

Lydian Payments Journal

Volume 1 | Issue 1 | November 2009



In This Issue

1	Introduction David Evans
3	SEPA: How Come and What's Up? Joris Barbas and Wim Westerman
12	A Review of Payment Card Economics Wilko Bolt and Sujit "Bob" Chakravorti
24	What Does the Credit Card Market Have In Common with a Peacock? Joshua M. Frank
41	Marketing Credit Cards in Postcommunist Russia Alya Guseva
51	Who Owns Mobile Money? Ari Hyytinen and Tuomas Takalo
56	Two-Sided Bank Card Payment Networks and Public Policy David VanHoose

Lydian Payments Journal

The Lydian Payments Journal publishes articles from thought leaders across the globe on one of the most important industries in the world — payments, the industry which makes trade, the source of all economic prosperity, possible.

Titled after the Kingdom of Lydia, a region that is now part of eastern Turkey and which is attributed with inventing coinage in 600 BC, the Lydian Payments Journal shapes and chronicles this important sector by focusing on a spectrum of topics: from policy issues such as competition, consumer protection, and interchange to disruptive innovation such as social lending, remittance products, and mobile commerce.

This e-journal leverages the power of the Web to provide germane and timely thought leadership to a broad and relevant audience. The online format also works to fuel interaction between readers and authors, extending the dialogue beyond just the articles. Visit the Lydian Payments Journal online at www.pymnts.com/journal.

For information about contributing to The Pymnts Journal, please contact editorial@pymnts.com



Introducing the Lydian Payments Journal

By David S. Evans[†]

The Lydian Payments Journal publishes articles from thought leaders on one of the most important industries in the world—an industry that makes trade, the source of almost all economic prosperity, possible. That will seem like an outrageous claim to some people. Making and receiving payments is so routine and trivial that most take it for granted. But flash back three millennia to the days when a lot of effort went into buying a chicken. Our ancestors had to figure out what they could offer the chicken owner in trade. Maybe a shirt or a clay pot, or as trading became more sophisticated a slug of some precious metal.

The Kingdom of Lydia, a region that is now part of eastern Turkey, started the payments industry at least in their part of the world in around 600 BC. The government stamped out a coin with a particular metal content and therefore value. It became a popular method of payment and soon created imitators. More innovations followed. All these innovations greased the way for trade.

In doing so the payments industry permitted the creation of vast wealth for society. Besides simply making it easier for people to exchange value it encouraged economic specialization and trade over ever expanding regions. The latest frontiers are e-commerce and mobile phones. Trading in some of the poorest parts of the world is now undergoing a revolution as a result of payments innovations made possible by wireless communications.

Today, a vast industry supports payments. It includes governments that make cash; networks that clear and settle transactions; banks that provide their customers with several payments devices; companies that issue credit cards; entities that help merchants take and collect payments in a variety of ways; manufacturers of point-of-sale terminals, ATM machines, and other hardware; software providers who sell applications that make many things work; and many more.

[†] David S. Evans is the author of *Paying with Plastic: The Digital Revolution in Buying and Borrowing* and has been a business advisor to many payment companies around the world. David is the founder of Market Platform Dynamics, a boutique consulting firm that helps businesses leverage economics and quantitative methods for growth and profit. David teaches part time at the University of Chicago where he is Lecturer and at the University College London where he is Visiting Professor and Executive Director of the Jevons Institute for Competition Law and Economics. David has a Ph.D. and undergraduate degree in economics from the University of Chicago. He also serves on the boards of several high-technology companies and is a longtime advisor to some of the largest platform-based companies in the world.



Until recently, the payments industry attracted relatively little serious thought, analysis, and discussion compared to its outsize economic importance. That has changed in the last decade. The new economics of two-sided markets has provided both an analytical framework for understanding the payment business and an incentive for apply their new tools to one of the most interesting two-sided businesses around. At the same time controversies over payment card interchange fees have resulted in a lot of attention paid to how payments are done.

An outpouring of work on strategy, economics, law and policy has come from the minds and pens of business, economic, legal and policy experts from around the world. If you are a payments professional, or a policymaker concerned with this critical industry, the Lydian Payments Journal is the place to come for cutting edge knowledge and commentary produced by the best thinkers from around the globe. If you are one of these thinkers the Lydian Payments Journal is a common place for you to present your research and exchange ideas with other experts from diverse disciplines and geographies who are also interested in payments as well as to make you and your work known to the broader payments community.

We hope you enjoy the first of many issues to come.



SEPA: How Come and What's Up?

