of the Federal Reserve.

Nevertheless, the Court's decision did not end the discussion about interchange fees. As such, payment card pricing seems to continue attracting economists' attention.[40] One overall conclusion of this literature might be summarized as recognition that a collective setting of the interchange fee seems beneficial, as it avoids a serious free-rider problem and high transaction costs of inter-bank negotiations. Moreover, it cannot be asserted that any chosen fee level is socially beneficial and it is unclear whether such a beneficial fee should be higher or lower than the one set by a profit maximizing network. This calls for caution in the regulating activity. Contrariwise, the Polish Office of Competition and Consumer Protection forbade a multilateral interchange fee setting in 2007, while several other countries have ordered (or entered into mutual agreements with) Visa and MasterCard to lower their interchange fees; often based on cost arguments (Bradford and Hayashi, 2008).[41]

Another significant part of the literature investigates payment card network pricing decisions. These are particularly interesting, because of the two-sidedness of the market at hand. Evans (2002) and Evans (2003) provide a detailed description of why two-sided markets lead to unexpected results. Firstly, two-sided markets exhibit positive network effects, meaning that the product or service becomes more valuable

---

[40] An extensive introduction into the payment card industry is given in Evans & Schmalensee (2005a).

[41] For a useful overview of the related literature and the payment card industry in general, refer to Schmalensee (2003), Evans & Schmalensee (2005a and 2005b), and Bolt & Chakravorti (2008).

as more customers are using it. This makes the purchasing decisions of consumers interdependent. Consumers only want to use Visa if merchants accept it and vice versa. Secondly, the market platform has to sell two distinct products to each end of the market. In normal markets, firms may also sell multiple products, e.g., due to economies of scope. In contrast, companies in two-sided markets must offer both products if they want to remain in business at all. In business environments, this problem is often referred to as getting both sides on board.

Rochet and Tirole (2003b) recognized that, because of these two features, firms have to choose a pricing structure, as well as a price level, to maximize profits. In other words, it is not enough to determine how much to charge in total, but which of the customer groups has to pay how much. Because each product benefits both customer groups, it does not make sense to apply standard microeconomic conditions, such as equalizing the marginal revenue to the marginal costs, on each side. If lawmakers forced companies to price according to costs, they would firstly neglect the positive externalities each customer group exhibits on the other, and secondly, inflict upon them the problem that an allocation of costs to one side is often hardly possible; e.g.: Are costs for the payment infrastructure to execute a payment due to the consumer or the merchant? This leads to the conclusion that a regulation on the basis of costs does not work effectively in two-sided markets.

One part of the price structure decision is whether to charge fixed fees or to charge fees on a per-transaction basis. As Armstrong (2006) notes, platforms may internalize some of the positive cross-externalities by demanding per-transaction fees. In this way, customers from one group have to pay every time a member from the other group actually exhibits a positive externality. In doing so, a fraction of the interaction benefit is accounted for and market inefficiencies are attenuated to. For payment card networks, this means charging merchants every time they benefit from the positive externality they receive from each consumer paying with the card.

Another pricing decision is based on the tendency to multi-home. Customers are multi-homing when they use multiple competing platforms—e.g., merchants often accept both Visa and MasterCard and many consumers obtain several different payment cards. If it is the case that one customer group multi-homes, while the other single-homes, the single-homers make the actual decision on which platform is being used. As a result, competing platforms have to make sure the single-homers choose them over the competition if they want to increase their revenues. Conversely, the multi-homing group draws less attention and ultimately receives fewer benefits.

In the payment card industry, many merchants accepting one card typically also accept at least one other. This is facilitated by the fact that e.g., Visa and MasterCard utilize the same technology, and thus, merchants only incur setup costs once when connecting to the network(s).[42] On the other hand, a significant number of consumers single-home and, even if they do have more than one payment card, they make the final decision on which card to use. This explains the actual pattern in the price structure that can be observed in the market for payment cards, even though transaction costs seem to be lower on the acquirer side than on the issuer side. There is quite some competition for the patronage of consumers, and rather than having to pay for the service, consumers regularly receive benefits in the form of extended credit or frequent flyer miles. This also has to do with the fact, that we typically are confronted with a buyers' market. With this price structure (i.e., consumers rather receiving benefits than being charged a fee) and the consumer's being the critical platform side, issuing appears far more profitable than acquiring.[43] Merchants do not pay large fixed fees, but face significant merchant usage fees, which are a fraction of

---

[42] Often, a network offers credit as well as debit cards. With the so called honor-all-cards rule, some networks oblige merchants to accept the debit as well as the credit cards of the network; even so, the fees typically are different between the two of them. Rochet and Tirole (2008) found that the honor-all-cards rule has a socially balancing effect.

[43] For details, refer to European Commission Competition DG (2006, pp. 62-77).

the value of each card transaction. In the remaining, the merchant usage fee is indicated by an *a*.

One can conclude that in the payment card industry, it is quite well understood why interchange fees are set collectively and to a far lesser degree how they may be set under different objectives. The existing literature has also persuasively laid out the reasons for per-transaction fees on the merchant side, while undercutting prices for consumers. However, the EU Commission Competition found that merchant usage fees vary significantly across merchant sectors (Figure 3-1). Additionally, the U.S. Government Accountability Office (2009) has identified four major factors that influence the price level of the payment card service:

1. The type of card: credit and debit cards, consumer, and commercial cards and rewards payment cards, all are associated with different fees.
2. The merchant category: the merchant's line of business and the competitive environment.
3. Merchant size or transaction volume: Merchants with a larger volume generally have lower rates.
4. Processing mode: whether the cards are PIN-based or have magnetic swipe strips, or whether the transaction is over the internet and translates to a different risk of fraud.

The second category could thus far only be insufficiently explained by the existing literature. As such, we want to add some insights to this point.[44]

---

[44] Simultaneously, and without being aware of each others' research, Wang (2010) presented a model with a fairly similar set-up as the present study.

In this study, we argue that the merchant usage fee depends on the characteristics of the downstream market. We employ a situation of Bertrand price competition in a duopoly of merchants where each merchant sells one product. The products from Merchants 1 and 2 may be either complements or substitutes to reflect a continuum of different competitive situations. The downstream market may also vary in market size and the price sensitivity of consumers.

We are interested in the effects of the characteristics of the downstream market on the merchant usage fee. Accordingly, we try to model other market features as simply as possible. We take as given the price structure that payment card networks choose in reality and do not model any other charges apart from the merchant usage fee. Furthermore, we avoid the discussion about interchange fees by looking at a monopolistic unitary payment network. Unitary networks are such where the acquirer, the issuer, and the payment network are one entity, so that we only have to worry about covering the costs of the whole transaction service. The workings of a unitary payment card network can be seen in Figure 3-2.

While such unitary networks exist, most notably American Express and Discover Card, we want to argue that our results extend, to some degree, to the case of multi-party networks. In such multi-party networks, acquirers face incentive structures similar to the unitary network of our model, because the acquirer simply passes on a fixed fraction of the merchant usage fee to the issuer through the interchange fee. As a result, the acquirer's share is simply reduced by a proportional amount.

So far, most of the governmental regulatory activity has been concerned with multi-party networks. While it is true that Visa and MasterCard are the largest networks and leading banks in their system are among the biggest players in the banking industry, they still face competition from other banks within their systems that have to play by the same rules. Even if the collective setting of interchange fees limits some form of competition, the problem is aggravated in the context of unitary networks. Not only do unitary networks not have to worry about interchange fees and can more easily cross-subsidize between the platform sides, neither do they face competition within their system. As a matter of fact, merchant usage fees of American Express have been higher than those from Visa and MasterCard since the beginning of these multi-party payment card networks. This has been attributed to the role of American Express as a first-mover and the fact that American Express is historically very familiar with the travelling business, which is one of the prime segments of the payment card industry.[45] Again, the downstream market may play a crucial role in this context.

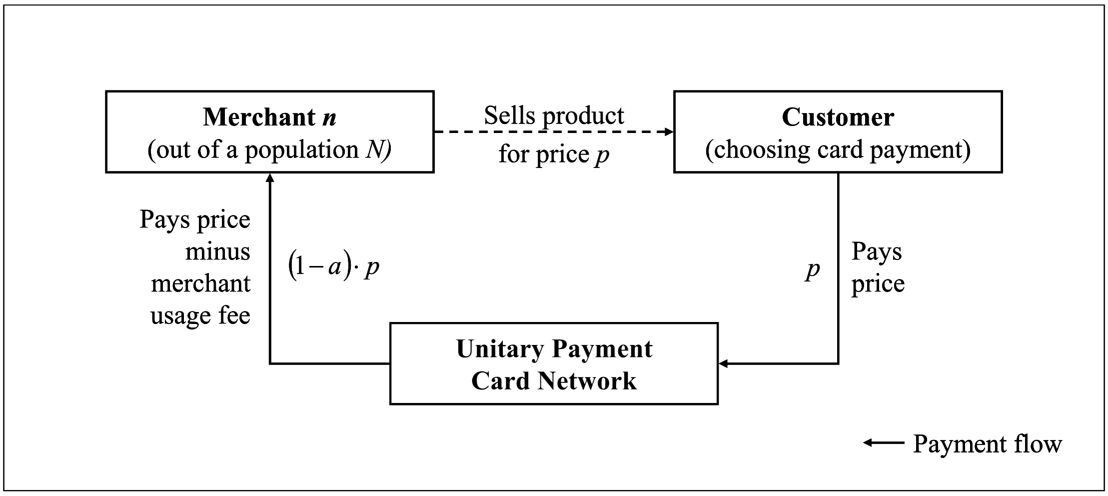

Figure 3-2: Organigram of a Payment Card Network (Cf Gans and King, 2003, p. 4).

In the next section, we present our model starting with the general outset in Section 3.3. We then look at the characteristics of the downstream market that are influencing the merchant usage fee on the consumer side in Section 3.3.1 and the equilibrium and merchant behavior in Section 3.3.1. In Section 3.3.3, we analyze how the monopolistic unitary payment network anticipates the results in the downstream market and adjust the merchant usage fee accordingly. Section 3.4 concludes.

---

45 This might be especially true for business travelers.

3.3 The Model

Our model economy is populated by two identical merchants who have some monopoly power as they engage in Bertrand price competition.[46] Each of the merchants sells one product to a continuum of consumers. Payments are made by using either cash or a payment card. A fixed proportion of consumers will only purchase when they are allowed to use a payment card, while all other consumers prefer to pay cash. We do not take into account any strategic considerations on either side of accepting or using payment cards. Let there be one monopolistic unitary payment card provider that charges a merchant usage fee while letting consumers use the card without any charge. Therefore, we take the price structure adopted by the payment network as given and focus on the determinants of the merchant usage fee.[47]

Furthermore, consider the timing of the game as follows:

1. The payment card network sets a profit maximizing merchant usage fee.

2. Merchants decide whether to accept card payments or not.

---

[46] The assumption of Bertrand price competition enables the analysis of heterogeneous products, which allows for an analysis of product substitutability as, what we find, a downstream market determinant of the merchant usage fee. The intuition of product substitutability can be projected on a continuum which ranges from the products of different merchants being substitutes (e.g., fuel) via no substitutes at all (i.e., the monopolistic merchant) to the products being complements (video game consoles and the respected games).

[47] Our chosen price structure resembles the characteristics of payment cards in many countries where consumers usually only have to pay a small annual fixed fee and frequently enjoy additional benefits of card usage. For example, according to the semi-annual Federal Reserve Survey, most U.S. payment cards do not charge membership fees and none charge on a per-transaction basis for domestic purchases (Board of Governors of the Federal Reserve System, 2010).

3. Merchants maximize profits by setting a price for their product and anticipating the demand behavior of consumers.
4. Consumers make their purchasing decision.

The game is solved by backward induction.

3.3.1 Consumers

In the last stage, we model consumer behavior based on a standard Bertrand price competition model analogous to Singh and Vives (1984) and we also stick to the nomenclature of this section. Accordingly, we have the following demand functions:

$$q_1 = A_1 - B_1 p_1 + \tau \cdot p_2 \tag{3.1}$$

$$\text{and } q_2 = A_2 - B_2 p_2 + \tau \cdot p_1 . \tag{3.2}$$

Hence, A is a measure of the market size, B represents the sensitivity of the demand to own-price changes and τ reflects the sensitivity to changes in the price of a comparable product from a different merchant. Furthermore, we assume without loss of generality that $B \geq \tau$.[48]

We also derive the price elasticity of demand ε:[49]

$$|\varepsilon_{q_1,p_1}| = \frac{B_1 p_1}{A_1 - B_1 p_1 + \tau \cdot p_2} . \tag{3.3}$$

---

48 In other words, the price that a merchant charges has at least as much influence on the purchasing decision as the price of the other merchant.

49 Note that in contrast to Section 2 and 4, in this section, ε indicates the price elasticity of consumer demand and is defined in the standard way as being the absolute value.

It is easily verified that our model is consistent with standard assumptions about demand behavior, as indicated by Lemma 3.1.

Proofs are delegated to the Appendix.

Lemma 3.1 *The more consumers react to the price changes of the first (second) merchant the higher (lower) the price elasticity of demand for the first product, since:*

$$\frac{\partial |\varepsilon|}{\partial B} > 0 \tag{3.4}$$

$$\text{and } \frac{\partial |\varepsilon|}{\partial \tau} < 0\,. \tag{3.5}$$

The first part is straightforward, as own-price sensitivity B and own-price elasticity are virtually synonymous. For the second part, note that with an increasing τ, consumers react more strongly to the prices of the other merchant. In addition, the first merchant sells more of his/her products at a given price level. In other words, the own-price elasticity of consumer demand is decreasing as they react less severely to own-price changes of this merchant, relative to the reaction to the prices of the other merchant.

3.3.2 Merchants

We take as given the matching process between buyers and sellers and assume that merchants maximize their profits by setting prices. Because of symmetry, it suffices without loss of generality to look at only one merchant. The resulting profit function provides:

$$\pi(p_1, p_1) = (1 - \gamma \cdot a) p_1 q_1(p_1, p_2) - c \cdot q_1(p_1, p_2)\,. \tag{3.6}$$

Where γ denotes the fixed proportion of consumers who will only purchase the product if they can effect a card payment. All other consumers will use a different payment method we refer to as cash. If consumers pay by card, the payment network has to pay the merchant usage fee a as a percentage of the transaction volume of each purchase. Accordingly, we have $0 < a < 1$, as well as $0 < \gamma < 1$. Assume that merchants incur constant variable costs c with the sale of every product, but do not have any fixed costs. The function thus defines profits as revenues minus payment card transaction costs and procurement costs.

Given this profit function, merchants face higher costs when consumers use payment cards.[50] For parsimony, and without loss of generality, we only include the merchant usage fee in our model and assume all other costs to be negligible.[51] In addition, consumers have to pay the same price no matter whether they pay by cash or by card.[52]

[50] Similarly, some studies suggest that the costs involved in accepting and managing cash payments are smaller than the costs of accepting payment cards (Chakravorti, 2003; Wang, 2010).

[51] An alternative interpretation would be to look at a as the differential between the merchant usage fee and the cost of accepting and managing cash. This, however, would make the following interpretation of a slightly more complicated without adding much value to the analysis.

[52] Some networks try to contractually oblige merchants not to surcharge for card payments (the so-called no surcharge rule). This has raised interest among competition authorities. Bolt & Chakravorti (2008) claim that such a rule makes consumers and merchants worse off, while banks benefit. However, there is strong empirical evidence that most merchants do not surcharge, even in the absence of a no surcharge rule (for example, refer to Bolt et al., 2010, as well as European Commission Competition DG, 2000, for the Netherlands and for Sweden) aside from some peculiar business settings, such as online sales, where the surcharging of card payments is often found. Even so, surcharging is a rather uncommon practice. As Chakravorti & To (2007) have noticed, when merchants are willing to inflict these higher prices on all consumers, rather than surcharging for card payments, this strengthens any results in which they do accept payment cards. If they could surcharge, they would then certainly be willing to accept card payments.

We now assume, for our standard case, that merchants are not only symmetrical, but identical, such that we have $A_1 = A_2$ and $B_1 = B_2$. Substituting Equation (3.1) into Equation (3.6) yields the profits the unitary network will strive to maximize:

$$\max_{p_1} \left(A - B \cdot p_1 + \tau \cdot p_2\right) p_1 \left(1 - \gamma \cdot a\right) - c\left(A - B \cdot p_1 + \tau \cdot p_2\right). \tag{3.7}$$

Because of merchant symmetry, they will set the same price in equilibrium $p = p_1 = p_2$. In addition, at this stage, we take as given that both merchants accept card payments. Taking the first order condition, then setting prices equal and solving for p^* leads to Lemma 3.2:

Lemma 3.2 *The equilibrium in the downstream market is characterized by:*

$$p^* = \frac{1}{2B - \tau}\left(A + \frac{Bc}{1 - \gamma \cdot a}\right), \tag{3.8}$$

$$q^* = \frac{B}{2B - \tau}\left(A - \frac{(B - \tau)c}{1 - \gamma \cdot a}\right), \tag{3.9}$$

$$\text{and } Q^* = 2q^* = \frac{2B}{2B - \tau}\left(A - \frac{(B - \tau)c}{1 - \gamma \cdot a}\right). \tag{3.10}$$

The first thing to notice is that we have $\frac{\partial p^*}{\partial \gamma} > 0$ and $\frac{\partial q^*}{\partial \gamma} < 0$. All consumers, card paying as well as cash paying, bear the additional costs due to card payments. Hence, with a constant merchant usage fee, the more consumers use their payment cards, and thus inhibit these transaction costs on merchants, the higher the equilibrium price will be and the lower the equilibrium quantity.

At the second stage of the game, merchants decide whether to accept card payments or not. There is a certain proportion of consumers who will only purchase the product if the merchant accepts card payments, thus, the necessary and sufficient condition for card acceptance is that merchants make non-negative profits with respect to these consumers. We denote the threshold merchant usage fee that yields this result as $\bar{a}$. Any fee above this level will make merchants reject card payments, while they would accept it for any fee below the threshold.[53]

In order to calculate the maximum acceptable discount $\bar{a}$, we have to compare the profits a merchant makes when accepting card payments to those when he/she does not accept it. As we can see from Equations (3.8 and 3.9), the equilibrium price and quantity will differ in these cases. We assume that the payment network sells no other products to merchants than the card service. Therefore, there is neither a subsidy of other banking services, nor any other motivation than payment services for merchants to make business with the network. Hence, merchants decide whether to accept the payment service solely based on the level of the merchant fee. The profits are thus given by:

$$\pi_{nc}(p_1, p_2) = (1-\gamma)q^*_{1,\text{nc}}\left(p^*_{1,\text{nc}} - c\right) \tag{3.11}$$

$$\text{and } \pi_c(p_1, p_2) = q^*_{1,\text{nc}}\left(p^*_{1,\text{c}} - a \cdot \gamma \cdot p^*_{1,\text{c}} - c\right), \tag{3.12}$$

where $_c$ indicates card acceptance and $_{nc}$ non-acceptance of payment cards.

The maximum merchant usage fee $\bar{a}$ is then implicitly provided by:

$$\pi_c = q^*_{1,\text{nc}}\left(p^*_{1,\text{c}} - \bar{a} \cdot \gamma \cdot p^*_{1,\text{c}} - c\right) \overset{!}{=} (1-\gamma)q^*_{1,\text{nc}}\left(p^*_{1,\text{nc}} - c\right) = \pi_{nc} \tag{3.13}$$

---

53 With regard to this, it has to be stated that we are ignoring that card acceptance might be a competitive instrument, and thus, overstating merchant resistance. Guthrie & Wright (2007) provide a rationale for $\bar{a}$, since they find that with homogenous merchants regarding the benefits and costs of the card services competing payment card schemes will strive for achieving the optimum merchant usage fee provided merchants still accept cards (Chakravorti, 2003). This will be true for pretty much all situations where the merchant usage fee is lower than the equilibrium merchant usage fee a^*, which will be derived in the remaining.

and the merchants will not accept any fee that is above this threshold, so we have $a \leq \bar{a}$.[54]

3.3.3 The Payment Card Network

Going back to the first stage of the game, the payment card network anticipates the merchant behavior p^* as well as the resulting market equilibrium $Q^*=2q^*$. We assume that the network does not incur any variable costs, but only fixed costs C. The rationale is that large investments are necessary to introduce and operate a payment network, while the actual per-transaction costs are minimal and can be ignored. The network chooses the actual merchant usage fee and makes a take-it-or-leave-it offer to the merchants. In other words, merchants do not have any bargaining power.[55]

---

54 The assumption that consumers indeed have a strict preference for card payments and will not purchase at all if a merchant rejects it appears to be very strong. An alternative modeling setup could include both consumers with a weak preference (who will pay cash if that's the only accepted method) and with a strong preference (who will refrain from purchasing altogether). The result would be a lower threshold for the maximum acceptable merchant usage fee $\bar{a}$. However, once a merchant decides to accept card payments and there are a proportion of consumers paying by card, all the other interpretations would still hold. We thus have chosen to stick with the simpler approach. As long as the maximum threshold is not crossed, the distinction between a weak and a strong preference for card payments is immaterial and does not make a difference. Furthermore, the strict preference can be justified by the fact that merchants experience increased sales after accepting payment cards (Chakravorti, 2003). As payment card transactions are more costly than cash transactions, there would not be any incentive for merchants to accept payment cards if sales did not increase.

55 The assumption that merchants do not have any bargaining power may seem striking at first. And it seems reasonable to argue that large merchants do have some leverage to reduce fees. The European Commission Competition DG (2006) does recognize size as one factor of bargaining power and found that the fee is sometimes only negotiated with larger merchants. However, even large merchants are relatively small compared to the market size of the large payment card networks with transaction volumes of $465 billion of American Express and $1,370 billion of Visa and MasterCard in the U.S. in 2008 (Federal Reserve Bank of Kansas City, 2009). Consequently, for the U.S., the U.S. Government Accountability Office (2009) found that ”representatives from some of the large merchants with whom we spoke [...] reported that their inability to refuse popular cards and network rules (which prevent charging more for credit card than for cash payments or rejecting higher-cost cards) limited their ability to negotiate payment costs.” In addition, in multi-party networks, the largest chunk of the total merchant fee is the

However, merchants do have a maximum willingness-to-pay that the network anticipates and sets the usage fee accordingly, so that we have $a \leq \bar{a}$.

As network profits amount to fee revenues minus fixed costs, the profit function can be written as:

$$\max_a \ \Pi(a) = p^*(a)Q^*(a)\gamma \cdot a - C \tag{3.14}$$

or equivalently, after inserting the equilibrium values:

$$\max_a \ \Pi(a) = \left(\sigma + \frac{\phi}{1-\gamma \cdot a} - \frac{\xi}{(1-\gamma \cdot a)^2} \right) \gamma \cdot a - C\,, \tag{3.15}$$

where: $\sigma = \dfrac{2A^2B}{(2B-\tau)^2}$, $\phi = \dfrac{2AB\tau \cdot c}{(2B-\tau)^2}$, and $\xi = \dfrac{2B^2c^2(B-\tau)}{(2B-\tau)^2}$.

The corresponding first order condition gives the optimal merchant usage fee.

Lemma 3.3 *The optimal merchant usage fee level* a^* *is a function of* (A, B, τ, γ). *It is implicitly defined by:*

$$\left[\sigma + \frac{\phi}{1-\gamma \cdot a^*} - \frac{\xi}{(1-\gamma \cdot a^*)^2} \right] \gamma + \left[\frac{\gamma\phi}{(1-\gamma \cdot a^*)^2} - \frac{2\gamma\xi}{(1-\gamma \cdot a^*)^3} \right] \gamma \cdot a^* = 0\,. \tag{3.16}$$

---

interchange fee set collectively by the card network. Acquirers negotiating with large merchants hence only have a limited scope for negotiations. We thus think it is a supportable assumption to model the merchant usage fee as a take-it-or-leave-it offer of the card network to the merchants. Again, as in the current paper, we are not interested in the complicated interrelationships of acquirers, issuers, or the card network service: we treat these as one entity. For multi-party networks, the incentives for the acquirer are the same as in the unitary network case. While the interchange fee is fixed by the network, the acquirer sets the additional fee optimally, resulting in the final merchant usage fee. A more complicated model including a multi-party network should yield similar results. Nevertheless, our setup is naturally the best fit when looking at unitary payment networks.

The optimal merchant usage fee is implicitly defined by the first order condition of the profit function. The main idea behind this study is that the parameters (A, B, τ, γ) are the driving forces behind the determination of the merchant usage fee. As the characteristics in the downstream market change, the payment network should adjust their pricing strategy accordingly.

Proposition 3.1. *The optimal merchant usage fee is increasing in the market size and the impact of other merchant's prices, but decreasing in the own-price sensitivity of demand, as well as in the fraction of consumers paying by card:*

$$\frac{\partial a^*}{\partial A} > 0, \tag{3.17}$$

$$\frac{\partial a^*}{\partial B} < 0, \tag{3.18}$$

$$\frac{\partial a^*}{\partial \tau} > 0, \tag{3.19}$$

$$\text{and } \frac{\partial a^*}{\partial \gamma} < 0. \tag{3.20}$$

The main interpretation of these results is that with a more competitive market environment in the downstream market, the payment service networks have to lower the merchant usage fee. Essentially, the payment card network obtains a share of the economic rents. If competition in the downstream market is higher and profits of merchants are lower, there is a smaller total fee revenue to be shared. Consequently, the price the network can charge decreases.

A gives the size of the market the merchants penetrate. If merchants serve a larger market in an otherwise unchanged competitive environment, they obtain larger profits. But remember that the payment network is in a strong bargaining position and can make take-it-or-leave-it offers. It will consequently want to earn a major share of these profits and will charge a higher price for its services. The caveat is that, with increasing size, merchants are likely to have a larger bargaining power. Since we do not look at bargaining power issues in our model, we do not find this to likely be a countervailing effect. However, we still think that the result is applicable to a wide range of markets, where merchant bargaining power is not a major determining factor, if at all.

On the other hand, there is the price sensitivity of consumer demand. For *B*, this refers to the own price the merchant sets while τ measures a sensitivity to the price of the other merchant. Regarding the former, when the own-price sensitivity of demand is higher, less of the fee can be passed on. As a result, the optimal fee level decreases. The opposite logic applies for τ. If τ increases, the higher prices of the other merchant will have a stronger impact on the first merchant's sales. This also translates into a higher sales quantity of the first merchant at given prices. It therefore increases this merchant's minimum sales quantity, no matter his own price, and provides each merchant with more monopoly power. The overall result is that a higher merchant usage fee can be passed on.

The last statement claims that the optimal merchant usage fee is decreasing in the fraction of consumers who prefer to pay by card. We know that without surcharging, the merchant usage fee is passed on to all consumers, whether they are paying with cash or a card. We have already established that the equilibrium market price is increasing in γ as well as in *a* ($\frac{\partial p^*}{\partial \gamma} > 0$, $\frac{\partial p^*}{\partial a} > 0$), while the equilibrium quantity is decreasing in γ, as well as in *a* ($\frac{\partial q^*}{\partial \gamma} < 0$, $\frac{\partial q^*}{\partial a} < 0$). The rationale is that merchants are increasing product pricing when facing the higher costs of more consumers paying by card. As a result, the equilibrium quantity will go down. However, the payment card network anticipates this behavior. As can be seen in Equation (3.15), the network's profits are given by p^*, q^*, a^*, and γ. In addition, if γ increases, p^* is increased by the merchant and q^* decreases in the market. We find that the payment card network's reaction is to decrease *a*, as long as it does not cross the threshold $a^* \leq \bar{a}$.

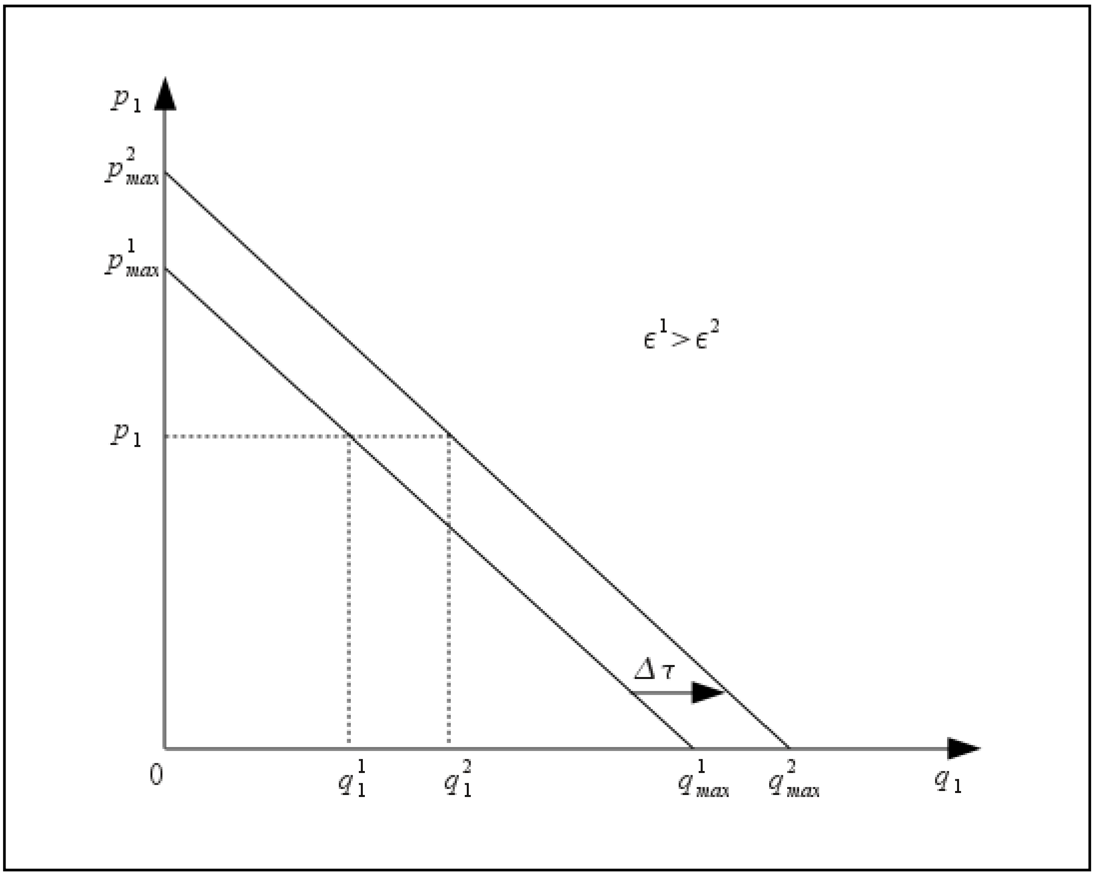

Figure 3-3: The Effects of an Increase in τ ; at a Given Price, Sales Quantity Increases while the Price Elasticity Decreases.

Another way to look at the optimal relationship of Proposition 3.1 is through the price elasticity of demand. In this way, we obtain:

Proposition 3.2. *The merchant usage fee is decreasing in the price elasticity of demand, since:*

$$\frac{\partial a^*}{\partial |\varepsilon|} < 0 . \tag{3.21}$$

From Lemma 3.1, we know that the elasticity is decreasing in τ ($\frac{\partial|\varepsilon|}{\partial\tau}<0$). As the optimal merchant usage fee is increasing in τ, it has to be decreasing in the elasticity $\frac{\partial a^*}{\partial|\varepsilon|}<0$. The interpretation of this result is analogous to $\frac{\partial a^*}{\partial B}$. The own-price sensitivity of consumers is equivalent to the price elasticity. This relationship is illustrated in Figure 3-3.

Regarding the foregoing demand function (Eq. 3.1 and 3.2), Singh and Vives (1984) derive a measure for product substitutability from the corresponding inverse demand function, in the remaining indicated by *θ*. *"The goods are substitutes, independent* [i.e., the monopolistic merchant]*, or complements, according to whether* ..[*θ*]$\gtreqless$*0."* (Singh and Vives, 1984, p. 548.) Since product substitutability is derived from the (inverse) demand function (Eq. 3.1 and Eq. 3.2), *θ* is a function of (A,B,τ). By comparative statics regarding the equilibrium merchant usage fee and product substitutability, we derive the next Proposition:

> **Proposition 3.3.** *The optimal merchant usage fee is decreasing in the product substitutability, since:*
>
> $$\frac{\partial a^*}{\partial\theta}<0. \tag{3.22}$$

For details of how *θ* is composed and the proof, see the Appendix. Note, that this finding is consistent with our other results in this section, since a higher product substitutability increases the price elasticity of demand.

3.4 Discussion and Conclusions

We have shown that the merchant usage fee depends on the factors of the downstream market. While Langlet (2009) had established this for the case of

Cournot quantity competition among merchants, we have extended his approach to Bertrand price competition and can thus also identify effects resulting from heterogeneous products. We find the optimal merchant usage fee is increasing in the downstream market size and decreasing in the products substitutability, as well as the fraction of consumers strictly preferring card payment. The rationale is that the payment card network is extracting part of the rent from the merchants. As a result, the merchant usage fee is an increasing function of the merchant's rent.

A natural way to look at the competition between the merchants would be in terms of a Hotelling setup, where the two merchants are located at each end of a continuum and consumers are uniformly distributed along the spectrum. Every consumer has to pay the price of the product and transport costs in order to overcome the distance and make a purchase at each of the merchants. An extension of this model could argue along these lines and investigate size and distribution effects. Another complication arises if we look at a market with more than two merchants. When there are more than two merchants, the monopoly power of a single merchant decreases and thus, each merchant's profits decreases, as the market resembles the case of perfect competition more closely. Consequently, the merchant usage fee should be lower in such a setup.

Our model also provides a rationale for lower merchant usage fees when more consumers are using payment cards. The effects of an increased number of consumers paying by card are balanced by a decreased merchant usage fee. Chakravorti and To (2007) find the opposite result. In their model, the more people use credit cards, the higher the fee the payment network can charge. This is because, in their model, the driving force is the easier access to credit that credit cards enable, which then gives liquidity constrained consumers the ability to consume in an earlier period. Thus, merchants increase their sales in the present at the expense of other merchants' sales in the future. However, if all merchants accept credit cards, the result is a prisoner's dilemma, where all merchants face the higher costs of credit cards, while no actual sales increase takes place. The U.S. Government

Accountability Office (2009) reports that merchants now claim that the increased sales through card acceptance cannot make up for the rising fees they have to pay. As card acceptance in the U.S. is very widespread, such a prisoner's dilemma situation may actually have occurred.

In contrast, we do not include a liquidity constraint and a multiple period perspective in our model. Alternatively, another explanation seems plausible. We have assumed that the payment networks make a take-it-or-leave-it offer to the merchants. Nevertheless, it is reasonable to assume that merchants do have some bargaining power over the fee level. The more consumers insist on card payments and, consequently, the higher the positive externality that payment card users exhibit on merchants, the lesser the merchant is able to refuse accepting card payments. Put differently, the merchant's bargaining power is decreasing in the fraction of payment card users. We can therefore identify two countervailing effects, one leading to higher merchant fees and one to lower merchant fees as more consumers prefer card payments. As such, it remains an empirical question which effect dominates.

The bargaining power of merchants can also be thought of as a function of the merchant's size. Even though size does not seem to be the overarching explanatory factor for the merchant usage fee, it does seem to be an influential one. This influence might also be a worthwhile object of a different article. The payment card network's objective function in our framework is optimized by setting the merchant fee. If this fee is reduced because a large merchant has more bargaining power, the network may devote its efforts to optimizing the opposite market side: increasing the number of card paying consumers that will then increase fee revenues and, over time, reduce the merchant's bargaining power. Such a framework may explain developments in the payment card industry over time.

Another reason for a higher bargaining power of merchants may be competition among different payment card networks. Because of such competition, a single network may be unable to make take-it-or-leave-it offers and will have to engage in direct negotiations with merchants. The actual market reality will be influenced by at least two factors: the degree of multi-homing of card users and the competitive environment of the payment card network. If consumers single-home and only have one payment card available which they want to use for their purchases, card competition only has a limited effect on the merchant's ability to refuse single cards. Regarding the competitive environment for the payment card network, another model could deviate from a monopolistic unitary payment network. In reality, we do see competition among different networks and in multi-party networks among different acquirers of one network. According to standard economic intuition, a first guess would be that network competition should reduce the equilibrium merchant usage fees. In contrast, the multi-party payment card networks claim that competition might even lead to an increase in the payment card network price; they argue that further network competition will lead to higher interchange fees, since the networks will have to compete for issuers. In this regard, analyzing the dynamics between the downstream and the upstream payment market promises further results.

Lastly, our model provides a rationale for the vast differences in merchant usage fees between merchant sectors. The intuition behind this finding is that the payment networks extract a part of the economic rent of the merchants. Thus, the margins of the merchant and the payment card network add up, which is similar to the double marginalization of two subsequent monopolies, which raises the suspicion of a potential market failure. However, in this regard, it remains unclear what might help to constitute socially optimal payment card prices or, on the other hand, a market failure. In order to evaluate the question of social optimality, a model of overall welfare would be necessary; we have not constructed this in this particular study. Such a consideration might be a worthwhile extension of this study.