By Joris Barbas[†] and Wim Westerman[§]

Introduction

The Single Euro Payments Area (SEPA) is gradually coming into force. The aim of SEPA is to harmonize the Euro payment market; citizens and corporations in Europe already are able to make Euro payments anywhere within Europe as easily as they make their national payments. With the first banks launching their SEPA Direct Debit in November 2009, the foundations of SEPA are laid. It is expected that SEPA will foster greater transparency and competition, resulting in a cheaper and more efficient payments system in Europe.

To put the development of SEPA opportunities for European firms in a broader context, we discuss the roots of liberalization and deregulation within the Euro Zone, which finally led to the idea of SEPA. This understanding is essential. Only in a historical context is it clear that SEPA is not an isolated event, but a small part of a much larger path leading towards the harmonization of the European payment market. Therefore, firms cannot just seize opportunities overnight in a one-size-fits-all way, but should assess the opportunities as they arise.

We will show the events leading to the launch of the Euro first, followed by a discussion of the European banking landscape at the turn of the century. We follow with the Euro payments infrastructure development at the time. After this, we describe the European Union regulation towards an integrated European payment area. Finally, we take a closer look at the potential for firms to benefit from competition between European payment service providers.

[†] Joris Barbas studied business administration at the University of Groningen, the Netherlands. Currently he is employed as a Business Analyst with Deloitte Financial Advisory Services in Amsterdam where he advises firms on a strategic SEPA implementation.

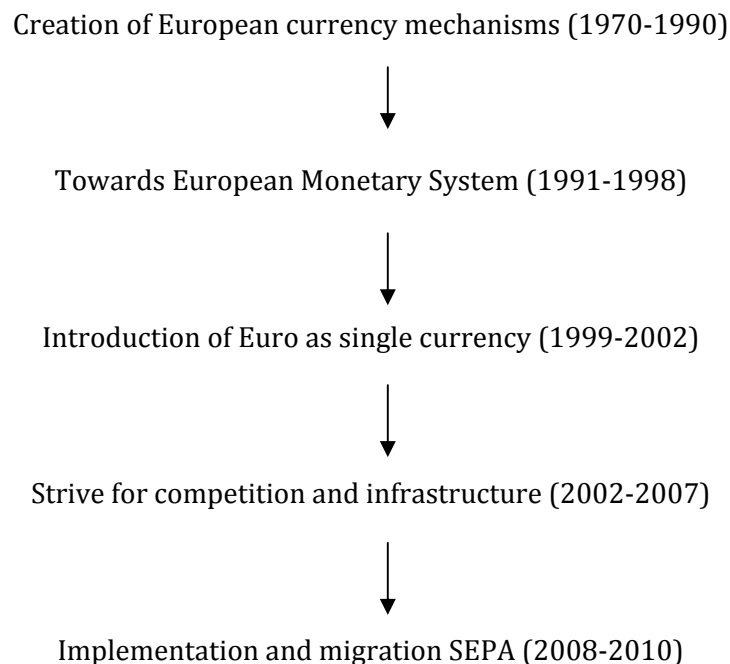
[§] Wim Westerman is a fellow of the University of Groningen's Center for International Banking, Insurance and Finance. Wim specializes in treasury management, capital budgeting, and management control in the Netherlands, the rest of Europe, and South Africa.



Towards the Introduction of the Euro

In the history of the development of an integrated European market for goods and services, the introduction of Economic and Monetary Union (EMU) and the Euro as a single currency in 1999–2002 has been thought of as a key factor and a precondition for a European payment market. (See Figure 1.) Yet, the harmonization of the European money market started earlier. In 1861, the Portuguese Carlos Morato Roma was the first man known to propose a reform of the European monetary system through the adoption of a common European currency.¹

Figure 1: Main Movers European Payment Market



Nevertheless, it took until 1970, when the Werner Report was published, for the wheels to be set in motion. This document on monetary integration recommended the development of the European Currency Unit (ECU), a centralized European credit policy, a unified capital market policy, and the gradual narrowing of exchange-rate fluctuations. In 1977, the European Commission (EC) reactivated its efforts at establishing a

¹ J. L. Cardoso, "A Proposal for a European Currency in 1861: The Forgotten Contribution of Carlos Morato Roma," *History of Political Economy* 36(2): 273-293 (summer 2004).



European Monetary Union (EMU). These efforts led to the birth of the EMU in 1979. At that time, few economists gave it much chance of survival.² It not only survived, however, it grew and prospered.

At the Hannover Summit of 1988, the European Council agreed to adopt the Single European Act, with the implicit objective of monetary integration. Mr. Delors, then President of the European commission, recommended a three-stage plan to coordinate more economic and monetary policies with the intention of creating a European single currency under the stewardship of a European Central Bank. The three stages in de Delors plan incorporate the coordinating economic policy, achieving economic convergence and finally the adoption of the euro, the EU's single currency. This report was known as the Delors Plan.