4 Multi-Party Network with Quantity Competition among Merchants[56]

with Clemens Buchen

In this study, a multi-party payment card network model is constructed to explore the impact of downstream market characteristics (i.e., the market where merchants and consumers interact) on the merchant and the interchange fee, thereby extending the literature's observations of payment systems. The focus of this analysis is on the impact of the price elasticity of demand, the relative frequency of card usage, and the competitive condition of merchants. To accomplish this, two scenarios are analyzed: a network with all-purpose banks that serves both sides of the payment market in an equal manner (i.e., acquiring and issuing cards) and specialist banks that exclusively serve as either issuers or acquirers. This section yields two significant results. Firstly, similar to related studies of unitary payment card networks, we find the merchant and the interchange fee decreasing in the consumer price elasticity and the relative frequency of card usage. Second, regulating the interchange fee and the merchant fee respectively is found particularly useful since, in markets with less competition where consumers already face inefficiently high prices, the payment card fees will also tend to be higher.

[56] For full reference, refer to Buchen & Langlet (2010).

4.1 Introduction

The first significant anti-trust lawsuit against the assumed abusive pricing behavior of a payment card network, the U.S. case of NaBanco against Visa in 1979[57], led to extensive discussions among economists, policy-makers, and industry experts in relation to what constitutes efficient payment card pricing. Payment cards include debit, credit and charge cards, enabling consumers to pay for purchases without cash. Moreover, from 1979 onwards, the importance of payment cards as a payment method has grown tremendously; card usage in the U.S. grew from about 3% in 1986 to 25% in 2000 (i.e., the ratio of payments conducted by cards relative to the payment volume over cash; Evans & Schmalensee, 2005b, p. 3).[58] Consequently, a significant amount of research has been conducted in the field of payment card pricing, and more particularly, on the constitution of efficient payment card pricing. Still, there remains a gap in the understanding of the determinants of socially and privately optimized payment network pricing.

In 2006, the European Commission Competition Directive Generale[59] (2006, pp. 40-51) observed that payment card fees varied significantly among countries and merchant sectors, although there is a high correlation when comparing the network fees of different systems.[60] Thus far, neither the European Commission, nor the theoretical literature, has sufficiently explained this observation.[61] For example,

---

57 National Bancard Corp. (NaBanco) vs. Visa U.S.A., Inc., 596 F. Supp. 1231, 1265 (S.D. Fla. 1984), *aff'd.*, 779 F.2d 592 (11th Cir. 1986), *cert. denied*, 479 U.S. 923 (1986). For details, refer to Evans & Schmalensee (2005a).

58 Evans & Schmalensee (2005b) refer to the original sources: HSN Consultants (1987, 1991, 1992, 2001a, and 2001b), and the U.S. Department of Commerce (2003).

59 From here on referred to as European Commission of Competition DG.

60 Blending is a widespread phenomenon of payment card networks, among others, an observation of European Commission of Competition DG (2006, pp. 64-65).

61 Simultaneously, and without being aware of each others' research, Wang (2010) presented a model with a fairly similar set-up as the present study.

Armstrong (2006) analyzes competition in two-sided markets, revealing three key determinants of platform pricing: the level of cross-group network externalities, the mix of fixed and per-transaction prices, and the level of multi-homing among participants. On the other hand, Section 2 analyzes the determination of payment card prices of a unitary network and finds three significant determinants of payment card usage fees that help to explain the European Commission Competition DG observation: inverse consumer price elasticity, the relative frequency of card usage, and the competitive condition of merchants.

In doing so, in Section 2, a model of a unitary payment card network is constructed with no surcharging, which includes the influence of the payment card network fees onto consumer demand and also allows for competition among merchants. Moreover, by assuming that the payment card system and the merchants engage in Cournot quantity competition[62], equilibrium product prices and sales quantities are derived, as well as an equilibrium merchant usage fee. In doing so, fee regulation is found to be particularly efficient, since it reduces merchant usage fees, especially for less competitive markets where consumers already face inefficiently high prices for the products and services they purchase. Even so, unitary payment networks make up the smaller segment of the payment card market; e.g., Visa and MasterCard, the two big multi-party payment card networks, alone account for about 80% of the U.S. credit card market (Wang, 2004, p. 86). Thus, to test the generality of the findings of Section 2, the current analysis extends the model in Section 2 to the case of a multi-party payment card network with merchants under Cournot quantity competition. As with the unitary network, the aim of this study is to explore the determinants of merchant and interchange fees. The focus is placed on the impact of price elasticity

---

[62] The assumption of Cournot quantity competition allows for an isoelastic demand function which enables a focused analysis of the relationship between the price elasticity of consumer demand and, first and foremost, the merchant usage fee and the interchange fee.

of demand, the relative frequency of card usage, and the competitive condition of merchants. As a result, the model is designed as a two-stage game. In the first stage, the issuers and acquirers negotiate an interchange fee. In the second stage, the acquirers and merchants negotiate the merchant usage fee.

In this section, two scenarios of a multi-party payment network are analyzed. In the first scenario, all-purpose banks both issue and acquire cards. In the second, specialist banks are exclusively either issuers or acquirers. In relation to all-purpose member banks, for each of these banks, the total of paid and received interchange fees will be zero. Hence, with merchants having no bargaining power, this outcome will virtually resemble the unitary network outcome. In contrast, the second case of distinction between issuing and acquiring is characterized by the double marginalization of issuers and acquirers. In reality, all-purpose banks and specialist banks are both examples of actual business models in the industry. Evans and Schmalensee (2005a, pp. 16-18) list the ten largest issuers and acquirers in the U.S. market in 2002, including the unitary networks American Express and Discover (Table 1-1). Among the eight issuers associated with multi-party payment card networks, there are mere two examples of the same company issuing and acquiring to a similar extent; thus, JP Morgan Chase and Bank of America are examples of large all-purpose banks. On the other hand, there are firms that specialize in either acquiring or issuing.

This section provides a generalizability of the findings of Section 2. The determinants of the interchange fee and the merchant usage fee rate of multi-party payment card networks are found to be the same as the determinants of the merchant fee for unitary networks. Furthermore, in this study, interchange fee regulation is, for the most part, found to be particularly useful in the case of inefficiently high merchant usage fees. This is so, since for the first scenario banks negotiate profits from the acquiring side to the issuing side of the network to undermine the bargaining position of the merchants.[63] And in the second scenario, there is a double marginalization of issuers and acquirers in some bargaining cases.

The literature has made many contributions regarding the pricing of payment networks. Schmalensee (2005b) and Bolt and Chakravorti (2008) discussed a valuable overview of the literature on payment card networks. Much of this literature focuses on the private and social optimization of interchange fees, therefore analyzing the costs and benefits of card services. In general, there is still little agreement on how to achieve efficient payment card pricing (Bolt, & Chakravorti, 2008, p. 15).

Baxter (1983) was the first to approach the pricing of multi-party payment card networks. In his model, Baxter assumes perfect competition among merchants and the member banks of the network, as well as homogenous consumers. Baxter explains that, in the social optimum, the total marginal costs of the payment service provision would match the total marginal benefits of merchants and consumers, which would lead to their willingness to pay for the service. Thus, Baxter argues that for an efficient payment service provision, the costs of the payment service should be reallocated between the issuers and acquirers, hence, making an interchange fee necessary. Even so, with the assumption of the zero profits of the member banks, this model does not allow for an analysis of the optimal interchange fee. Consequently, extensions of Baxter's model relax this assumption of perfect competition among banks and merchants.

Schmalensee (2002) presents one such extension by providing market power to banks, but not to merchants. His results support Baxter's findings that interchange

---

[63] For details, refer to Wright (2004).

fees might be necessary to balance the revenues of acquirers and issuers, according to the allocation of the costs. Although the results of Schmalensee illustrate that the privately optimized interchange fee of a payment network might bring about socially optimal network pricing, such evidence requires a number of strong assumptions (e.g., one issuer and one acquirer, no fixed costs, and linear demand curves).

Carlton and Frankel (1995a, 1995b; cp. Evans, & Schmalensee, 1995) find that in a perfectly competitive system with costless surcharging, the fee structure of the payment card networks would be irrelevant for network efficiency. Gans and King (2003), in turn, find that costless surcharging alone would enable merchants and consumers to negotiate away the fee structure imposed upon them, thus causing the fee structure to be irrelevant. Even so, for the most part, surcharging is not costless, even when a payment network does not impose the so-called ‘no surcharge rule’ on merchants (Bolt, Jonker, & van Renselaar, 2008; Frankel, 1998; IMA Market Development AB, 2000).

Rochet and Tirole (2002) extend the existing literature by considering strategic interactions between merchants and consumers. They assume that issuers have market power and that there is perfect competition among acquirers. This leads acquirers to simply pass any interchange fee onto the merchants. Another assumption in the model of Rochet and Tirole is that changes in network prices do not influence consumer demand, in this way there is a fixed volume of payment transactions. Hence, Rochet and Tirole reveal that, depending on the issuer’s profit margin and the consumer’s surplus from card payments, the privately optimized interchange fee will be equal to, or higher than, the socially optimal interchange fee.

By considering different industries with merchants expecting a variety of benefits from card services, Wright (2004) extends the work of Rochet and Tirole (2002). Wright’s approach better represents merchant behavior, since a change in the merchant fee will affect the number of merchants accepting cards on a continual basis, rather than the binary case of Rochet and Tirole, where all merchants either accept cards or do not. Wright also assumes imperfect competition among issuers and acquirers. As a result, in contrast to Rochet and Tirole (2002), Wright finds that the privately optimized interchange fee could be higher, or lower, than the socially optimal interchange fee, since there is a trade-off between consumer surplus and merchant acceptance.

The present study diverges from other approaches in the literature. Wang (2010), in a similar set-up to ours, assumes perfect competition among merchants on the downstream market. In contrast to that, we allow for different degrees of competition. Similarly, Wang considers acquirers to be under perfect competition, thus, acquirers play no role in the study, other than the fact that they simply pass the merchant fee onto the issuers (in the form of the interchange fee). For that reason Wang's approach simultaneously applies to multi-party and unitary payment card networks. In contrast, we model acquirers not under perfect competition, thus our below analysis concerns multi-party payment card networks more realistically. Even so, we find that unitary networks are a subclass of a multi-party payment card network. Finally, our model incorporates less complexity than Wang's approach.

This section is organized as follows. In the next subsection, the model of a multi-party payment network will be constructed. The last subsection will discuss the findings from the model, conclude this essay, and provide recommendations for future work.

4.2 The Model

In Section 2, payment card pricing and the determination of merchant fees for a unitary network are explored, including corporations such as American Express and Discover Card. We extend this by analyzing a multi-party payment card network. A multi-party payment network, in the standard case, involves five parties per transaction; a merchant, a consumer, the payment network association (e.g., Visa or MasterCard), the issuing bank and the acquiring bank. Figure 4-1 illustrates a stylized overview of the workings of a multi-party network. The consumer receives a payment card from the issuing bank, which collects price *p* if the consumer makes a purchase from a merchant. The acquiring bank of the merchant collects the payment from the issuer, but is charged the interchange fee $f \cdot p$ from the issuer. The acquirer then reimburses the merchant, who receives the product price *p,* while the merchant usage fee $a \cdot p$ is deducted from the payment. In some cases, the issuing and the acquiring institution will be the same bank.

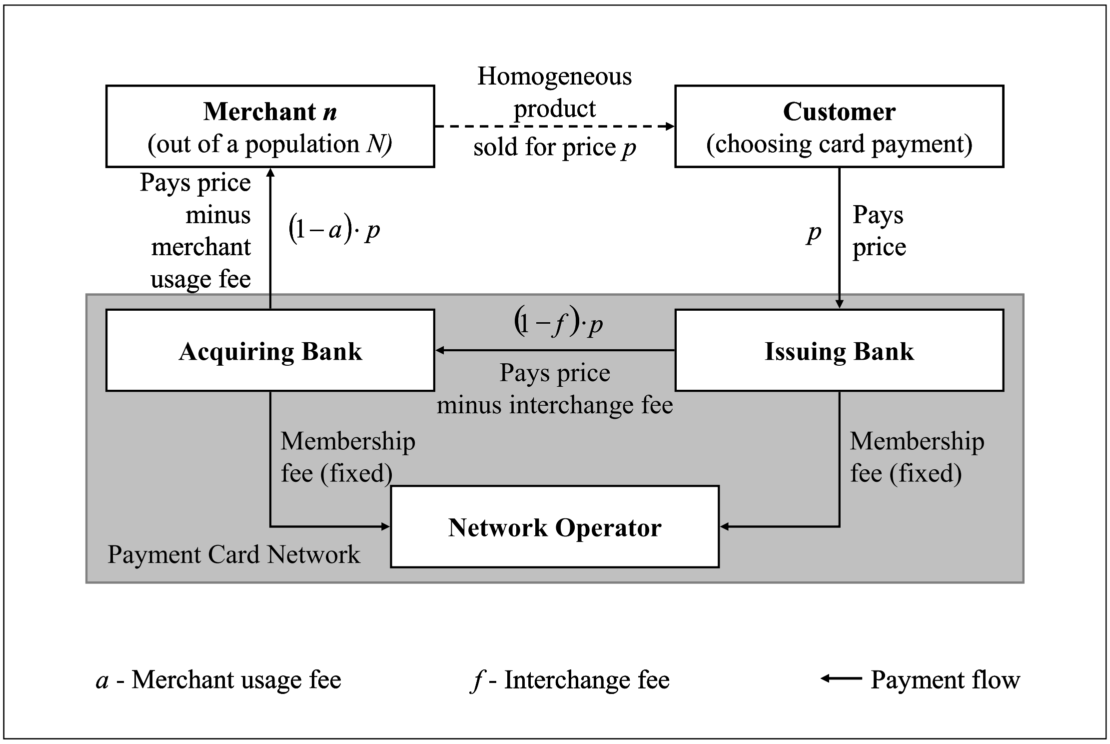

Figure 4-1: Payments in a Multi-Party Payment Card Network (Cf. Gans & King, 2003, p. 4).

We model the behavior of the actors in a dynamic game setting. The timing of the game is as follows: first, the issuing banks set the interchange fee rate *f.* Then, the acquiring banks set the merchant fee rate *a*.[64] In this study, we assume the merchant has no bargaining power, i.e., the acquirers offer the merchant usage fee as a take-it-or-leave-it offer. In the last step, merchants set the quantity q based on the merchant fee; corresponding to common economic theory, the product price p corresponds. In general, the game will be solved by backward induction.

In this general setting, we will distinguish between two types of banks. On the one hand, we analyze all-purpose banks, which issue credit cards and acquire payments. On the other hand, we look at specialist banks, which exclusively function as either issuers or acquirers.

As will become clear, the interchange fee is particularly relevant in the case of specialist banks. We will model the setting of the interchange fee as the outcome of a Nash bargaining game. In this context, we will consider two scenarios of bargaining. The first scenario involves sequential bargaining, in which the issuer and the acquirer first negotiate over f, and then the acquirer sets a. The second scenario models the bargaining as a simultaneous game, in which the issuer and the acquirer simultaneously negotiate over the merchant fee a and the interchange fee f. For parsimony, but without of loss of generality, we consider merchants with no bargaining power.

4.2.1 Consumers and Merchants

A homogenous product is supplied by N identical merchants, indexed by n at market price p. Let q_n indicate the quantity of products sold by merchant n and q_{-n} be the

---

[64] Similar to Wang (2010), by assuming proportional fees instead of fixed per-transaction fees, our model differs from previous models. In doing so, we model the reality of pricing controversies of proportional fees. Even so, the findings of this study remain stable when considering per-transaction fees (refer to Shy and Wang, 2008).

quantity sold by all other merchants. We assume an isoelastic inverse demand function with the absolute value of the elasticity given by:[65]

$$p(q_n, q_{-n}) = Q^{-\varepsilon} = (q_n + q_{-n})^{-\varepsilon} \text{ with } 0 < \varepsilon < 1 . \quad (4.1)$$

Merchants incur variable costs, c, and no fixed costs. Merchants may offer two payment methods to consumers: payment by cash or payment by card. There is a fixed portion of consumers, γ, exclusively choosing to pay by card, whereas the fraction of consumers (1-γ) prefers to pay with cash.[66] The member banks of the multi-party payment card network operate as card issuers and/or acquirers. Issuers provide payment cards to consumers. On the other hand, acquirers connect merchants to the card network, consequently enabling merchants to claim receipts for card payments. These network member banks do not impose any transaction fees upon card holders, the consumers, instead request a merchant usage fee. In particular, acquiring banks charge merchants a certain percentage of the transaction volume referred to as the *merchant usage fee a*. Neither merchants nor consumers incur extra benefits from sales paid for by card or extra costs of card transactions other than the merchant usage fee. The network association does not influence the level of the merchant usage fees, other than obliging acquirers, to impose a no-surcharge rule upon merchants. Thus, merchants do not surcharge consumers who pay by card.

---

65 Note that in contrast to Section 3, here ε indicates the inverse price elasticity of consumer demand.

66 This assumption might seem strict at first. Even so, relaxing this assumption does not change the findings of this study, e.g., when consumers hadn't such a strong preference regarding the payment instrument, i.e., some consumers who prefer cards, but might still pay with cash when card payments were no option. Wang (2010) relaxed this assumption by introducing two types of consumers: cash users, who never pay by card, and card holders, who might or might not pay by card. He derives a threshold price for which the latter group also exclusively pays by cash. In contrast, our analysis is more simplified. Our assumption appears innocuous, because the introduction of a preference of card payment ultimately only alters the threshold merchant usage fee at which the transaction still goes through. For further details, refer to Footnote 55.

The maximization problem of the merchant can thus be written as:

$$\max_{q_n} \pi_n(q_n, q_{-n}) = q_n p(q_n, q_{-n}) - q_n p(q_n, q_{-n})\gamma a - q_n c, \tag{4.2}$$

with $c > 0$ and $0 < \gamma < 1$, as well as $0 < a < 1$ and $0 < \varepsilon < 1$.

We assume Cournot quantity competition[67]; hence, merchant n maximizes profits over quantity. As a result, assuming merchants to be identical, quantity and prices in the market are given by (for details, refer to Section 2.2.1):

$$Q^* = \sum_{n=1}^{N} q_n = \left(\frac{\left(1 - \frac{\varepsilon}{N}\right)\left(1 - \gamma a\right)}{c} \right)^{\frac{1}{\varepsilon}}, \tag{4.3}$$

$$\text{and } p^* = \frac{c}{\left(1 - \frac{\varepsilon}{N}\right)\left(1 - \gamma a\right)}. \tag{4.4}$$

The maximum merchant fee $\overline{a}$, the highest fee a merchant will accept, must now be calculated. In order to accept a fee, the profits of the merchant from a card transaction must at least be zero:

$$\pi_n^{card} = q_n p(q_n, q_{-n}) - q_n p(q_n, q_{-n}) a - q_n c = 0. \tag{4.5}$$

---

67 With Cournot quantity competition, a merchant maximizes profits over quantity. In doing so, the merchant anticipates the behavior of all other merchants. We assume Cournot quantity competition, since one focus of the determinant is the price elasticity of consumer demand. Thus, it is especially appealing to consider isoelastic demand, which Cournot quantity competition allows for.

Consider solving Eq. (4.5) for *a* having substituted the Cournot quantity from Eq. (4.3) that determines the cut-off merchant usage fee as:[68]

$$\bar{a} = \frac{\varepsilon}{N - N\gamma + \varepsilon\gamma}. \tag{4.6}$$

Later in the model, acquirers propose the merchant usage fee *a* in the manner of a take-it-or-leave-it offer. With the take-it-or-leave-it offer, merchants will deny card payments if $a > \bar{a}$, whereas for $a \le \bar{a}$, merchants make non-negative profits from sales paid for by card. Hence, Eq. (4.6) determines the upper boundary for the merchant usage fee.

4.2.2 The Banks

Having established the outcome in the downstream commodity market, we now turn to the second step of determining the merchant usage fee *a* and the interchange fee *f*. This will involve two scenarios. In the first scenario, we model all-purpose banks, which act both as acquirers of cards and issuers. As will become clear, in this scenario, the merchant usage fee is independent of *f,* because each bank pays the same amount of interchange fees as costs as the bank receives as revenue.[69] The

---

[68] This condition brings about an approximation for the cut-off usage fee. For parsimony, this condition implies the assumptions of cash and card paying consumers to be equally distributed along the continuum of the sales quantity *q*, thus, the cash and card paid segments can be viewed as separate markets. More importantly, let the revenue of cash paid sales remain constant whether the merchant accepts cards or not. Even though the latter is a simplification of reality, this assumption does not conflict with the generality of the model. In contrast to this approximation, in Section 2, the maximum merchant usage fee is thoroughly and implicitly derived.

[69] This outcome requires the assumption of no bargaining power of merchants. In contrast, with e.g., Nash-bargaining over the merchant fee, the interchange fee will serve as a lower bound to the Nash-equilibrium merchant fee. In fact, without strategic considerations, the acquirers will not accept a merchant fee lower than the interchange fee since the acquirer would otherwise be making a negative profit margin from transactions where another issuer is involved. For details, refer to Section 4.3.

second scenario assumes banks to be specialists, or, in other words, to be either an acquirer or an issuer.

4.2.2.1 All-purpose banks

Suppose a multi-party network consists of two all-purpose banks, bank *B1* and bank *B2,* which both issue cards to consumers and act as acquirers for merchants.[70] Other than the interchange fee, which the acquiring banks pay to the issuers, the network banks incur no variable costs and only fixed operational costs C_B[71]. In regard to a random transaction, the probability *p(B1,B2)* of bank *B1* being the issuer and bank *B2* being the acquirer is identical to the probability *p(B2,B1)* of bank *B2* being the issuer and bank *B1* being the acquirer. Consequently, each of the banks' total of all received and all paid interchange fees will be zero. The average merchant usage fee of the payment market with two all-purpose banks, a_{2B}, is the total of the banks' market fee weighted by the banks' market share s_{Bi} with $i = \{1,2\}$:

$$a_{2B} = s_{B1}a_{B1} + s_{B2}a_{B2} \text{ with } s_{B1} + s_{B2} = 1 . \tag{4.7}$$

Simply put, in this scenario, the interchange fee can be disregarded, because it occurs for each bank both as revenue and profits and cancels out. In this case, we assume banks sell no other products to merchants than the payment card services. Therefore, there is neither a subsidy of banking services, nor any other motivation than payment services, for merchants to do business with banks. Hence, merchants only decide whether to buy the card service at a certain bank based on the level of the merchant usage fee. Let there be a high price transparency on the payment market, thus there is

---

70 Other oligopoly cases can be derived from this simple duopoly case.

71 For parsimony, let the fixed costs C_A include network membership fees

perfect blending[72] regarding the merchant usage fee of bank *B1* and bank *B2*, such that:

$$a_{2B} = a_{B1} = a_{B2}\,. \tag{4.8}$$

The representative bank 1 maximizes profits over the merchant usage fee *a*, so that the maximization problem can be formulated as:

$$\max_{a_{B1}} \Pi_{B1}(a_{B1}, a_{B2}) = p(a_{B1}, a_{B2})Q(a_{B1}, a_{B2})s_{B1}a_{B1} - C_{B1}\,. \tag{4.9}$$

Finding the first-order condition and solving for *a* results in the merchant usage fee in the case of two all-purpose banks:

$$\hat{a}_{2B} = \frac{\varepsilon}{\gamma}\frac{2}{1+\varepsilon}\,. \tag{4.10}$$

Remember that there is an upper limit on the merchant fee given by Eq. (4.6). Hence, the merchant usage fee is given by Eq. (4.10), but cannot exceed $\bar{a}$ (Eq. 4.6).[73] This leads to the first Proposition.

Proposition 4.1. *In the case of all-purpose duopoly banks, the equilibrium merchant usage fee will be:*

$$a^*_{2B} = \min\left\{\frac{\varepsilon}{\gamma}\frac{2}{1+\varepsilon}; \frac{\varepsilon}{N - N\gamma + \varepsilon\gamma}\right\}. \tag{4.11}$$

---

[72] Keep in mind that blending is a common strategy of payment card networks.

[73] We find that the level of competition among merchants determines the merchant usage fee when merchants are at the edge of denying payment cards. In addition, we find that in a less competitive environment, the competitive condition of merchants (i.e., monopoly power) does not impact the merchant usage fee. Even so, this finding might depend on the no bargaining power of the merchant(s) in regard to payment network prices, since the competitive condition of the merchant will most likely influence his/her bargaining power.

Proofs are delegated to the Appendix.

Hence, the bank will choose to charge $\hat{a}$ from Eq. 4.10) if $\hat{a}<\bar{a}$ from Eq. (4.6) and $\bar{a}$ otherwise. This is depicted in Figure 4-2, which plots $\bar{a}$ decreasing in the number of merchants *N* and $\hat{a}$ as a constant, i.e., independent of *N*. At the intersection of the two curves, we find the number of merchants that allow the banks to still make maximum profits (i.e., obtain the optimal fee) while the merchants' profits are zero:

$$N_{2B}^{*} = \frac{\gamma}{1-\gamma}\frac{1-\varepsilon}{2}. \tag{4.12}$$

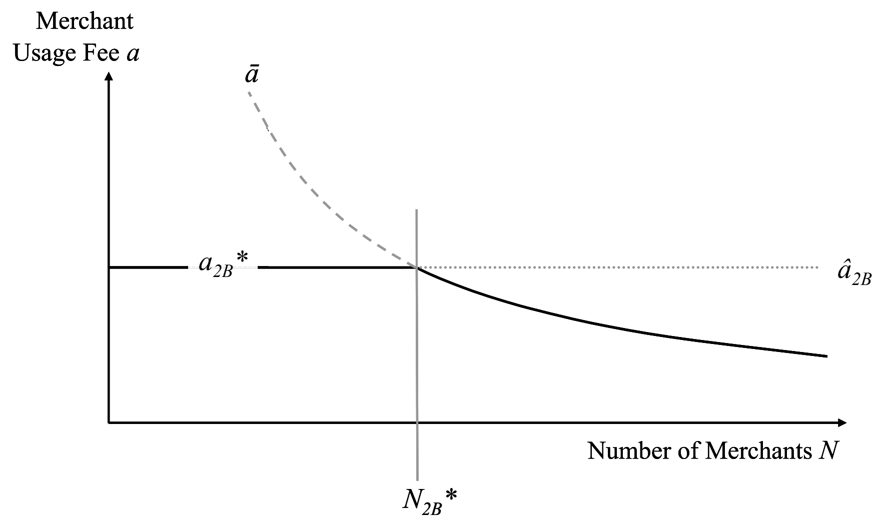

Figure 4-2: The All-Purpose Bank's Merchant Fee.

While observing the U.S. payment card market, the U.S. Government Accountability Office (2009, pp. 18-21) concludes that payment card prices might increase when there is more competition among the networks. This increase relates to the fact that the multi-party payment card networks, when having to compete for issuers, might become motivated to raise interchange fees. Therefore, increasing interchange fees would simultaneously lead to higher merchant fees.

In this section, we find another rationale for an increase of the payment card fees in network competition. For a moment, suppose the merchant usage fee is a value smaller than the cut-off value from Eq. (4.6) (i.e., $\hat{a}<\bar{a}$) and consider merchants with no bargaining power. In this case, it can be easily shown that the outcome merchant fee is identical to the networks with all-purpose banks and unitary payment card networks. As such, the above effect becomes relevant when comparing the equilibrium merchant usage fee with a monopolistic unitary network $\hat{a}_M = \frac{\varepsilon}{\gamma}$ (Eq. 2-19) and the equilibrium merchant usage fee with duopolistic unitary networks $\hat{a}_{2D} = \frac{\varepsilon}{\gamma} \cdot \frac{2}{1+\varepsilon}$ (Eq. 4-10), i.e., $\hat{a}_M < \hat{a}_{2D}$ for all $0<\varepsilon<1$. Hence, for this range of values, the (inverse) demand elasticities adding unitary networks increases the merchant usage fee.

4.2.2.2 Specialist Banks

In this section, the case of specialist banks will be considered. The multi-payment network consists of banks that only act as acquirers of card payments or only issue cards to consumers. While the interchange fee f could be ignored in the forgoing case of all-purpose banks, here it will be a determinant of the merchant usage fee. The single acquiring bank in this model sets a merchant usage fee, while the issuing bank and the acquiring bank, as members of the card network, bargain over the interchange fee. We will model the latter as two separate Nash-bargaining solutions. First, we will look at the results when the bargaining is modeled in sequential steps, and in a second part, we analyze simultaneous bargaining. The respective bargaining power of the banks determines the degree by which the overall surplus of the card network is shared among the parties.

In general, we assume the interchange fee to be the outcome of a Nash-bargaining between the representative issuing and acquiring bank within the card network. In general, Nash-bargaining solutions are the standard concept of solving distributive problems between parties. The card network generates a total surplus, which by choice of a and f, is distributed between the parties. More specifically, Nash bargaining posits that the interchange fee and the merchant usage fee will be chosen such that each party receives a share α and (1-α) respectively, of the entire surplus of the relationship in addition to their outside option.[74] We assume the outside options to be zero in this case, because if the negotiation fails, the card network simply does not come into existence and any profits from issuing and acquiring cards would be non-existent. Hence, α will represent the bargaining power of the issuer, whereas (1-α) illustrates the bargaining power of the acquirer. These parameters attempt to capture a host of possible factors that can influence the bargaining power of a bank, such as the size, capitalization, strategic considerations, and scope for cross-financing.

Generally, there are two concepts to think about in the bargaining process within the card network. One can assume that the member banks bargain simultaneously over the merchant usage fee a and the interchange fee f. in this case, both banks sit down to bargain over a "package deal" including both fees. As an alternative, one can model the bargaining as a sequential process, in which the issuer first sets f and then the acquirer determines a.

Sequential bargaining

We begin with the acquirer. As before, we consider a single acquiring bank. For every card payment the issuer charges, the acquirer pays a fixed percentage of the transaction volume, the *interchange fee rate f* (Figure 4-1). On the basis of the fee f, the acquirer sets the merchant usage fee. By way of backward induction, we begin with the latter. Let the acquirer incur no variable costs other than the interchange fee,

---

74 For an overview, see Muthoo (1999).

and furthermore, only fixed operational costs C_A. In a monopoly with no variable costs other than the interchange fee, acquirer *A*'s profits will be the merchant usage fee revenue minus the total interchange fee paid to the issuing banks and fixed costs C_A.[75] Hence, the maximization problem of the acquirer is to choose the merchant usage fee such that profits are maximized:

$$\max_a \Pi(a) = p(a)Q(a)(a - f)\gamma - C_A . \tag{4.13}$$

The representative acquirer maximizes its profits over the merchant usage fee *a*, taking into account the fee chosen by the other acquirer. This leads to Lemma 4.1.

Lemma 4.1. *When $a \leq \bar{a}$, the monopolistic specialist acquirer maximizes profits by setting the merchant fee as:*

$$\hat{a} = (1 - \varepsilon)f + \frac{\varepsilon}{\gamma} . \tag{4.14}$$

It is straightforward to see that $\hat{a}$ is decreasing in γ: for a smaller rate of card purchases, the merchant usage fee decreases. The effect of the elasticity of (inverse) demand crucially depends on the interchange fee *f*. The Nash-solution over *f* now will be found as the maximization of the Nash product over the choice of *f*:

$$\max_f N = (pQf(a)\gamma)^{\alpha} (pQ(a - f(a))\gamma)^{1-\alpha} . \tag{4.15}$$

Note that this disregards fixed costs (e.g., fixed infrastructure costs, fixed membership fees of the payment network) because, at the time of negotiation, these costs are sunk.

---

[75] For parsimony, let the fixed costs C_A include network membership fees

Proposition 4.2. *In the case of bilateral monopolistic specialist banks and sequential Nash bargaining over a and f, with α indicating the bargaining power of the issuer, the equilibrium interchange fee is given by:*

$$f_{Seq}^{N} = \alpha \frac{\varepsilon}{\gamma} \text{ for all } a \leq \bar{a} \tag{4.16}$$

The proof of Proposition 4-2 involves finding the value for f that maximizes Eq. (4.15) (Appendix). The foregoing Proposition and Lemma 1 lead to the next Proposition about the merchant usage fee:

Proposition 4.3. *In the case of bilateral monopolistic specialist banks and sequential Nash bargaining over a and f, with α indicating the bargaining power of the issuer, the equilibrium merchant usage fee is given by:*[76]

$$a_{Seq}^{*} = \min\left\{\frac{\varepsilon}{\gamma} + \frac{\varepsilon}{\gamma}\alpha(1-\varepsilon); \frac{\varepsilon}{N - N\gamma + \varepsilon\gamma}\right\}. \tag{4.17}$$

As before, the cut-off point for the maximum merchant usage fee is given by Eq. (4.6), which is the second argument in Eq. (4.17). The first argument is found by substituting Eq. (4.16) for f in Eq. (4.14). It can be seen that the merchant usage fee increases the bargaining power of the issuer.

[76] Keep in mind that we consider merchants to have no bargaining power.

By setting the merchant usage fee in Eq. (4.17) equal to the cut-off fee, we can find the maximum number of firms (i.e., the most competitive environment of merchants) where the merchants still accept the optimum merchant fee $\hat{a}$ from the perspective of the acquirer as:

$$N_{Seq}^{*} = \frac{\gamma}{1-\gamma}(1-\varepsilon)\frac{1-\alpha\varepsilon}{1+\alpha(1-\varepsilon)} . \tag{4.18}$$

Simultaneous bargaining

The foregoing Nash-bargaining involves bargaining between the acquirer and the issuer over the interchange fee f in a scenario where a is set by the acquirer after the bargaining over f. An alternative view of the bargaining process entails the simultaneous bargaining over both a and f. Formally, this warrants the maximization of the Nash-product over the two choice variables a and f; keeping in mind that the merchant fee a is a function of the interchange fee f:

$$\max_{a,f} N = \left(pQf\gamma\right)^{\alpha}\left(pQs\left(a(f)-f\right)\gamma\right)^{1-\alpha} . \tag{4.19}$$

The solution to the maximization problem in Eq. (4.19) lets us formulate Lemma 4.2:

> **Lemma 4.2.** *In the case of bilateral monopolistic specialist banks and simultaneous Nash bargaining over a and f, with α indicating the bargaining power of the issuer, the equilibrium interchange fee is given by:*
>
> $$f_{Sim}^{N} = \alpha\, a \text{ for all } a \leq \bar{a}. \tag{4.20}$$

The proof of Lemma 4.2 is found by setting the first-partial derivative of the Nash-product Eq. (4.19), with respect to *f,* equal to zero.[77] The intuition of Lemma 4.2 is straightforward: the higher the bargaining power of the issuer, the more of the total merchant fee revenue the issuer is able to appropriate.

We now need to determine *a* by finding the partial derivative of Eq. (4.19) with respect to *a* and solving the system of equations, which will lead to the final Proposition:

Proposition 4.4. *In the case of bilateral monopolistic specialist banks and simultaneous Nash bargaining over a and f, the equilibrium merchant usage fee is given by:*

$$a^*_{Sim} = \min\left\{\frac{\varepsilon}{\gamma};\frac{\varepsilon}{N - N\gamma + \varepsilon\gamma}\right\}, \tag{4.21}$$

Refer to the Appendix for the proof. Again, by setting the respective merchant usage fee in Eq. (21) equal to the cut-off fee, we find the maximum number of firms for which the acquirer can still charge the profit-maximizing merchant fee *â*:

$$N^*_{Sim} = \frac{\gamma}{1-\gamma}1 - \varepsilon\,. \tag{4.22}$$

Discussion

The difference between the two Nash-bargaining solutions becomes clear when comparing the merchant usage fees for both, the simultaneous and the sequential case. In the latter case of simultaneous bargaining, the multi-party network acts as a

[77] Again, we consider the banks' and the networks' fixed costs as sunk costs.

unitary network. The merchant usage fee is set as ε / γ and the bargaining power solely determines the distribution of this gain between the two parties.[78] In the extreme case with the perfect bargaining power of the issuer ($\alpha = 1$), the per-unit profits of the acquirer entirely shifts to the issuer. In the other extreme, $\alpha = 0$, the interchange fee will be set to be zero, hence, the acquirer receives all the gains from the network. In the first case of sequential bargaining, the acquirer observes *f* and then adds a premium to the fee. This premium depends on the acquirer's bargaining power and the market conditions given by elasticity ε. It is easy to see from Eq. (4.17) that for a higher bargaining power of the issuer (i.e., when α increases) the merchant usage fee also increases. If, on the other hand, the acquirer has more power, there is less need to increase *a* relative to *f*, because, anyway, more of the gain remains with the acquiring bank. This is the case, because for a lower α, the interchange fee decreases (Eq. 4.16). The intersection of the optimum $\hat{a}$ and the cut-off merchant usage fees $\bar{a}$ in the sequential case (Eq. 4.18) differs from the equivalent intersection in the simultaneous case Eq. (4.22) only in the factor $\frac{1-\alpha\varepsilon}{1+\alpha(1-\varepsilon)}$, which lies in the interval [0,1]. This factor is 1 if $\alpha = 0$; for an increasing bargaining power of the issuer, the factor decreases, since the merchant usage fee also increases in α. Figure 4-3 illustrates this context.

The model discusses two variants of bargaining between parties. The practical relevance differs for both. The simultaneous case assumes that the network as an entity maximizes the overall surplus from the network. In other words, network parties first attempt to generate the maximum possible revenue, which is then distributed according to the respective stance of each party. In the sequential case,

---

78 Refer to Section 2.2.3 and Eq. (2.20) for details regarding the equivalent results in the case of a unitary payment card network.

the total surplus can be smaller, because the merchant fee increases depending on the bargaining power of the parties.