The first stage of the Delors Plan began in 1990, and the European Council convened in Maastricht in 1991. There the Heads of State signed the Maastricht Treaty, which set out the tough economic convergence criteria that had to be met to qualify for the single currency. Next, the currencies of the non-participating countries (pound sterling, the drachma, and the escudo) were brought into the exchange-rate mechanism (notably, the pound sterling left the mechanism not much later). The second phase of the Delors plan was fuelled in 1994, with the creation of the so-called European Monetary Institute (EMI).

It was decided in May 1998 that 11 EU countries (Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain) satisfied the convergence criteria. Greece met these criteria very soon afterwards, so that it was ready to introduce the Euro on January 1, 2002 as the twelfth country. Also in 1998, the European Central Bank (ECB), the successor of the EMI, was erected. It became operational in 1999.

The final stage of the Delors Plan, the adoption of the Euro, started on January 1, 1999. The parity rates were fixed irreversibly. As a consequence, national currencies were simply denominations of the Euro, but continued to be used as a matter of convenience. Full monetary union came into existence on January 1, 2002 when the Euro was introduced in its physical form and when the national currencies were taken out of circulation. At this time, Slovenia (2007), Cyprus (2008), Malta (2008), and Slovakia (2009) have joined the Euro Zone, with EU members Denmark and the Baltic countries legally pegging their currencies to the Euro.

There are many implications of the Euro adaptation.³ The Euro led to reduced transaction costs (due to less need for currency conversion and hedging), simplified cash management, reduced cost of borrowing,

² B. Eichengreen. and C. Wyplosz. "The Unstable EMS," *Brookings Papers on Economic Activity* 1 (1993): 51.



provided greater price competition, increased standardization of products, improved ability to compete, and simplified accounting and financial reporting. Yet, two items were not covered by the Euro. The first was the integration of the banking sector, which could be resolved by the second, the infrastructure in the Euro arena. To these items SEPA would contribute.

The Drive for Competition Between Banks

The “Lisbon Agenda” of 2000 aimed to make the EU the most dynamic and competitive knowledge-based economy in the world by 2010. One of the key goals of the Lisbon Agenda was to increase the competitiveness of both the European businesses and the financial sector. This competitiveness is achieved through the adoption of common technical standards and of common business practices. This effectively requires a Pan-European payment arena.

One of the general goals of the EU was to generate competition between service providers and have a level playing field across all aspects of the financial services industry throughout the Union. Introducing open, non-proprietary, and independent standards for an integrated payment area would generate new levels of competition. Payments services providers would seek more cost-efficient solutions and, thanks to common standards, customers would be better able to see pricing differentials between payment services providers.

To get a better understanding of the need for a single payment infrastructure in Europe, it is interesting to note the greatly segmented banking landscape of the Euro Zone in 2002. As an example, Table 1 shows the market shares of foreign branches and subsidiaries in the different national markets in 2002.

Table 1: Market Shares of Branches and Subsidiaries of Foreign Credit Institutions as a Percentage of Total Assets of Domestic Credit Institutions (as of year-end 2002)

Belgium	21.6
Germany	4.9
Greece	17.5
Spain	8.7
France	10.6
Ireland	37.3
Italy	3.7
Luxembourg	94.2
Netherlands	8.9
Austria	21.1

³ R. Grant. and S. Goldberg, “Euro Benefits for Cash Management,” *Journal of Corporate Accounting and Finance*. 18(1): 29-32 (2006).



Portugal	24.3
Finland	8.5
Eurozone	12.4

Source: European Commission (2004)

With Luxembourg as an exception, the countries all had a market share of 25% or less of foreign banks or other credit institutions. For the Euro Zone as a whole, foreign banks on average had a market share of less than 15%. This low average mainly reflects the very low market shares of foreign banks in the largest Euro Zone countries (Germany, France, Italy, and Spain) which were typically less than 10%. This lack of integration of the banking sector contrasted greatly with other sectors of the economy.⁴

Although EU regulation has become evermore effective, the presence of foreign banks has not grown remarkably in most (old) EU countries. Also, many EU countries already have a high concentration of banks (for example, Sweden, Belgium, and Portugal), with apparent additional growth from 2002 until recently. Furthermore, apart from a small but growing number of notable exceptions, M&A activity largely takes place within national boundaries. An early example to the contrary was the Swedish/Finnish Merita Nordbanken merger of 1998.

Euro Payment Infrastructure Development

Another item not covered by the introduction of the Euro was the EU payment infrastructure. Parallel to the introduction of the Euro, regulation was implemented to liberalize, deregulate, and harmonize financial markets and institutions with the aim of cutting the costs of cross-border payments. In the past, cash flows between European countries were obstructed by high transfer payments by banks.⁵ With intra-EU cash transactions, often both the payer and the receiver were charged. Maybe almost as bad, there were administrative burdens as well.