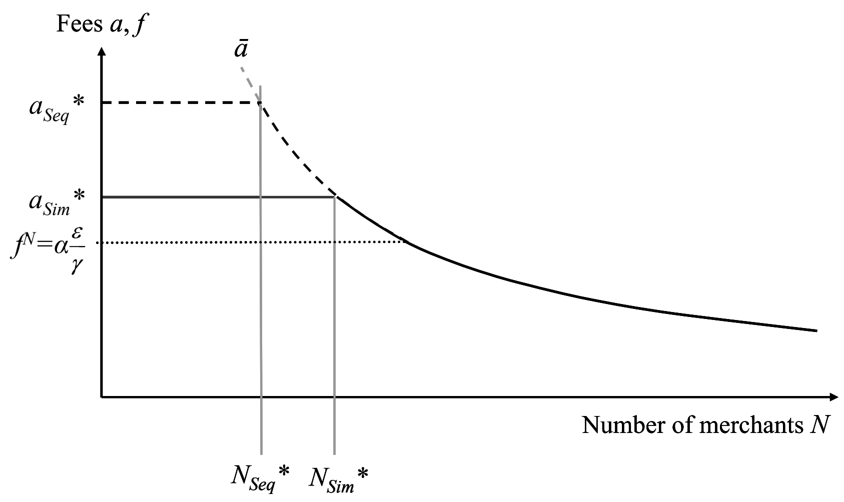

Figure 4-3: Specialist Bank's Merchant Fee.

The literature tends to think about the inner workings of the card network more as a sequential bargaining process. Wang (2010, p. 89), for instance, models the acquirers as moving after the interchange fee is set. Since, in his model, the acquirer market is perfectly competitive, the merchant fee is driven down until the profits of acquirers are zero. In our model, this is not the case, but the intuition that follows is similar. In the simultaneous case, the network banks are able to transfer profits to the issuing side, where they cannot be competed away, regardless of the competitive situation on the acquiring market (Wright, 2004). In reality, even though merchants seem to be successful in bargaining the merchant fees, the majority of merchants claim a limited ability to negotiating interchange fees (U.S. Government Accountability Office, 2009, p. 36).

In this study, the merchant and the interchange fees are found to be decreasing in the frequency of card usage. This might appear as a contradiction, since e.g., over the past years, the interchange fees, as well as the frequency of card usage, have simultaneously been increasing.[79] First, for parsimony, we assumed merchants to have no bargaining power. In reality, this is obviously not the case. One can expect bigger retailers to be able to cut a better deal with payment card networks. Thus, an increase in the frequency of card usage might weaken the merchants' ability to deny cards, and thus, enable the acquirer to negotiate a higher merchant fee. Second, the foregoing model illustrates a mature payment market. In contrast, the speed of growth of the payment card markets might be an indicator that these markets have not yet reached a stage of maturity. From the perspective of the networks and the banks interchange and merchant fee might not yet be in the optimum; hence, fees and the card usage are still growing instead of balancing each other out. It remains an empirical question which of these effects might prevail.

Consider, for example, the simultaneous bargaining case. Ultimately, when $a<\bar{a}$ the optimum merchant usage fee is chosen to be anti-proportional to the relative frequency of card usage; i.e., $\hat{a}_{sim}*=\varepsilon/\gamma$. Mathematically speaking, the elasticity of a with respect to γ is constant negative unity. In Section 2, we learned, that when $a=\varepsilon/\gamma$, the profits of the merchants, as well as the network, is thus constant over the relative frequency of card usage; card usage and the optimum merchant fee $\hat{a}_{sim}*$ balance each other out. Figure 4-4 plots this context, i.e., the optimum merchant usage fee over the frequency of card usage. The three rectangles in Figure 4-4 are the fee revenues for different combinations of γ and a; each leading to the same

79 For example, Wang (2010, p. 87) in the U.S. reports an increase in the interchange fee from about 1.3% in 1996 to 1.8% in 2006, regarding a $100 non-supermarket transaction. Regarding the percentage of payment cards used for in-store purchases in the U.S., Bolt & Chakravorti (2008, p. 15) refer to an increase from 43% in 1999 to 56% in 2005 (original source: American Bankers Association and Dove Consulting, 2005).

(optimum) fee revenue for the network and banks respectively. In fact, with no surcharging, the merchant will consider $\gamma \cdot a$ as a cost when calculating his/her products.

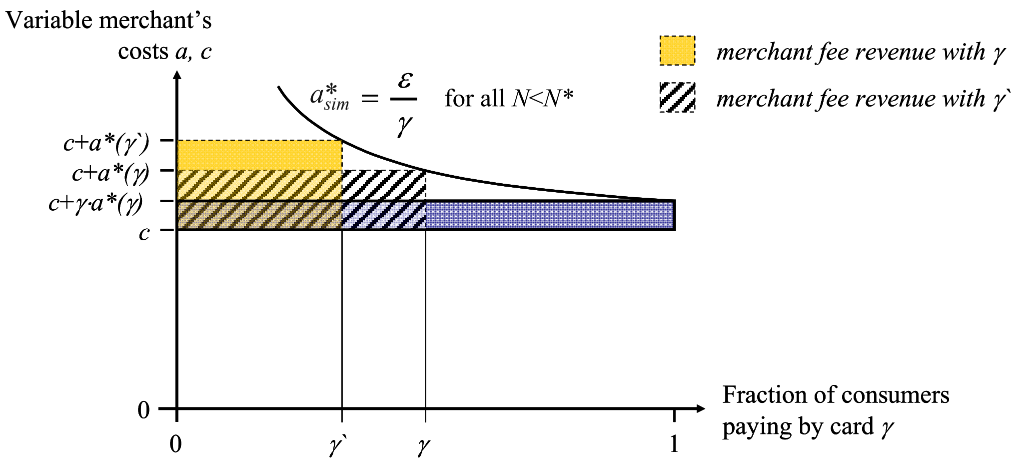

Figure 4-4: Effect of Anti-Proportionality of γ and ε: Constant Merchant Fee Revenue.

4.3 Discussion and Conclusions

The foregoing model attempts to determine merchant usage fees and interchange fees in different settings. We find a broader generality to the findings in Section 2, i.e., for the latter case of a multi-party payment card network. We find the interchange fee decreasing in the price elasticity of consumer demand[80] and in the relative frequency of card usage.

As a first step, the merchant usage fee for duopoly all-purpose banks, which engage in both issuing and acquiring cards, is derived. For our purposes, it seems natural to disregard the interchange fee at this point, because payments received and paid in interchange fees cancel each other out. The result is similar to the unitary network case discussed in Section 2.

The model then moves on to examine the polar opposite case of specialist acquirers and issuers. The interchange fee has an important role in this setting, because it serves as revenue for one party (issuer), while constituting a cost for the other (acquirer). Hence, the choice of the fee has implications for the distribution of the total fee revenue generated by the card network. As such, it warrants the use of Nash-bargaining for modeling. In the first case, we model the choice of the merchant usage fee and the interchange fee as the outcome of a sequential game: the issuer sets the interchange fee and then the acquirer sets the merchant fee, i.e., depending on the outcome of the prior negotiations. The acquirer, depending on the outcome of the negotiations over the interchange fee, will extract an additional margin by setting the merchant usage fee accordingly higher. Hence, the issuer will have to anticipate the behavior of the acquirer, who will set the merchant fee during a future time period. For the second case, both fees are negotiated simultaneously as a "package deal" of the multi-party network. Note that, only for the second case, the acquirers with no bargaining power regarding the interchange fee will make zero profits.

---

80 Currently, the card networks differentiate interchange fees for different categories. The differentiation relates to fraud costs (e.g., face-to-face purchases with the card present vs. online-purchases), rewards, etc. (Wang, 2010, p. 92).

The interchange fee is revenue to the issuers and a cost to the acquirers. Thus, the acquirer is interested in a rather low interchange fee rate, while the issuer strives to optimize profits over the interchange fee. The all-purpose banks are also interested in optimized interchange fees, since the interchange fee is a means to transfer profits to the issuing side, where they are less likely to be competed away. Consider that power in a payment card network corresponds to the transaction volume. In this case, with the natural symmetry of issuing and acquiring, for any given multi-party payment card network, the associated all-purpose banks and issuers make up the majority of the network banks. In addition, all-purpose banks and specialist issuers are in a powerful market position, since they are the upstream players in the payment card market and most of their downstream markets are buyer's markets. Remember that, in the past, the two big multi-party payment card networks were set-up as associations, with the member banks appointing the network's executive board.[81] Not surprisingly, payment card networks seem to be set up to maximize the interchange fee revenue (Wang, 2010, p. 89).[82]

The models results deserve further discussion. First, it has been shown that in the case of the simultaneous Nash-bargaining of the issuer and the acquirer over the merchant and the interchange fees, the resulting merchant fee is the same as in the case of a unitary network (refer to Propositions 2.1 and 4.4). Hence, a simultaneous bargaining of the two variables works like a unitary network, because both variables are chosen to maximize the overall network's surplus. The internal distribution of the revenue is then only dependent on the relative bargaining power of the parties. Furthermore, the double marginalization associated with multi-party networks

---

81 Even so, the organizational form of Visa and MasterCard was changed to publicly traded companies; there is reason to presume that this did not necessarily change the objectives of the card network, i.e., the scheme (refer to Wang, 2010, p. 92).

82 The payment card networks obtain much of their revenue in the form of the so called Transaction Processing Fee.

vanishes in this case. Only when we model the bargaining process as a sequential game, the acquirer when setting the merchant fee will include an additive premium on top of the interchange fee. This premium is a percentage of the interchange fee that is decreasing in elasticity: the less elastic demand, the less severe the double marginalization (refer to Proposition 4.3). This also depends on the bargaining power of the issuer: the more asymmetrically the bargaining power is vested with the issuer, the higher the premium, and therefore, the marginalization in the market. Thus, for the sequential bargaining case, the multi-party payment card network might provide higher merchant usage fees than the unitary network or the multi-party network in the case of simultaneous bargaining.[83]

With no doubt, there is a correlation between failure in the downstream market and the above determinants of the merchant usage fee, i.e., the elasticity of consumer demand and the competitive condition of the merchants. Consequently, regulating the interchange fee and the merchant fee, respectively, is found to be particularly useful, since in markets with less competition where consumers already face inefficiently high prices, the payment card fees will also tend to be higher. Such regulations might help to reduce the inefficient double marginalization of acquirers and issuers. On the other hand, regulating the interchange fee limits the networks' ability to negotiate profits from the acquiring side of the payment card network to the issuing side. In doing so, the merchants' bargaining position will be strengthened, as

[83] The above study suggests that multi-party payment networks due to potential double marginalization within the network tend to request higher merchant fees than unitary payment card networks. When comparing this result to reality, there is a conflict, since the biggest unitary payment card network, American Express, is known to charge significantly higher rates than its biggest competitors, Visa and MasterCard. This conflict will be discussed in the remaining. Firstly, the foregoing model did not concern the strategic considerations of the networks. Secondly, in the case of multi-party payment card networks, house banks offer card services to their clients, which involves cost opportunities (synergies to other banking services and processes) and strategic advantages (e.g., when judging credibility). Thirdly, American Express is historically very familiar with certain consumer segments, e.g., business travelers, which are one of the prime payment card segments (i.e., characterized by rather high price elasticities). The foregoing study explains why merchants in such segments are likely to accept fairly high fees.

they are in an underdog position. We remain to discuss further shortcomings of the model, which could be addressed in future extensions of the model. The below possible main extensions are related to questions of the bargaining power of merchants and the competitive situation of multi-party networks.

Our model assumes no bargaining power on the part of the merchants. Acquiring banks simply offer a take-it-or-leave-it proposition on a merchant usage fee to the merchant. If, for this fee, profits from card sales are non-negative for the merchant, the fee will be accepted, otherwise, the cut-off fee with zero profits is chosen. If one were to relax this assumption of a merchant with no bargaining power, several considerations will come into play. First, insofar as merchants have no bargaining power, the assumption that the interchange fee in the case of all-purpose banks is irrelevant can easily be defended. This, however, changes whether we introduce a bargaining process between all-purpose banks and merchants. Now, for the banks, the interchange fee serves as a way to extract more profits from the merchants, since it constitutes a cost that credibly warrants a higher merchant fee to keep the network alive. The threat point of the bank is always that for too low a merchant fee, the bank's acceptance of purchases made with other cards withers. This would segment the market effectively into a number of unitary networks. Note that there would be a significant loss of transactions (i.e., all the transactions with different issuers and acquirers). In effect, one could observe the shifting of profits to the interchange part of the business, where there is no incentive for other banks to compete for market shares. This is a result discussed in the literature (for example, Wright, 2004).

Currently, card networks charge different interchange fees depending on the fraud costs of the transaction category, e.g., the networks generally charge a lower interchange fee in the case of face-to-face purchases with the card present than for online-purchases (Wang, 2010, p. 92). In addition, our study suggests that card networks maximize profits by differentiating the interchange fee depending on the above downstream determinants. This can be clearly observed when analyzing the development of interchange fees. Table 4-1 reports statistics on the change of interchange fees between 1991 and 2009. Not only that the number of interchange fee categories has grown dramatically, but also the spread of fees has been expanded. To provide a particularly stark example, MasterCard increased the number of categories of interchange fees from 4 in 1991 to 243 in 2009; the spread between the highest and the lowest fee has increased as well. In less competitive markets (i.e., when consumers already face inefficiently high prices), payment card fees bear the potential to be similarly high, thus adding to the problem. Thus, together with these numbers, our model suggests that policy-makers observe such behavior carefully.

Changes in rates from 1991 to 2009	Visa	MasterCard
Number of interchange rate categories 1991	4	4
Number of interchange rate categories 2009	60	243
Range of interchange fee rates 1991	1.3%-1.9%	1.3%-2.1%
Range of interchange fee rates 2009	1.0%-3.0%	0.9%-3.3%
Percentage of rates that increased	43	45
Percentage of rates that stayed the same	45	45
Percentage of rates that decreased	12	10

Table 4-1: Changes of Visa and MasterCard Credit Card Interchange Fee between 1991 and 2009 (Source: U.S. Government Accountability Office, 2009, Table 2, p. 15)

Another way to modify and extend the model involves introducing greater competition and heterogeneity among card networks. So far, the focus is on a duopoly of all-purpose banks and a bilateral monopoly of two specialist banks. It seems natural to think through cases where either the number of networks increases, maybe even competition among unitary and multi-party payment card networks, or the number of banks within a network increases. Also, increasing the heterogeneity of networks in the sense that we would have both all-purpose banks and specialist banks operating alongside could offer yet richer results.

5 Discussion and Conclusions

The payment card pricing literature primarily focuses on the upstream payment market.[84] In contrast, there is reason to believe that payment card fees do depend on the conditions of the downstream markets they serve, because a payment card network will extract a portion of the merchants' economic rent. In addition, the existing literature cannot sufficiently explain certain variations in the merchant fee (e.g., regarding different merchant sectors[85]), there remains a gap in understanding the determination of payment card prices. Thus, the purpose of this thesis is to investigate the relationship of downstream market conditions (i.e., characteristics of the product(s) and the competitive conditions of the market the merchants face) and payment card prices, including the merchant usage fee and the interchange fee.

More specifically, the present thesis consists of three self-contained studies analyzing different settings of mature payment card markets.[86] While the first two settings concern unitary payment card networks, the third incorporates a multi-payment card network. As such, the second and the third models are extensions of the first study. Not surprisingly, it is discovered that the determinants of the merchants' profits do similarly impact the level of payment card prices. To conclude, this thesis provides a new perspective into payment card pricing, by developing a better understanding of the dynamics between the (original) downstream market and the upstream payment card market.

---

84 For an overview of the related literature and its history, refer to Bolt & Chakravorti (2008) and Evens & Schmalensee (2005b). In contrast, literature which considers the impact of downstream market conditions onto payment card network prices are the essays accumulated in this thesis, i.e., Langlet (2009), Buchen & Langlet (2010), and Langlet & Uhlenbrock (2010), as well as Wang (2010).

85 For details on the variation in payment card network prices, refer to the European Commission Competition DG (2006) and U.S. Government Accountability Office (2009).

86 In this study, strategic considerations of the network are not taken into account. By referring to equilibrium prices (i.e., privately optimized prices), I consider a mature payment card market.

In Section 2, a model of a monopolistic unitary payment card network is constructed (Langlet, 2009). Consumer demand is assumed to be isoelastic and merchants are considered under Cournot quantity competition.[87] Results revealed that the merchant usage fee is decreasing in the price elasticity of consumer demand. Furthermore, this study provides a rationale for lower payment card fees when more consumers pay with a card. Chakravorti and To (2008) find the opposite result which relates to the credit function of some cards. In addition, when more consumers use cards, the merchants can less likely deny cards; i.e., the bargaining power of the networks increases in card usage.[88] Lastly, a mature card market is considered, whereas even the recent increase of card fees and usage indicates that the payment card market has not yet reached a stage of maturity. Concluding, it remains an empirical question which of these effects dominates. Caution must be taken, however, as the observation of an immature payment card market implies the threat of a continued growth of payment card prices, i.e., a significant potential for market inefficiencies, especially since policy-makers, even today, are concerned about the level of payment card prices. This finding offers some justification for the recent concern and actions of policy-makers.

Regarding the optimal merchant usage fee from the perspective of the payment card network, I find a threshold for which merchants are at the edge of denying cards, the maximum merchant usage fee, which depends on the competitive condition of the merchants. In reality, we find that the networks "*...have developed lower rates to encourage certain merchants to accept their cards.*" (U.S. Government

---

87 This allows for a focused analysis of the impact of the price elasticity of consumer demand and the relative frequency of card usage onto the merchant usage fee, but requires the assumption of homogeneous products being exchanged in the downstream market.

88 Note the network externalities; the more consumers a network succeeds to attract, the more interesting it becomes to merchants and the easier it becomes for the network to attract further consumers. On the other hand, this effect makes it harder for merchants to deny cards, i.e., the bargaining power of the network increases.

Accountability Office, 2009, p. 9.) Not surprisingly, this price differentiation seems to be closely followed by a rise of the payment card prices.[89]

The setting in Section 3 (Langlet & Uhlenbrock, 2010) is similar to the first study, only in that the underlying model concerns merchants under Bertrand price competition, which requires a different demand function.[90] In return, heterogeneous products can be regarded, which allows for a study of the impact of product substitutability. As a matter of fact, we find the merchant usage fee is decreasing in the product substitutability.[91] For example, fuel is a fairly homogeneous product.[92] Hence, petrol stations face fairly strong competition and drivers are rather price sensitive. In contrast, restaurants offer rather inhomogeneous products, spreading from fast food to exclusive gourmet cuisine. Hence, competition tends to be weaker among different kinds of restaurants.[93] Consequently, the observation that restaurants, on average, pay a merchant usage fee of 2% and petrol stations pay 1% of the transaction value[94] supports the above finding. Finally, this study provides a broader generality for the findings of the first model.