In a discussion paper⁶ published in 1990, the then European Commission made it clear that establishing an equitable payment infrastructure within Europe was critical to the success of its overall goal: *“The prospect of economic and monetary union which will lead to a further increase in intra-community trade makes it all the more urgent to ensure that Europe is equipped with structures which provide as cheap, as rapid and as reliable a payments service between different Member States of the Community as already exists within them.”*

⁴ P. De Grauwe. *Economics of Monetary Union*. (Oxford: Oxford University Press, 2005).

⁵ H. Von Eije. and W. Westerman, “Multinational Cash Management and Conglomerate Discounts in the Euro Zone,” *International Business Review*. 11(4):. 453-464 (2002).

⁶ European Commission, *Making Payments in the Internal Market*, COM (90) 447:1, 26 September 1990.



Shaping this vision into reality, several EU directives were implemented concerning the deregulation of the financial markets and institutions. In 1993, the second banking coordination directive⁷ came into force. Part of this directive was the "Single Banking License" that made it possible for banks that were authorized in one European country to establish affiliates in the other member states. This made it possible for foreign banks to become domestic and to issue financial instruments without having limitations because of their country of origin. This led to more international competition on markets that were previously considered to be only domestic, but competition still remained low.

Furthermore, in 1997, the EU imposed legislation to reduce the European banking tariffs,⁸ As a consequence, European banking tariffs were indeed reduced, but differences remained. Even after the introduction of the Euro, cross-border retail payments were still more costly than national ones.⁹ (See Table 2.) In some countries, there were significant entry barriers to banks moving into the sector, including access to payment systems and credit databases. Also, the cost of joining clearing systems for interbank payments in some countries was prohibitively high. To overcome these problems and develop a new pan-European payment infrastructure, the disparity between legislation, systems, and standards across Member States and the European financial community had to be eliminated.

Table 2: Cross-border Payment Cost-raising Factors

1. Anticompetitive behavior by banks
2. Structural problems within markets
3. Market failures

Source: De Grauwe (2005)

It was only when the introduction of the Euro was secured that the political attention turned to making improvements to a fully integrated payments market. It was understood that greater integration in the financial markets was crucial to the success of the Euro and the European Economic and Monetary Union. The Financial Services Action Plan (FSAP) was published by the EC in 1999 and endorsed by the Lisbon Council in 2000. The FSAP represents the European agenda to create an integrated financial marketplace by 2010. One of the key components of the FSAP and the Lisbon Agenda was the Single Euro Payments Area, or SEPA for short.

⁷ European Commission, *The Banking Coordination (Second Council Directive) Regulations 1992*.

⁸ European Commission, *Council Directive 97/5/EC on Cross Border Credit Transfers*, OJ EC L43, 14 February 1997.

⁹ Grant. and Goldberg,. "Euro Benefits," 29-32.



In order to achieve the given goal, in June 2002 the European Payment Council (EPC) was set up. The EPC is recognized as the decision-making and coordinating body of the European banking industry in relation to the implementation of SEPA. Its primary purpose is to support the creation of SEPA. The EPC is a self-regulatory body encompassing more than 65 European banks, including the three European credit sector associations (among them the European Associations of Corporate Treasurers) and the Euro Banking Association.

The EPC's primary deliverables are the development of SEPA Credit Transfer (SCT) and SEPA Direct Debit (SDD) "Schemes." These schemes are a set of core inter-bank rules, practices, and standards around which banks can compete to offer credit transfer and direct debit products to their customers and potential value-added services at a national community level. Furthermore the EPC is responsible for the development of the SEPA cards framework. This framework spells out high level principles and rules which, when implemented by banks, schemes, and other stakeholders, will enable European customers to use general purpose cards to make Euro payments and cash withdrawals throughout the SEPA area with the same ease and convenience as in their home country.

EU Regulation and ECB Pressure

As of 2001, the EU started to develop regulation to ensure that the vision of SEPA would be realized as the first regulation came into force.¹⁰ At that time, the basic principle was adopted that fees applied to payments must be the same regardless of sender or recipient account location within the EU. This regulation covers card payments, cash withdrawals, and credit transfers.

In 2005, after extensive discussions with the various SEPA stakeholders (banks, companies, trade associations, and consumers), the European Commission issued its proposal for the Payments Services Directive (PSD), the underlying legal framework for the creation of an EU-wide single market for payments. The PSD is the necessary legal framework in which all payment service providers will operate. In April 2007, the European Parliament adopted the PSD proposal and it came into force on December 25, 2007.¹¹

The first SEPA pan-European payment instruments became operational parallel to domestic instruments on January 28, 2008. The PSD had to be transposed into the national legislation by all EU Member States by

¹⁰ European Commission, *Regulation 2560/2001 of the European Parliament and of the Council of 19 December 2001 on Cross-border Payments in Euro*, OJ L 344, December 28, 2001.