89 For further details, refer to Section 4.3 and U.S. Government Accountability Office (2009, pp. 14-18).

90 In this regard, we consider a linear demand function based on Singh & Vives (1984).

91 Note that product substitutability can serve as a measure of the different competitive conditions of the merchant(s). As such, the regarded products can be characterized on a continuum which ranges from (perfect) substitutability (i.e., a competitive environment) via no substitutability at all (i.e., the monopolist) to (perfect) product complementarity.

92 Note that, due to industry standards, the product substitutability of fuel in most countries is perfect. This fact is highlighted when observing that petrol companies try to differentiate their fuel by the color, which is an addition to the same basic product with no technological effect.

93 Even so, there might be exceptions to this; e.g., one burger might be very similar to the burger of another fast food restaurant.

94 This is an observation of the European Commission Competition DG (2006, p. 41).

In Section 4, we extend the case of Section 2, such that it models a multi-party payment card network (Buchen & Langlet, 2010).[95] We find that the network banks can serve in two different settings, each leading to different outcomes. Firstly, the banks can be all-purpose banks that issue and acquire in an equal extent. Due to this symmetry, each of these banks' balance of paid and received interchange fee revenues is zero. Therefore, the equilibrium merchant fees for this case is similar to the unitary payment card network's outcome, except that the multi-party payment card network, by using the interchange fee, might be able to transfer profits from the acquiring to the issuing side of the network. In this way, the network reduces the bargaining power of the merchants' regarding the merchant usage fee, since the interchange fee serves as lower bound for the merchant fee.

Secondly, banks might be specialist banks, i.e., some banks specialize in issuing cards and other institutions are specialist acquirers. The latter case might lead to a double marginalization, even within the payment card network.[96] In reality, we find all-purpose banks and specialist banks are both examples of actual business models in the industry. As a matter of fact, in accordance with the net interchange fee balance, any bank associated with a multi-party payment card network can be plotted on a continuum; specialist issuers on one end, all-purpose banks in the center, and specialist acquirers on the other end. By considering simultaneous Nash bargaining of the acquirer and the issuer regarding the interchange and the merchant fee, the sub case of a unitary network can be modeled for this setting. In contrast, a sequential setting, where the issuer and the acquirer first bargain (Nash) over the interchange

95 This setting is identical to the first model (Section 2), other than the network being a multi-party payment card network; consider a homogeneous product and isoelastic consumer demand, merchants shall be under Cournot quantity competition. Even though for simplicity we assume two banks associated with the network; other oligopoly settings can easily be derived. In addition, there shall be a uniform interchange fee.

96 Thus, in total, there might result triple marginalization when considering that the merchant, the acquirer, and the issuer want to obtain a profit margin.

fee and then the acquirer sets the merchant fee, models the common practice of multi-party payment card networks. In addition, by giving bargaining power to merchants, the setting could be extended in order to model reality more precisely.

In addition, we find that interchange fees are set to maximize the interchange fee revenue, thus generating revenue for the networks and driving up the profits of specialist issuers and all-purpose banks. Note that specialist issuers and all-purpose banks make up the 'majority'[97] of the associated banks of a multi-party network, which is due to symmetric acquiring and issuing, i.e., the nature of the network. Therefore, and not surprisingly, issuing is found to be, by far, more profitable than acquiring.[98]

Last, but not least, this third study provides a broader generality of the findings of the foregoing analysis'. Again, we find the merchant usage fee and here, the interchange fee, decreasing in the price elasticity of consumer demand and in the fraction of card usage. Not surprisingly, the maximum merchant usage fee is also a threshold to the interchange fee that a card network can obtain.

Concluding, it can be stated that the multi-party payment card networks make the impression to be better suited for competing in the payment market than the unitary payment card networks. This is highlighted by the fact that the Visa and MasterCard networks alone account for about 80% of the U.S. credit card market (Wang, 2010, p. 86). Thus, with reason, policy-makers watch over the pricing of multi-party payment

---

97 'Majority' here refers to associated banks weighed by their interchange fee revenue. Thus, there is just as much interchange fee revenue 'net' issuers receive as 'net' acquirers pay. Consequently, I refer to 'net' issuers and all-purpose banks, together, being the 'majority' of a multi-party payment card network.

98 For further detail, refer to European Commission Competition DG (2006, pp. 62-77) and U.S. Government Accountability Office (2009, p. 21-24).

card networks with more concern than the fees of the unitary networks.[99] The advantages of the multi-party payment card networks over the unitary networks can be highlighted by the fact that Discover, originally a unitary network, currently seems to pursue a conversion to a multi-party payment card network. Just recently, Discover has been successful in offering card-issuing agreements with other financial institutions and "*...is moving from a single rate for each merchant that applies to all of their cards to a tiered interchange fee model, with higher interchange fees for rewards and corporate cards.*" (U.S. Government Accountability Office, 2009, pp. 17-18.)

In this study, the author finds reason to believe that multi-party payment card networks are more likely to bring about market inefficiencies than unitary payment card networks. Potential market failure, for example, is more likely inherent in the potential double marginalization over acquiring and issuing. The networks' ability to transfer profits to the issuing side of the network by the collective setting of an interchange fee has a similar effect, since undermining the bargaining power of the merchants. In addition, the fact that issuing is a highly profitable business enables issuers to give card holders extra incentives for card usage, which bears the latent risk to boost card over-usage. Another impression is that multi-party payment card networks are more successful in price differentiating merchant fees through a differentiated interchange fee, which seems to be closely followed by an increase in payment card prices (Table 4-1).

In general, the author finds that downstream markets with less competition are typically characterized by higher payment card fees, i.e., due to the cut-off fees in highly competitive markets, as well as the fact that payment network prices decrease

---

[99] Visa and MasterCard have generated more concern among policy-makers than the unitary networks (most notably American Express and Discover), for the most part, since being more widely spread (U.S. Government Accountability Office, 2009, p. 17).

in the price elasticity of consumer demand as well as in the product substitutability. Therefore, regulating the merchant fee and the interchange fee, respectively, seems to be particularly useful, since in less competitive environments, merchants already face inefficiently high prices. Furthermore, fee regulation appears particularly useful to help reduce the inefficient double marginalization of acquirers and issuers. Especially, regulating the interchange fee can help to limit the ability of the multi-party networks to transfer profits to the issuing side, thus strengthening the bargaining power of the merchants who tend to be in the underdog position. Finally, interchange fee regulation might prove useful, as it concerns the multi-payment card networks, thus implicitly strengthening the unitary networks. Remember the above discussion of the multi-party payment card networks that seem to dominate the payment card markets. Thus, this study, to some extent, might justify recent attempts of policy-makers regulating down payment card fees, especially the interchange fee. However, fee regulation requires a thorough and ongoing observation of the payment card market, especially since it lacks adaptability and self-regulatory features. In addition, policy-makers need to be aware of the fact that fee regulation might render unintended consequences, such as encouraging card over-usage. Remember, in the foregoing study, we learned that networks can approach maximum profits by adapting fees and/or card usage.

However, the question of judging a potential failure on the payment card market, for the most part, remains unsolved. This calls for further research regarding the determination of privately and socially optimized payment card networks. In this regard, the importance of gaining a deeper understanding can hardly be overemphasized, since, firstly, the payment card networks apparently are in a very powerful competitive position (e.g., due to network externalities). Secondly, the economics of payment cards is fairly complex. Thirdly, this need for an improved understanding is highlighted when observing that the payment card market in itself is huge and that card payments concern a decent portion of a country's GDP, namely the retail industry and related markets.[100]

Consider a brief illustration: In the U.S. in 2008, merchants spent over $60 billion in merchant usage fees (Federal Reserve Bank of Kansas City, 2009, p. 1). Thus, plain card services, on average, add up to annual merchant fees of about $200 per U.S. citizen. Add the card fees. Now compare the amount to an average yearly telephone bill, a service that supposedly creates much more consumer surplus than the payment function of card services over cash, checks, etc.[101] Beyond, remember that payment card prices and revenues in the future might continue to grow.

In the remaining, four potential extensions of the foregoing study will be pointed out. Firstly, the main aim of this study is to investigate the determination of privately optimized payment card prices. Thus, a worthwhile extension would concern the social efficiency of the payment card market bearing the potential to further assist in the discussion of potential policy interventions. In this regard, exploring the dynamics between the downstream and upstream payment market and the corresponding effects appears particularly appealing. For example, by introducing the costs and benefits of card usage, the upstream market could be modeled more realistically. A possible aim could be to investigate the effects of card usage rewards which would require relaxing the assumed card preference of the consumers. Thus, alternative approaches to regulate payment card markets could be tested. A leading hypothesis could be, for example, the intuition that the merchants might be better equipped to control the social effects of card usage rewards than the networks, since

[100] This is highlighted by the fact that, for example, in 2009, the purchase volume of U.S. general purpose cards was about $3 trillion (an approximation based on data from the HSN Consultant, 2010, p. 7). GDP refers to gross domestic product.

[101] Note, that the above value only concerns the payment function of cards, since it does not include the costs of interest when using cards for credit.

probably being more sober about card over-usage. In contrast, issuers tend to encourage card over-usage, since card usage expands their potential revenues, and in doing so, strengthens their bargaining position. Thus, exploring the dynamics between the downstream and the upstream payment market seems appealing.

Secondly, a natural extension of the foregoing study would consider competition among several networks. With coexisting unitary and multi-party payment card networks, effects such as blending will have to be analyzed in more detail.[102] This study suggests that when isolated, multi-party payment card networks tend to obtain higher merchant fees than the unitary networks, i.e., due to double marginalization even within the multi-party network and by transferring profits through the interchange fee from the acquiring to the issuing side of the network. Thus, an extension of the foregoing study incorporating competition among different kinds of networks seems appealing. In addition, the multi-party payment card networks claim that payment card prices might increase with more competition among the networks, since the networks' competition for issuers would motivate higher interchange fees (U.S. Government Accountability Office (2009, pp. 18-21).

Interestingly, in the foregoing study, I find another rationale for an increasing merchant usage fee in the network competition, even for unitary networks. This effect is due to the complex dynamics between the upstream payment market and the downstream market, i.e., where products are being exchanged between consumers and merchants. This claims that, in contrast to other markets, increasing network competition might not be the answer to obtaining more efficiency in the payment card markets. Consequently, a highly important question is how regulation might be effectively approached, especially with the currently low number of big payment

[102] Blending is a widespread phenomenon of payment card networks. This is, among others, an observation of the European Commission of Competition DG (2006, pp. 64-65).

card networks. For example, creating competition to the payment card networks through encouraging innovation regarding alternative payment instruments (mobile payments, etc.), maybe even under public control, might play a significant role in such an approach.

Thirdly, there is a lack of empirical analysis, thus, conclusive guidelines for policy makers. This observation is highlighted when reviewing the different opinions and policy-makers approaches regarding the interchange fee all around the world, e.g., ranging from the general prohibiting of interchange fees in Poland to great controversy and hesitance in the U.S., even so, network fees in the U.S. are among the highest in the world (Bradford & Hayashi, 2008, pp. 1-2). One barrier regarding empirical study of the payment card markets is the lack of available data. This leads to a lack of transparency, mostly to the advantage of the networks and the disadvantage of merchants (and consumers). That is, even increasing transparency and raising public scrutiny regarding payment card pricing might be a worthwhile regulatory instrument (Wang, 2010, p. 95). On the other hand this transparency might provide economists and policy-makers with data necessary for empirical analysis and further needed theory, a potential stepping stone toward more efficiency on the payment card markets. In regard to the present thesis, such data could be helpful to empirically test the foregoing results, such as e.g. correlation analyses regarding fee level and the downstream market characteristics (e.g., price elasticity of demand, card usage).

Fourthly, the foregoing study could be adapted in such a way to model other two-sided markets. Such an extension seems possible, especially for market makers (e.g., internet auction platforms) and demand coordinators (e.g., supplier collaboration platforms), but might be harder to apply for audience makers (e.g., information internet portals). In this regard, an observation of the dynamics between the downstream and upstream market appears similarly appealing as in the foregoing study. For example, consider a supplier collaboration platform serving different industries (automotive, aviation, machine building, etc.). An observation similar to the foregoing study might conclude that such a supplier collaboration platform would differentiate fees depending on the conditions of the downstream industry, such as the price elasticities of demand.

References

American Bankers Association and Dove Consulting (2005). *Study of the consumer preference*. Report, Washington/DC, October.

Armstrong, M. (2006). Competition in two-sided markets. *The Rand Journal of Economics*, 37 (3), 668-691.

Baxter, W. F. (1983). Bank exchange of transactional paper: legal and economic perspectives. *Journal of Law and Economics*, 26 (3), 541-588.

Board of Governor of the Federal Reserve System (2010). *Survey of credit card plans*. July, 31, 2010. Accessed October 14th, 2010, from http://www.federalreserve.gov/creditcard/survey.html

Bolt, W., & Chakravorti, S. (2008). Economics of payment cards: A status report. *Economic Perspectives.* Federal Reserve Bank of Chicago, 4th quarter, 15-27.

Bolt, W., Jonker, N., & van Renslaar, C. (2008). Incentives at the counter: An empirical analysis of surcharging card payments and payment behavior in the Netherlands. *DNB Working Paper* 196 (December 2008).

Bradford, T., & Hayashi, F. (2008). *Developments in interchange fees in the United States and abroad*. Federal Reserve Bank of Kansas City, Payments System Research Brienfings.

Buchen, C., & Langlet, M. J. (2010). Merchant and Interchange fee determination in payment card networks with quantity competition among merchants. *Mimeo*. EBS Universität für Wirtschaft und Recht Wiesbaden.

Caillaud, B., & Jullien, B. (2003): Chicken & egg: Competition among intermediation service providers. *RAND Journal of Economics*, 34(2), 309-328.

Carlton, D., & Frankel, S. (1995a). The antitrust economics of credit card networks. *Antitrust Law Journal*, 63 (2), 643-668.

Carlton, D., & Frankel, S. (1995b). The antitrust economics of credit card networks: Reply to Evans and Schmalensee. *Antitrust Law Journal*, 63 (3), 903-915.

Chakravorti, S. (2003). Theory of credit card networks: A survey of the literature. *Review of Network Economics*, 2 (2), 50-68.

Chakravorti, S., & To, T. (2007). A theory of credit cards. *International Journal of Industrial Organization*, 25(3), 583-595

Chang, H., & Evans, D. (2000): The competitive effects of the collective setting of interchange fees by payment card systems. *Antitrust Bulletin*, 45(3), 641.

European Commission Competition DG. (2006). *Interim report I – Payment cards*. Accessed October 11, 2007, from http://ec.europa.eu

Evans, D. (2002). The antitrust economics of two-sided markets. *Joint Center AEI-Brookings Joint Center for Regulatory Studies*. Accessed April 26, 2008, from http://aei-brookings.org

Evans, D. (2003). It takes two to Tango. *The Payment Card Economics Review*. 1 (Winter), 13-24.

Evans, D., & Schmalensee, R. (1995). Economic aspects of payment card systems and antitrust policy toward joint ventures. *Antitrust Law Journal*, 63 (3), 861-901.

Evans, D., & Schmalensee, R. (2005a). *Paying with plastic*. 2nd Ed., Cambridge, Massachusetts: MIT Press.

Evans, D., & Schmalensee, R. (2005b). The economics of interchange fees and their regulation: An overview. *MIT Sloan Working Paper* 4548-05 (July 8, 2005). Accessed April 14, 2008, from http://dspace.mit.edu

Evans, D., & Schmalensee, R. (2007). The industrial organization of markets with two-sided platforms. *Competition Policy International*, 3(1), 151-179.

Federal Reserve Bank of Kansas City (2009). *An update on interchange legislation in the United States*. By Christian Hung. Accessed October 2, 2010. from http://www.kansascityfed.org/Publicat/PSR/Briefings/PSR-BriefingDec09.pdf

Frankel, A. (1998). Monopoly and competition in the supply and exchange of money. *Antitrust Law Journal*, 66 (2), 313-361.

Gans, J., & King, S. (2003). The neutrality of interchange fees in payment systems. *Topics in Economic Analysis & Policy*, 3 (1). 1-16.

Guthrie, G., & Wright, J. (2007): Competing payment schemes. *Journal of Industrial Economics*, 55(1), 37-67.

HSN Consultants (1987). *The Nilson Report*. No. 406 (June).

HSN Consultants (1991). *The Nilson Report*. No. 406 (June).

HSN Consultants (1992). *The Nilson Report*. No. 406 (June).

HSN Consultants (2001a). *The Nilson Report*. No. 406 (June).

HSN Consultants (2001b). *The Nilson Report*. No. 406 (June).

HSN Consultants (2010). *The Nilson Report*. No. 955 (August). Accessed December, 2010, from http://www.nilsonreport.com

Hunt, R. (2003). An introduction to the economics of payment card networks. *Review of Network Economics*. 2 (2), 80-96.

ITM Research (2000). *The abolition of the no-discrimination rule*. Accessed August 10, 2009, from
http://www.creditslips.org/files/netherlands-no-discrimination-rule-study.pdf

IMA Market Development AB (2000). *Study Regarding the Effects of the Abolition of the Non-Discrimination Rule in Sweden for the European Commission Competition Directorate General*. Final report, 2000-02-29.

Jullien, B. (2005). Two-sided markets and electronic intermediaries. *CESifo Economic Studies*, 51 (2-3), 230-260.

Katz, M. (2001). *Reform of Credit Card Schemes in Australia II: Commissioned Report.* RBA Public Document. August.

Langlet, M. (2009). Pricing of unitary payment card networks: The relationship between consumer price elasticity and merchant fees. *Oxford Journal of Business and Economics*, Fall 2009, 8 (1), 93-106.

Langlet, M. & Uhlenbrock, J. (2010). Merchant fee determination in unitary payment card networks with price competition among merchants. *Mimeo*. EBS Universität für Wirtschaft und Recht Wiesbaden.

Muthoo, A. (1999). *Bargaining Theory with Applications*. Cambridge: Cambridge University Press.

Rochet, J., & Tirole, J. (2002). Cooperation among competitors: Some economies of payment card associations. *RAND Journal of Economics*, 33 (4), 549-570.

Rochet, J., & Tirole, J. (2003a). An economic analysis of the determination of interchange fee in payment card systems. *Review of Network Economics*, 2 (2), 69-79.

Rochet, J., & Tirole, J. (2003b): Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4), 990_1029.