¹¹ European Commission, *Directive 2007/64/EC of the European Parliament and of the Council of 13 November 2007 on Payment Services in the Internal Market*, OJ L 319, December 5, 2007.



November 2009 at the latest, and only Sweden will not be able to implement the PSD in 2009. Expectations are that SEPA could become fully operational by the end of 2010 and hopes are to attain a critical mass of SEPA payment instruments by the end of 2011. However, this is still a major challenge. In July 2009 – almost one and a half years after the SCT launch – only 4.4 % of credit transfers used SEPA standards.¹² The economic slowdown may influence the speed of parties implementing SEPA solutions. However, it also seems that some parties recognize that the opportunities of major efficiency gains and cost savings are quite welcome in these uncertain times.

In order to have the usage up, three types of payment have to be realized with SEPA:

- 1) **SEPA Credit Transfer (SCT)** – already accepted by most clearinghouses, but the uptake by the market is slow.
- 2) **Card payments** – to be effective a European card scheme is needed. At the moment, several card schemes have been developed, however, no initiative is underway to take one of the schemes to the market.
- 3) **SEPA Direct Debits (SDD)** – the first banks will be ready for SDD as of November 2009 and the SDD schemes will be fully realized by November 2010.

Furthermore, the development of frameworks for SEPA E-payments, M-Payments, and E-invoicing can speed up the use of SEPA payments. An E-payment is an internet banking payment, with payment confirmations arriving in real time. An M-payment is initiated on a mobile phone and could be used for both proximity and remote transactions.

In September 2009 the ECB identified six actions to be completed by all stakeholders (EU and national authorities, industry and users) over the next three years, following six priorities:¹³

- (1) Foster migration;
- (2) Increase awareness and promote SEPA products;
- (3) Design a sound legal environment and ensure compliance;
- (4) Promote innovation

¹² European Central Bank, "SEPA in Use," available at <http://www.ecb.int/paym/sepa/timeline/use/html/index.en.html>

¹³ European Commission, *Completing SEPA: a roadmap for 2009-2012*, COM(2009)471 Final, September 10, 2009.



(5) Achieve standardisation and interoperability; and

(6) Clarify and improve SEPA project governance.

SEPA Potential for Firms

Financial market integration is urgent for the EMU to function smoothly, mainly so that it can function as an insurance mechanism enabling adjustment to asymmetric shocks. Substantial impediments that have led to a situation in which the banking sector within the Euro Zone is among the least integrated sectors of the economies continue to exist.¹⁴ And since so many of the credit needs within the Euro Zone depend on banks, it also implies that the introduction of the Euro has failed to integrate an important part of the credit markets. However, the implementation of SEPA, and consequently the adoption of open industry standards may have a positive effect on the competition between financial service providers.

First of all, because of the open standards, it becomes easier for customers to switch between banks. Indeed because of the open standards, the consumer-to-bank communication is the same throughout Europe and no changes are needed in the customer's software. Secondly, even more important, in principle, it is possible to reach any bank account in the Euro Zone with only one bank relation/account. Without SEPA, in order to properly serve business partners abroad and to provide them with local and familiar payment services, a firm had to maintain at least one bank account per country where it was active. Because of the open standards of SEPA this is not needed anymore. Thirdly, yet not part of the basic SEPA agenda, automation of accounts payables and receivables will change the payment industry. Standardized interfaces and payment related messages, E-invoicing, and E-reconciliation will save a lot of money.

In short, competition will increase because payments institutions will be able to offer their propositions, products, and services throughout SEPA more easily, which will provide customers with more choice and therefore lower prices and lower internal administrative costs. Although the implementation of and migration to SEPA has been slow, SEPA is with us and will continue towards full realization (albeit perhaps not by the end of 2010). Firms that adapt to the new standards in time may reap strategic benefits in terms of cost and speed. This may even enable greater revenues, if suppliers and customers also enter the payment chain.

¹⁴ De Grauwe, *Economics of Monetary Union*.



A Review of Payment Card Economics

Wilko Bolt[†] and Sujit Chakravorti[§]

The proliferation of payment cards—that is, debit, credit, and prepaid cards—has dramatically changed the way we shop and merchants sell goods and services. Today, payment cards are indispensable in most advanced economies. For Europe, Bolt and Humphrey report that the number of card payments increased by 140 percent across 11 European countries (Belgium, Denmark, France, Finland, Germany, Italy, the Netherlands, Norway, Spain, Sweden, and the United Kingdom) during the period 1987–2004.¹ Recently, some merchants have started to accept only card payments for safety and convenience reasons. For example, American Airlines began accepting only payment cards for inflight purchases on all its domestic routes starting June 1, 2009. Also, many quick service restaurants and coffee shops now accept payment cards to capture greater sales and increase transaction speed.

As more consumers and merchants adopt payment cards, providers of these products may benefit from economies of scale and scope. Some European payment providers might enjoy these benefits in the future as greater cross-border harmonization occurs with the introduction of the Single Euro Payments Area (SEPA). The potential advantages of SEPA are increased competition among a greater number of payment

[†] Wilko Bolt is a Senior Economist in the research department at De Nederlandsche Bank in Amsterdam. He received his PhD in economics from the Vrije Universiteit of Amsterdam. His current research focuses on the economics of payments and the theory of two-sided markets and its antitrust implications. He has published in journals such as *American Economic Review*, *European Economic Review*, *International Journal of Industrial Organization*, and *Journal of Banking and Finance*. His book ‘Credible Threats in Negotiations: A Game-theoretic Approach’, written jointly with Harold Houba, was published by Kluwer Academic Publishers in 2002. In 2007, Bolt was awarded the biannual Hennisman Prize by the Dutch Royal Economic Association for best article in its journal *De Economist* during 2005–2006.

[§] Sujit “Bob” Chakravorti is a senior economist in the Financial Markets Group at the Federal Reserve Bank of Chicago and has been analyzing retail and wholesale payment systems for over 20 years. He has published over 30 articles in industry, academic, and Federal Reserve publications including the *Antitrust Bulletin*, *Journal of Money Credit and Banking*, *Journal of Payments Strategy and Systems*, and the *Review of Network Economics*. Prior to joining the Federal Reserve System, he worked at KPMG as an international economist advising foreign governments on financial market policy. In addition, he has been a visiting scholar at De Nederlandsche Bank (Dutch central bank), the European University Institute, the International Monetary Fund, and the University of Granada. He received a B.A. in economics and genetics from the University of California- Berkeley and M.A. and Ph.D. in economics from Brown University.

¹ Wilko Bolt and David B. Humphrey, “Payment Network Scale Economies, SEPA, and Cash Replacement,” *Review of Network Economics* 6(4): 453–473 (December 2007).



providers and the realization of scale economies and more efficient payment instruments. Beijnen and Bolt provide estimates of scale economies that quantify the potential benefits of SEPA arising from consolidation of electronic payment processing centers across the euro zone.²

The increased usage of cards has increased the value of payment networks, such as Visa Inc., MasterCard Worldwide, Discover Financial Services, and others. In 2008, Visa Inc. had the largest initial public offering (IPO) of equity, valued at close to \$18 billion, in U.S. history.³ One potential reason for Visa to change its corporate structure from a card association to a publicly traded company is to reduce antitrust scrutiny by regulators and to lower the threat of lawsuits filed by certain payment system participants.⁴

Some industry observers have suggested that the high profitability of payment card providers has increased scrutiny by public authorities in many jurisdictions. Several U.S. merchants have filed lawsuits against MasterCard and Visa regarding the setting of interchange fees. Interchange fees are generally paid by the merchant's bank to the cardholder's bank and are set by the network. In December 2007, the European Commission (EC) ruled that the (multilateral) interchange fees for cross-border payments in the European Union applied by MasterCard Europe violated Council Regulation (European Commission) No. 1/2003. Since July 1, 2009, MasterCard Europe established interchange fees for consumer card transactions that, on average, do not exceed 30 basis points for credit cards and 20 basis points for debit cards. However, the EC stressed that it will continue its antitrust investigation against Visa.

To date, there is still little consensus—either among policymakers or economic theorists—on what constitutes an efficient fee structure for card-based payments. In this article, we discuss several economic models that analyze whether intervention by public authorities might improve the welfare of payment system participants.

Payment Flows in a Card Network

The two-sided market literature has been used to analyze the structure of fees paid by consumers and merchants. Rochet and Tirole define a two-sided market as a market where end-users are unable to negotiate prices based on costs to participate on a platform and the price structure affects the total volume

² Christine Beijnen and Wilko Bolt, "Size Matters: Economies of Scale in European Payments Processing," *Journal of Banking and Finance* 33(2): 203-210 (February 2008).