Rochet, J., & Tirole, J. (2006a). Two-sided markets: a progress report. *Rand Journal of Economics*, 37 (3), 645-667.

Rochet, J., & Tirole, J. (2006b). Externalities and regulation in card payment systems. *Review of Network Economics*, 5 (1), 1-14.

Rochet, J., & Tirole, J. (2008). Tying in two sides of the market and the honor all cards rule. *International Journal of Industrial Organization*, 26 (2008), 1333-1347.

Rysman, M. (2007). An empirical analysis of payment card usage. *Journal of Industrial Economics*, 55 (1), 1-36.

Scheerer, M. (2007). EU greift MasterCard und Visa an. *Handelsblatt*, 243 (December 17th), 21.

Schmalensee, R. (2002). Payment systems and interchange fees. *Journal of Industrial Economics*, 50 (2), June, 103-122.

Schmalensee, R. (2003). Interchange fees: A review of the literature. *The Payment Card Economics Review* 1.

Singh, N., & Vives, X. (1984). Price and quantity competition in a differentiated duopoly. *RAND Journal of Economics*, 15 (4), 546-554.

Shy, O. & Wang, Z. (2008). *Why do card issuers charge proportional fees?* Mimeo. Federal Reserve Bank of Kansas City.

United States Department of Commerce: Bureau of Economic Analysis (2003). *Imputations in the National Income and Product Accounts*. September 29.

United States Government Accountability Office (2009). *Credit Cards - Rising Interchange Fees Have Increased Costs for Merchants, but Options for Reducing Fees Pose Challenges*. Report to Congressional Addressees.

Wang, Z. (2010). Market structure and payment card pricing. What drives the interchange? *International Journal of Industrial Organization*. 28 (2010), 86-98.

Wright, J. (2003). Pricing in debit and credit card schemes. *Economic Letters*. 80 (3), 305-309.

Wright, J. (2004). The determinants of optimal interchange fees in payment systems. *Journal of Industrial Economics*, 52 (1), 1-26.

Appendix

Proofs of Section 2

Proof of Lemma 2.1

Merchant n will maximize profits π_n over sales quantity q_n

$$\max_{q_n} \ \pi_n \tag{2.7}$$

$$\text{with } \pi_n = q_n \cdot (q_n + q_{-n})^{-\varepsilon} \cdot (1-\gamma \cdot a) - q_n \cdot c \tag{2.6}$$

The first order condition is:

$$\frac{\partial \pi_n}{\partial q_n} = -\varepsilon \cdot (q_n + q_{-n})^{-1-\varepsilon} \cdot q_n \cdot (1-\gamma \cdot a) + (q_n + q_{-n})^{-\varepsilon} \cdot (1-\gamma \cdot a) - c \overset{!}{=} 0$$

$$\text{with } Q = q_n + q_{-n} \tag{2.2}$$

The solution of the first order condition requires the assumption of all merchants behaving according to the Nash-Cournot equilibrium concept, therefore maximizing profits and choosing identical sales quantities, thus:

$$q_n = \frac{q_{-n}}{N-1} = \frac{Q}{N} \tag{2.8}$$

$$\Rightarrow q^* = q_n^* = \frac{1}{N} \cdot \left(\frac{\left(1 - \frac{\varepsilon}{N}\right) \cdot (1-\gamma \cdot a)}{c} \right)^{\frac{1}{\varepsilon}} \tag{2.9}$$

Consequently,

$$\Rightarrow Q^* = N \cdot q^* = \left(\frac{\left(1 - \frac{\varepsilon}{N}\right) \cdot (1 - \gamma \cdot a)}{c} \right)^{\frac{1}{\varepsilon}} \qquad (2.10)$$

$$\Rightarrow p^* = Q^{*-\varepsilon} = \frac{c}{\left(1 - \frac{\varepsilon}{N}\right) \cdot (1 - \gamma \cdot a)}. \qquad (2.11)$$

The second order condition proves this solution is a maximum

$$\frac{\partial^2 \pi_n}{\partial^2 q_n} = \left(1 - \frac{\varepsilon}{N}\right) \cdot (-\varepsilon) \cdot (Q^*)^{-\varepsilon-1} \cdot (1 - \gamma \cdot a) = \frac{c^{\frac{\varepsilon+1}{\varepsilon}}}{\left(1 - \frac{\varepsilon}{N}\right)^{\frac{1}{\varepsilon}} \cdot (1 - \gamma \cdot a)^{\frac{1}{\varepsilon}}} \cdot (-\varepsilon) < 0$$

with $\left(1 - \frac{\varepsilon}{N}\right)^{\frac{1}{\varepsilon}} > 0$, since $0 < \varepsilon < 1$ and $N \geq 1$,

furthermore, $c^{\frac{\varepsilon+1}{\varepsilon}} > 0$, since $c > 0$, as well as

$(1 - \gamma \cdot a)^{\frac{1}{\varepsilon}} > 0$, since $0 < \gamma < 1$ and $0 < a < 1$.

Q.E.D.

Proof of Lemma 2.2

When *a=ā,* merchants make normal profits, i.e., zero profits from selling products paid for by the card. If the scheme would request any merchant usage fee *a*> *ā,* merchants would be better off not accepting the card. Thus,

$$\pi_n^{card} = \frac{Q}{N} \cdot Q^{-\varepsilon} \cdot (1-\overline{a}) - \frac{Q}{N} \cdot c \overset{!}{=} 0 \text{ with } N \geq 1 \text{ and } Q > 0, \tag{2.12}$$

$$\text{with } Q^* = \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-\gamma \cdot a)}{c} \right)^{\frac{1}{\varepsilon}} \text{ from Eq (2.10),}$$

$$\Rightarrow \overline{a} = \frac{\varepsilon}{N - N \cdot \gamma + \varepsilon \cdot \gamma}. \tag{2.13}$$

Q.E.D.

Proof of Lemma 2.3

The unitary payment card network will maximize profits Π over the transaction fee *a*:

$$\max_a \ \Pi \tag{2.18}$$

$$\text{with } \Pi = p \cdot Q \cdot \gamma \cdot a - C = Q^{1-\varepsilon} \cdot \gamma \cdot a - C = \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-\gamma \cdot a)}{c} \right)^{\frac{1-\varepsilon}{\varepsilon}} \cdot \gamma \cdot a - C, \tag{2.17}$$

The first order condition of the networks' profits is:

$$\frac{\partial \Pi}{\partial a} = \frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(-\gamma)}{c} \cdot \frac{1-\varepsilon}{\varepsilon} \cdot \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-\gamma\cdot\hat{a})}{c} \right)^{\frac{1-2\varepsilon}{\varepsilon}} \cdot \gamma \cdot \hat{a} + \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-\gamma\cdot\hat{a})}{c} \right)^{\frac{1-\varepsilon}{\varepsilon}} \cdot \gamma \overset{!}{=} 0 \cdot$$

Solving the first order condition leads to:

$$\Rightarrow \hat{a} = \frac{\varepsilon}{\gamma}. \tag{2.19}$$

The second order condition also proves this solution to be a maximum:

$$\frac{\partial^2 \Pi}{\partial^2 a} = \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(-\gamma)}{c} \right)^2 \cdot \frac{1-\varepsilon}{\varepsilon} \cdot \frac{1-2\varepsilon}{\varepsilon} \cdot \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-\gamma\cdot a)}{c} \right)^{\frac{1-3\varepsilon}{\varepsilon}} \cdot \gamma \cdot a + 2 \cdot \frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(-\gamma)}{c} \cdot \frac{1-\varepsilon}{\varepsilon} \cdot \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-\gamma\cdot a)}{c} \right)^{\frac{1-2\varepsilon}{\varepsilon}} \cdot \gamma$$

with $\hat{a} = \frac{\varepsilon}{\gamma}$

$$\Rightarrow \frac{\partial^2 \Pi}{\partial^2 a} = \left(\frac{1}{c}\right)^{\frac{1-\varepsilon}{\varepsilon}} \cdot \left(1-\frac{\varepsilon}{N}\right)^{\frac{1-\varepsilon}{\varepsilon}} \cdot (1-\varepsilon)^{\frac{1-2\varepsilon}{\varepsilon}} \cdot \gamma^2 \cdot \frac{1}{\varepsilon} \cdot (-1) < 0$$

with $\left(\frac{1}{c}\right)^{\frac{1-\varepsilon}{\varepsilon}} > 0$, since $0 < c < \infty$ and $\left(1-\frac{\varepsilon}{N}\right)^{\frac{1-\varepsilon}{\varepsilon}} > 0$, since $1-\frac{\varepsilon}{N} > 0$, as well as $(1-\varepsilon)^{\frac{1-2\varepsilon}{\varepsilon}} > 0$, since $1-\varepsilon > 0$, furthermore $\gamma^2 > 0$ and $\frac{1}{\varepsilon} > 0$, since $\varepsilon > 0$.

Q.E.D.

Proof of Lemma 2.2

N=N* is the competitive condition of merchants where (i) merchants make zero profits (or only normal profits) from sales paid for by a card; therefore, $a = \overline{a}$ with N=N*, and (ii) since with N=N* the network will be able to achieve $\hat{a}$ (according to Eq. 2.19). Consequently, in the case of a take-it-or-leave-it offer, and with N=N*,

$$\hat{a} = \overline{a},$$

$$\text{with } \overline{a} = \frac{\varepsilon}{N - N \cdot \gamma + \varepsilon \cdot \gamma} \text{ (Eq. 2.13) and } \hat{a} = \frac{\varepsilon}{\gamma} \text{ (Eq. 2.19);}$$

$$\Rightarrow N^* = \frac{1-\varepsilon}{\frac{1}{\gamma} - 1}. \tag{2.21}$$

Q.E.D.

Finally, N* proves to be independent from the network that makes a take-it-or-leave-it offer, or the case of a Nash bargaining situation among merchants and the respective network, since:

$$\hat{a} = \overline{a},$$

$$\text{with } \overline{a} = \alpha_i \cdot \frac{\varepsilon}{N - N \cdot \gamma + \varepsilon \cdot \gamma} \text{ and } \hat{a} = \alpha_i \cdot \frac{\varepsilon}{\gamma};$$

$$\Rightarrow N^* = \frac{1-\varepsilon}{\frac{1}{\gamma} - 1}. \tag{2.21}$$

Q.E.D.

Furthermore, consider that:

$$\frac{\partial \hat{a}}{\partial N} = 0 \text{ for all } N \leq N^*, \tag{2.29}$$

$$\text{and } \frac{\partial \overline{a}}{\partial N} = \frac{-\varepsilon \cdot (1-\gamma)}{\left(N - N \cdot \gamma + \varepsilon \cdot \gamma\right)^2} < 0 \text{, since } 0 < \varepsilon < 1,\ 1-\gamma > 0 \text{ and } 0 < \gamma < 1. \tag{2.30}$$

Thus, we know that $\hat{a} < \overline{a}$ for all N<N* and $\hat{a} > \overline{a}$ for all N>N*. Consequently, a payment card system will define the merchant fee such that $a = \hat{a}$ for all N<N* and $a = \overline{a}$ for all N<N*.

Proofs of Section 3

Proof of Lemma 3.2

The first order condition of the profit function is:

$$\frac{\partial \pi_1}{\partial p_1} = \left(A - 2Bp_1 + \tau p_2\right)\left(1 - \gamma a\right) + Bc = 0 \qquad (3.23)$$

Because of the symmetry of merchants, we know they will set the same price $p_1 = p_2 = p$. Solving for p provides the result.

Proof of Proposition 3.1

First Statement: $\frac{\partial a^*}{\partial A} = -\frac{\frac{\partial^2 \Pi}{\partial A \partial a}}{\frac{\partial^2 \Pi}{\partial a^2}} > 0$

The denominator is a second order condition of a maximization problem and thus always negative. It therefore suffices to find $\frac{\partial^2 \Pi}{\partial A \partial a} > 0$. We have:

$$\frac{\partial^2 \Pi}{\partial A \partial a} = \left[\frac{\partial \sigma}{\partial A} + \frac{\frac{\partial \phi}{\partial A}}{1 - \gamma a} - \frac{\frac{\partial \xi}{\partial A}}{\left(1 - \gamma a\right)^2}\right]\gamma + \left[\frac{\gamma \frac{\partial \phi}{\partial A}}{\left(1 - \gamma a\right)^2} - \frac{2\gamma \frac{\partial \xi}{\partial A}}{\left(1 - \gamma a\right)^3}\right]\gamma a \qquad (3.24)$$

and as $\frac{\partial \sigma}{\partial A} = \frac{4AB}{\left(2B - \tau\right)^2} > 0$, $\frac{\partial \phi}{\partial A} = \frac{2B\tau c}{\left(2B - \tau\right)^2} > 0$, and $\frac{\partial \xi}{\partial A} = 0$ the statement follows.

Second Statement: $\frac{\partial a^*}{\partial B} = -\frac{\frac{\partial^2 \Pi}{\partial B \partial a}}{\frac{\partial^2 \Pi}{\partial a^2}} < 0$

Again, it is enough to find $\frac{\partial^2 \Pi}{\partial B \partial a} < 0$. This provides:

$$\frac{\partial^2 \Pi}{\partial B \partial a} = \left[\frac{\partial \sigma}{\partial B} + \frac{\frac{\partial \phi}{\partial B}}{1-\gamma a} - \frac{\frac{\partial \xi}{\partial B}}{(1-\gamma a)^2} \right] \gamma + \left[\frac{\gamma \frac{\partial \phi}{\partial B}}{(1-\gamma a)^2} - \frac{2\gamma \frac{\partial \xi}{\partial B}}{(1-\gamma a)^3} \right] \gamma a < 0 \qquad (3.25)$$

and $\frac{\partial \sigma}{\partial B} = -\frac{2\tau A^2}{(2B-\tau)^3} < 0$, $\frac{\partial \phi}{\partial B} = 2A\tau c \frac{-\tau - B}{(2B-\tau)^3} < 0$, and

$$\frac{\partial \xi}{\partial B} = \frac{2Bc^2}{(2B-\tau)^3}\left(\tau^2 + 2B^2 - 3B\tau\right) > 0 .$$

Third Statement: $\frac{\partial a^*}{\partial \tau} = -\frac{\frac{\partial^2 \Pi}{\partial \tau \partial a}}{\frac{\partial^2 \Pi}{\partial a^2}} > 0$

For this to be true, we need the numerator to be positive $\frac{\partial^2 \Pi}{\partial \tau \partial a} > 0$. This is given by

$$\frac{\partial^2 \Pi}{\partial \tau \partial a} = \left[\frac{\partial \sigma}{\partial \tau} + \frac{\frac{\partial \phi}{\partial \tau}}{1-\gamma a} - \frac{\frac{\partial \xi}{\partial \tau}}{(1-\gamma a)^2} \right] \gamma + \left[\frac{\gamma \frac{\partial \phi}{\partial \tau}}{(1-\gamma a)^2} - \frac{2\gamma \frac{\partial \xi}{\partial \tau}}{(1-\gamma a)^3} \right] \gamma a > 0 \qquad (3.26)$$

because we have $\frac{\partial \sigma}{\partial \tau} = \frac{4A^2 B}{(2B-\tau)^3} > 0$, $\frac{\partial \phi}{\partial \tau} = \frac{2ABc}{(2B-\tau)^2} + \frac{4ABc\tau}{(2B-\tau)^3} > 0$, and

$$\frac{\partial \xi}{\partial \tau} = -\frac{2B^2 c^2 \tau}{(2B-\tau)^3} < 0 .$$

Fourth Statement: $\frac{\partial a^*}{\partial \gamma} = -\frac{\frac{\partial^2 \Pi}{\partial \gamma \partial a}}{\frac{\partial^2 \Pi}{\partial a^2}} < 0$

The derivative with regard to γ yields:

$$\frac{\partial^2\Pi}{\partial\gamma\partial a}= \left[\sigma+\frac{\phi}{1-\gamma a}-\frac{\xi}{(1-\gamma a)^2}\right]\gamma+\left[\frac{a\phi}{(1-\gamma a)^2}-\frac{2a\xi}{(1-\gamma a)^3}\right]\gamma+ \left[\frac{\gamma\phi}{(1-\gamma a)^2}-\frac{2\gamma\xi}{(1-\gamma a)^3}\right]a+\left[\frac{\phi(1+\gamma a)}{(1-\gamma a)^3}-\frac{2\xi(1+2\gamma a)}{(1-\gamma a)^4}\right]\gamma a \tag{3.27}$$

Denote the different parts of the sum as *I*, *II*, *III*, *IV* from left to right. Note for *I*: $\left[\sigma+\frac{\phi}{1-\gamma a}-\frac{\xi}{(1-\gamma a)^2}\right]\gamma=\frac{\Pi}{a\gamma}$. We then find that *I*+*III* gives:

$$\left[\sigma+\frac{\phi}{1-\gamma a}-\frac{\xi}{(1-\gamma a)^2}\right]\gamma+\left[\frac{\gamma\phi}{(1-\gamma a)^2}-\frac{2\gamma\xi}{(1-\gamma a)^3}\right]a=\frac{\partial\Pi}{\partial a}\frac{1}{\gamma} \tag{3.28}$$

If the payment network is setting the optimal merchant usage fee (in other words, if we have $\frac{\partial\Pi}{\partial a}=0$, then in Eq. (3.28) , $I+III=0$). As such, the sign of $\frac{\partial^2\Pi}{\partial\gamma\partial a}$ is determined by *II*+*IV:*

$$\left[\frac{a\phi}{(1-\gamma a)^2}-\frac{2a\xi}{(1-\gamma a)^3}\right]\gamma+\left[\frac{\phi(1+\gamma a)}{(1-\gamma a)^3}-\frac{2\xi(1+2\gamma a)}{(1-\gamma a)^4}\right]\gamma a<0 \tag{3.29}$$

This has to be negative because *I*>*II:*

$$\left[\sigma+\frac{\phi}{1-\gamma a}-\frac{\xi}{(1-\gamma a)^2}\right]\gamma>\left[\frac{a\phi}{(1-\gamma a)^2}-\frac{2a\xi}{(1-\gamma a)^3}\right]\gamma \tag{3.30}$$

And *III*>*IV:*

$$\left[\frac{\gamma\phi}{(1-\gamma a)^2}-\frac{2\gamma\xi}{(1-\gamma a)^3}\right]a>\left[\frac{\phi(1+\gamma a)}{(1-\gamma a)^3}-\frac{2\xi(1+2\gamma a)}{(1-\gamma a)^4}\right]\gamma a \tag{3.31}$$

In this case, we have $I + III = 0$, and $II < I$, as well as $IV < III$. Consequently, the full term *I+II+III+IV* has to be negative: $\frac{\partial^2 \Pi}{\partial \gamma \partial a} < 0$. This directly yields the claim.