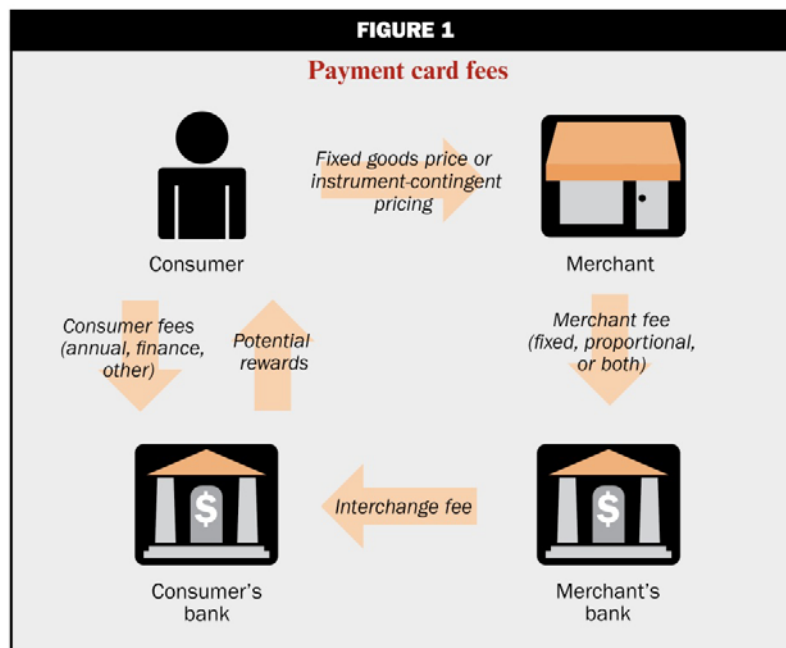
³ Katie Benner, "Visa's Record IPO Rings Up 28 percent Gain," CNNMoney.com, March 19 2008, available at http://money.cnn.com/2008/03/19/news/companies/visa_ipo_opens.fortune/index.htm.

⁴ David Enrich, "Visa Plans Restructuring, Sees IPO in 12-18 Mos.," *Dow Jones Newswires*, October 11, 2006.



of transactions.⁵ In the payments context, consumers and merchants generally do not negotiate prices of goods and services based on the payment instrument used to make a purchase. While not common, some merchants do post different prices based on the payment instrument used to make the purchase. For example, the prices are the same regardless of whether the consumer pays in cash or with a payment card. However, when merchant fees increase, some merchants might refuse to accept payment cards, resulting in fewer potential card transactions. Similarly, raising consumer fees may reduce consumer participation.

Figure 1: Costs and benefits of different payment methods



Payment card networks comprise of consumers and their banks (known as issuers), as well as merchants and their banks (known as acquirers). Issuers and acquirers are part of a network that sets the rules and procedures for clearing and settling payment card receipts among its members. In Figure 1, we diagram the four participants and their interactions with one another. First, a consumer establishes a relationship with an issuer and receives a payment card. Consumers generally do not pay per transaction fees but often pay annual membership fees to the banks that issue the payment cards. In addition, many payment card issuers give their customers per transaction rewards, such as cash back or other frequent-use rewards. Second, a consumer makes a purchase from a merchant. Generally, the merchant charges the same price regardless of the type of payment instrument used to make the purchase. Third, if a merchant has established a

⁵ Jean-Charles Rochet and Jean Tirole, "Two-sided Markets: A Progress Report," *RAND Journal of Economics* 37(3): 645–667 (Autumn 2006).



relationship with an acquirer, it is able to accept payment card transactions. The merchant either pays a fixed per transaction fee (more common for debit cards) or a proportion of the total purchase amount, known as the merchant discount fee (more common for credit cards), to its acquirer. In some instances, merchants are charged a fixed fee *and* a proportional fee. For credit cards, the merchant discount can range from 1% to 5% depending on the type of transaction, type of merchant, and type of card, as well as whether the card is present or not, among other factors. Fourth, the acquirer pays an interchange fee to the issuer.

Studying the costs to banks to provide payment services is difficult, given the proprietary nature of the cost data. However, there are some European studies that attempt to quantify the real resource costs of several payment services. In these studies, social cost refers to the total cost for society net any monetary transfers between participants, and reflects the real use of resources used in the production and usage of payment services. For the Netherlands in 2002, Brits and Winder report that the social costs of all point-of-sale (POS) payments (cash, debit cards, credit cards, and prepaid cards) amounted to 0.65% of gross domestic product (GDP).⁶ The social cost of payment services for Belgium in 2003 was 0.75% of GDP.⁷ Bergman, Guibourg, and Segendorff find that the social cost of providing cash, debit card payments, and credit card payments was approximately 0.4% of GDP in Sweden for 2002.⁸ Based on a panel of 12 European countries during the period 1987–99, Humphrey et al. conclude that a complete switch from paper-based payments to electronic payments could generate a total cost benefit close to 1% of the 12 nations' (Belgium, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom) aggregate GDP.⁹

These numbers confirm the widespread agreement that the ongoing shift from paper-based payments to electronic payments may result in large economic gains. Compared with cash, electronic payments also offer benefits in terms of greater security, faster transactions, and better recordkeeping; in addition, electronic payments offer possible access to credit lines. Some key benefits of using cash include privacy and anonymity that payment cards do not provide. Merchants may also benefit from increased sales or cost

⁶ Hans Brits and Carlo Winder, "Payments Are No Free Lunch," *Occasional Studies*, De Nederlandsche Bank 3(2) (2005).