Proof of Proposition 3.3

As a matter of preference, there exists an alternative way of presenting the above results starting from the inverse demand functions:

$$p_1 = \alpha_1 - \beta_1 q_1 - \theta q_2 \tag{3.32}$$

$$p_2 = \alpha_2 - \beta_2 q_2 - \theta q_1 \tag{3.33}$$

Using these, our exogenous variables from Eq. (3.1) and Eq. (3.2) were dependent on α, β, and θ, with $A_i = \frac{\alpha_i \beta_j - \alpha_i \theta}{\delta}$, $B_i = \frac{\beta_j}{\delta}$, $\tau = \frac{\theta}{\delta}$, and $\delta = \beta_i \beta_j - \theta^2$ with $i, j = 1{,}2$ and $i \neq j$. Thus, it is more clear that θ is a function of (A, B, τ). Without loss of generality, we assume that $\beta \geq \theta$; i.e., the price that a merchant charges has at least as much influence on the purchasing decision as the price of the other merchant. The demand function allows for products to account for any relationship in between being full substitutes or full complements based on the parameter *θ*. For *θ* =0, the two products are completely independent and each merchant is a monopolist, whereas for *θ*>0 (*θ*<0), the products are substitutes (complements). Products become perfect substitutes when $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2 = \theta$.

Because of the symmetry of merchants, we know that $\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 = \beta_2 = \beta$.

Because of the above symmetric merchants, the demand function (Eq. 3.1) can be re-written as:

$$q_1 = \frac{(\alpha - p_1)\beta - (\alpha - p_2)\theta}{\beta^2 - \theta^2}. \tag{3.34}$$

From this we can derive the price elasticity of demand

$$\left|\varepsilon_{q_1,p_1}\right| = \frac{p_1\beta}{(\alpha - p_1)\beta - (\alpha - p_2)\theta} \tag{3.35}$$

Note that the denominator has to be postive, otherwise Eq. (3.34) is negative and there would be a negative demand for q_1. Taking the derivative with regard to θ then yields:

$$\frac{\partial|\varepsilon|}{\partial\theta} = \frac{p_1\beta(\alpha - p_2)}{[(\alpha - p_1)\beta - (\alpha - p_2)\theta]^2} > 0 \tag{3.36}$$

Consequently, $$\frac{\partial a^*}{\partial\theta} < 0, \tag{3.22}$$

since $\frac{\partial a^*}{\partial|\varepsilon|} < 0$ (Eq. 3.21) and $\frac{\partial|\varepsilon|}{\partial\theta} > 0$ (Eq. 3.36).

The corresponding intuition is that the optimal merchant usage fee is decreasing in the substitutability of the products. This is consistent with our results, since greater product substitutability increases in the price elasticity of demand.

Proofs of Section 4

Proof of Proposition 4.1

In the case of a multi-party payment network with two identical duopoly all-purpose banks, one such bank (here bank *B1*) has profits:

$$\Pi_{B1}(a_{B1},a_{B2}) = p(a_{B1},a_{B2})\cdot Q(a_{B1},a_{B2})\cdot s_{B1}\cdot a_{B1} - C_{B1}, \tag{4.9}$$

$$\text{with } p = Q^{-\varepsilon} \text{ (Eq. 4.1)}, \ Q^* = \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-\gamma\cdot a)}{c}\right)^{\frac{1}{\varepsilon}}, \tag{4.3}$$

$$\text{and } a_{2B} = a_{B1} = a_{B2}. \tag{4.8}$$

The profit-maximizing duopoly network bank maximizes profits Π_{B1} over the merchant usage fee:

$$\max_{a_{B1}} \Pi_{B1}(a_{B1},a_{B2}) = \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\left(1-\gamma\cdot\left(s_{B1}\cdot a_{B1}+s_{B2}\cdot a_{B2}\right)\right)}{c}\right)^{\frac{1-\varepsilon}{\varepsilon}}\cdot\frac{\gamma}{2}\cdot a_{B1} - C_j,$$

by substitution of Eq. (4.1), Eq. (4.3) and Eq. (4.7) into Eq. (4.9).

The first derivative is:

$$\frac{\partial\Pi_{B1}}{\partial a_{B1}} = \frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\frac{-\gamma}{2}}{c}\cdot\frac{1-\varepsilon}{\varepsilon}\cdot\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\left(1-\frac{\gamma}{2}\cdot(a_{B1}+a_{B2})\right)}{c}\right)^{\frac{1-2\cdot\varepsilon}{\varepsilon}}\cdot\frac{\gamma}{2}\cdot a_{B1} + \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\left(1-\frac{\gamma}{2}\cdot(a_{B1}+a_{B2})\right)}{c}\right)^{\frac{1-\varepsilon}{\varepsilon}}\cdot\frac{\gamma}{2}$$

with $a_{2B} = a_{B1} = a_{B2}$ (Eq. 4.8), Thus, the first order condition is:

$$\frac{\partial \Pi_{B1}}{\partial a_{B1}} = \frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\frac{-\gamma}{2}}{c}\cdot\frac{1-\varepsilon}{\varepsilon}\cdot\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\left(1-\frac{\gamma}{2}\cdot a_{2B}\right)}{c}\right)^{\frac{1-2\cdot\varepsilon}{\varepsilon}}\cdot\frac{\gamma}{2}\cdot a_{2B}+\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\left(1-\frac{\gamma}{2}\cdot a_{2B}\right)}{c}\right)^{\frac{1-\varepsilon}{\varepsilon}}\cdot\frac{\gamma}{2}\stackrel{!}{=}0$$

$$\Leftrightarrow \hat{a}_{2B} = \frac{\varepsilon}{\gamma}\cdot\frac{2}{1+\varepsilon}. \qquad (4.10)$$

The second order condition proves this solution to be a maximum:

$$\frac{\partial^2 \Pi_{B1}}{\partial^2 a_{B1}} = \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\frac{\gamma}{2}}{c}\right)^2\cdot\frac{1-\varepsilon}{\varepsilon}\cdot\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\left(1-\gamma\cdot\left(a_{B1}+a_{B2}\right)\right)}{c}\right)^{\frac{1-3\cdot\varepsilon}{\varepsilon}}\cdot\left(\frac{\gamma}{2}\cdot\frac{1-2\cdot\varepsilon}{\varepsilon}\cdot a_{B1}-2\cdot\left(1-\gamma\cdot\left(a_{B1}+a_{B2}\right)\right)\right)$$

again, with $a_{2B} = a_{B1} = a_{B2}$ (Eq. 4.8),

$$\frac{\partial^2 \Pi_{B1}}{\partial^2 a_{B1}} = \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\frac{\gamma}{2}}{c}\right)^2\cdot\frac{1-\varepsilon}{\varepsilon}\cdot\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\left(1-\gamma\cdot a_{2B}\right)}{c}\right)^{\frac{1-3\cdot\varepsilon}{\varepsilon}}\cdot\left(\frac{\gamma}{2}\cdot\frac{1-2\cdot\varepsilon}{\varepsilon}\cdot a_{2B}-2\cdot\left(1-\gamma\cdot a_{2B}\right)\right)$$

with $\hat{a}_{2B} = \frac{\varepsilon}{\gamma}\cdot\frac{2}{1+\varepsilon}$ (Eq. 4.10),

$$\frac{\partial^2 \Pi_{B1}}{\partial^2 a_{B1}} = \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\frac{\gamma}{2}}{c} \right)^2 \cdot \frac{1-\varepsilon}{\varepsilon} \cdot \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\left(1-\gamma\cdot\frac{\varepsilon}{\gamma}\cdot\frac{2}{1+\varepsilon}\right)}{c} \right)^{\frac{1-3\cdot\varepsilon}{\varepsilon}} \cdot \left(\frac{\gamma}{2}\cdot\frac{1-2\cdot\varepsilon}{\varepsilon}\cdot\frac{\varepsilon}{\gamma}\cdot\frac{2}{1+\varepsilon} - 2\cdot\left(1-\gamma\cdot\frac{\varepsilon}{\gamma}\cdot\frac{2}{1+\varepsilon}\right)\right)$$

$$\frac{\partial^2 \Pi_{B1}}{\partial^2 a_{B1}} = \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\frac{\gamma}{2}}{c} \right)^2 \cdot \frac{1-\varepsilon}{\varepsilon} \cdot \left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\left(1-\frac{2\cdot\varepsilon}{1+\varepsilon}\right)}{c} \right)^{\frac{1-3\cdot\varepsilon}{\varepsilon}} \cdot \frac{1}{1+\varepsilon}\cdot(-1) < 0,$$

with $\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\frac{\gamma}{2}}{c} \right)^2 > 0$,

since $0 < \varepsilon < 1$ and $N > 1$, thus $1-\frac{\varepsilon}{N} > 0$, and since $\gamma > 0$ and $c > 0$,

with $\frac{1-\varepsilon}{\varepsilon} > 0$ since $0 < \varepsilon < 1$,

with $\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot\left(1-\frac{2\cdot\varepsilon}{1+\varepsilon}\right)}{c} \right)^{\frac{1-3\cdot\varepsilon}{\varepsilon}} > 0 \,\forall\, 0 < \varepsilon < \frac{1}{4}$,

since $1-\frac{\varepsilon}{N} > 0$, $1-\frac{2\cdot\varepsilon}{1+\varepsilon} > 0 \forall \varepsilon < 1$, $c > 0$, and $\frac{1-3\cdot\varepsilon}{\varepsilon} > 1 \forall \varepsilon < \frac{1}{4}$,

as well as with $\frac{1}{1+\varepsilon} > 0$ since $0 < \varepsilon$.

Q.E.D.

Proof of Equation (4.12)

$N=N_i$* is the competitive condition of merchants where (i) merchants make zero profits from sales paid for by card; therefore, $a=\overline{a}$ with $N=N_i$*, and (ii) since $N=N_i$*, the network will still be able to achieve $\hat{a}$ (Eq. 4.11). Consequently, in the case of a take-it-or-leave-it offer, and with $N=N_i$*,

$$\hat{a}_{2B}=\overline{a}\ ,$$

with $\overline{a}=\dfrac{\varepsilon}{N-N\cdot\gamma+\varepsilon\cdot\gamma}$ (Eq. 7) and $\hat{a}_{2B}=\dfrac{\varepsilon}{\gamma}\cdot\dfrac{2}{1+\varepsilon}$ (Eq. 4.11),

$$\Rightarrow \frac{\varepsilon}{\gamma}\cdot\frac{2}{1+\varepsilon}=\frac{\varepsilon}{N_{2B}*-N_{2B}*\cdot\gamma+\varepsilon\cdot\gamma}\ ,$$

$$\Leftrightarrow \frac{\gamma}{2}+\frac{\gamma\cdot\varepsilon}{2}=N_{2B}*-N_{2B}*\cdot\gamma+\varepsilon\cdot\gamma\ ,$$

$$\Leftrightarrow \frac{\gamma}{2}\cdot(1-\varepsilon)=N_{2B}*\cdot(1-\gamma),$$

$$\Leftrightarrow N_{2B}*=\frac{\gamma}{2}\cdot\frac{1-\varepsilon}{1-\gamma}\ . \qquad (4.12)$$

Q.E.D.

Proof of Lemma 4.1

Consider specialist banks and the sequential scenario, that is, when the issuer and the acquirer bargain over the interchange fee, and hereafter, the acquirer, sets the merchant usage fee. Thus, the acquirer maximizes profits with the interchange fee *f* being given:

$$\Pi(a) = p(a) \cdot Q(a)(a - f)\gamma - C_A, \qquad (4.13)$$

As before, with $p = Q^{-\varepsilon}$ (Eq. 4.1), $Q^* = \left(\frac{\left(1 - \frac{\varepsilon}{N}\right) \cdot (1 - \gamma \cdot a)}{c} \right)^{\frac{1}{\varepsilon}}$ (Eq. 4.3),

$$\max_a \Pi(a) = \left(\frac{\left(1 - \frac{\varepsilon}{N}\right) \cdot (1 - \gamma a)}{c} \right)^{\frac{1-\varepsilon}{\varepsilon}} \gamma(a - f) - C_A.$$

The first order condition is:

$$\frac{\partial \Pi}{\partial a} = -\frac{\left(1 - \frac{\varepsilon}{N}\right)}{c} \frac{1-\varepsilon}{\varepsilon} \cdot \left(\frac{\left(1 - \frac{\varepsilon}{N}\right) \cdot (1 - a\gamma)}{c} \right)^{\frac{1-2\cdot\varepsilon}{\varepsilon}} (a - f)\gamma^2 + \left(\frac{\left(1 - \frac{\varepsilon}{N}\right) \cdot (1 - a\gamma)}{c} \right)^{\frac{1-\varepsilon}{\varepsilon}} \gamma \overset{!}{=} 0$$

Solving for *a* yields

$$\hat{a} = (1 - \varepsilon) f + \frac{\varepsilon}{\gamma}. \qquad (4.14)$$

Local sufficient conditions: the second order derivative is given by:

$$\frac{\partial^2\Pi}{\partial^2 a}=\left(\frac{\left(1-\frac{\varepsilon}{N}\right)}{c}\right)^2\frac{1-\varepsilon}{\varepsilon}\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-a\gamma)}{c}\right)^{\frac{1-3\cdot\varepsilon}{\varepsilon}}(a-f)\gamma^3-2\frac{1-\varepsilon}{\varepsilon}\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-a\gamma)}{c}\right)^{\frac{1-2\varepsilon}{\varepsilon}}\gamma^2\frac{\left(1-\frac{\varepsilon}{N}\right)}{c}$$

$$\frac{\partial^2\Pi}{\partial^2 a}=\frac{1-\varepsilon}{\varepsilon}\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-a\gamma)}{c}\right)^{\frac{1-2\varepsilon}{\varepsilon}}\gamma^2\frac{\left(1-\frac{\varepsilon}{N}\right)}{c}\left(\left(\frac{\left(1-\frac{\varepsilon}{N}\right)\cdot(1-a\gamma)}{c}\right)^{-1}(a-f)\gamma\frac{\left(1-\frac{\varepsilon}{N}\right)}{c}-2\right)$$

Substituting (4.14) for a and some simplifications yields:

$$-\frac{1-\varepsilon}{\varepsilon}\gamma^2\left(\frac{\left(1-\frac{\varepsilon}{N}\right)}{c}\right)^{\frac{1-\varepsilon}{\varepsilon}}\left[(f\gamma-1)(\varepsilon-1)\right]^{\frac{1-2\varepsilon}{\varepsilon}}\frac{3\varepsilon-2}{\varepsilon-1}.$$

All terms in this derivative are positive with the qualification that for the last term $\varepsilon<\frac{2}{3}$ must be imposed. Hence, local concavity is ensured.

Proof of Proposition 4.2

Consider Nash bargaining over the interchange fee between the issuer and the acquirer, i.e.,

$$\max_f N=\left[\left(\frac{\left(1-\frac{\varepsilon}{N}\right)(1-\gamma a(f))}{c}\right)^{\frac{1-\varepsilon}{\varepsilon}}(a(f)-f)\gamma\right]^{1-\alpha}\left[\left(\frac{\left(1-\frac{\varepsilon}{N}\right)(1-\gamma a(f))}{c}\right)^{\frac{1-\varepsilon}{\varepsilon}}f\gamma\right]^{\alpha}$$

The derivative $\frac{\partial N}{\partial f}$ is given by:

$$\frac{\partial N}{\partial f} = -\frac{1-\varepsilon}{\varepsilon}\frac{f\gamma}{1-\gamma a(f)}\frac{da}{df} + (1-\alpha)\left(\frac{da}{df}-1\right)\frac{f}{a(f)-f} + \alpha$$

From Eq. (4.14), we know that:

$$a(f) = (1-\varepsilon)f + \frac{\varepsilon}{\gamma} \text{ and therefore:}$$

$$\frac{\partial a}{\partial f} = (1-\varepsilon).$$

Inserting $\partial N/\partial f = 0$ in the first-order condition yields the following, with some simplifications:

$$\frac{1}{\varepsilon}\frac{f\gamma - \alpha\varepsilon}{f\gamma - 1} = 0,$$

$$\Leftrightarrow f_{Seq} = \alpha\frac{\varepsilon}{\gamma}. \qquad (4.16)$$

The second derivative is:

$$\frac{\partial^2 N}{\partial^2 f} = (\alpha\varepsilon - 1)\frac{1}{\varepsilon}\frac{\gamma}{(f\gamma-1)^2} \leq 0,$$

which is negative, because $\alpha\varepsilon \leq 1$. Hence, concavity is proven.

Proof of Lemma 4.2 and Proposition 4.4

The scenario of simultaneous bargaining can be represented by the maximization of the Nash-product over variables *a* and *f*:

$$\max_{a,f} N = \left[\left(\frac{\left(1-\frac{\varepsilon}{N}\right)(1-\gamma a)}{c} \right)^{\frac{1-\varepsilon}{\varepsilon}} (a-f)\gamma \right]^{1-\alpha} \left[\left(\frac{\left(1-\frac{\varepsilon}{N}\right)(1-\gamma a)}{c} \right)^{\frac{1-\varepsilon}{\varepsilon}} f\gamma \right]^{\alpha}$$

This results in two first-order conditions:

$$-(1-\alpha)\frac{f}{a-f} - \alpha = 0$$

$$\Leftrightarrow f_{Sim} = \alpha a \, , \tag{4.20}$$

$$\text{and } (1-\alpha)\frac{f}{a-f} - \frac{f\gamma}{1-a\gamma}\frac{1-\varepsilon}{\varepsilon} = 0 \, .$$

Inserting *f* for the merchant usage fee yields:

$$\hat{a}_{Sim} = \frac{\varepsilon}{\gamma} \, .$$

Thus,

$$a_{Sim}^{N} = \min\left\{ \frac{\varepsilon}{\gamma} ; \frac{\varepsilon}{N - N\gamma + \varepsilon\gamma} \right\} \tag{4.21}$$

Together with the cut-off value $\bar{a}$, Eq. (4.6) results in the value given in Eq. (4.20). The Hessian of the Nash-product is given by:

$$N''(a,f)=\begin{pmatrix} -(1-\alpha)\dfrac{a}{(a-f)^2} & (1-\alpha)\dfrac{f}{(a-f)^2} \\ (1-\alpha)\dfrac{a}{(a-f)^2}-\dfrac{1-\varepsilon}{\varepsilon}\dfrac{\gamma}{1-a\gamma} & -(1-\alpha)\dfrac{f}{(a-f)^2}-\dfrac{1-\varepsilon}{\varepsilon}\dfrac{f\gamma^2}{(1-a\gamma)^2} \end{pmatrix}$$

Negative values on the diagonal ensure concavity. The determinant of the Hessian is found to be positive:

$$\det=(1-\alpha)\frac{1-\varepsilon}{\varepsilon}\frac{f\gamma}{(a-f)^2(a\gamma-1)^2}\geq 0\,.$$