⁷ Guy Quaden, "Costs, Advantages, and Disadvantages of Different Payment Methods," (report presented for the National Bank of Belgium, December 2005).

⁸ Mats Bergman, Gabriela Guibourg, and Björn Segendorff, "The Costs of Paying—Private and Social Costs of Cash and Card," working paper No. 212, Sveriges Riksbank, 2007.

⁹ David Humphrey et al., "Benefits From a Changing Payment Technology in European Banking," *Journal of Banking and Finance* 30(6): 1631–1652 (June 2006).



savings by accepting an array of electronic payment instruments. However, these benefits to consumers and merchants are often difficult to quantify.

Using U.S. retail payments data, Garcia-Swartz, Hahn, and Layne-Farrar attempt to quantify both the costs and benefits of POS payment instruments.¹⁰ They find that shifting payments from cash and checks to payment cards results in net benefits for society as a whole, but they also conclude that merchants may be paying a disproportionate share of the cost. Much of the payment card literature focuses on the proportion of the total price paid by merchants and consumers. In other words, economists are trying to answer the question: Do the sum of prices to end-users for card payments and their peculiar asymmetric structure reflect the exercise of market power by card providers or do they reflect the nature of the service provided? In the next section, we consider how the economics literature has attempted to answer this question.

Economic Models of Payment Cards

In this section, we review some important contributions to the theoretical payment card literature. The early models of payment cards ignored strategic interactions of consumers and merchants, and focused on the aggregate demand of each type of end-user and the level of the interchange fee. These models were extended to include explicit consumer and merchant interactions but assumed inelastic consumer demand for goods, no price differentiation by merchants based on the payment instrument used by consumers, and exogenous benefits from card usage. Several models extended this literature by considering merchants' ability to separate consumers by charging different prices. Another set of models expanded the literature by considering the ability of payment cards to increase sales because the cards provided greater security and eased consumers' liquidity and credit constraints. In addition, other models of payment cards have considered network competition, as well as competition among different types of payment instruments.

Models focusing on interchange fees

Here, we discuss the academic literature on interchange fees. Baxter argues that the equilibrium quantity of payment card transactions occurs when the total transactional demand for credit card services, which are determined by consumer and merchant demands jointly, is equal to the total transactional cost for credit card services, including both issuer and acquirer costs.¹¹ A consumer's willingness to pay is based on her

¹⁰ Daniel D. Garcia-Swartz, Robert W. Hahn, and Anne Layne-Farrar, "A Move Toward a Cashless Society: A Closer Look at Payment Instrument Economics," *Review of Network Economics* 5(2): 175–198 (2006).

¹¹ William F. Baxter, "Bank Interchange of Transactional Paper: Legal and Economic Perspectives," *Journal of Law and Economics* 26(3): 541–588 (October 1983).



net benefits received and is greater than or equal to the fee in equilibrium. Similarly, the merchant's fee is less than or equal to the net benefits it receives. Hence, pricing each side of the market based on marginal cost—as would be suggested by economic theory for one-sided competitive markets—need not yield the socially optimal allocation. To arrive at the socially optimal equilibrium, a side payment may be required between the issuer and acquirer.

Unfortunately, Baxter's framework does not allow us to study the optimal setting of interchange fees by banks, since their profits are zero regardless of the level of the interchange fee. Schmalensee extends Baxter's analysis by allowing issuers and acquirers to exercise market power but still assumes that merchants operate in competitive markets.¹² His results support Baxter's conclusions that the interchange fee balances the demands for payment services by each end-user type and the cost to banks to provide them. Schmalensee finds that the profit-maximizing interchange fee of issuers and acquirers may also be socially optimal.

Unlike Baxter and Schmalensee, Rochet and Tirole consider strategic interactions of consumers and merchants.¹³ They have three main results. The first result is that their socially optimal interchange fee is higher than the socially optimal Baxter interchange fee, since issuers exert their market power and capture merchants' surplus. Their second result is that the interchange fee that maximized profit for the issuers may be more than or equal to the socially optimal interchange fee, depending on the issuers' margins and the cardholders' surplus. Third, merchants are willing to pay more than their net benefit if they can steal customers from their competitors or retain their customers by accepting cards. However, overall social welfare does not improve when merchants steal customers from their competitors by accepting payment cards.

Wright extends Rochet and Tirole by considering a continuum of industries where merchants in different industries receive different benefits from accepting cards.¹⁴ His model is better able to capture the trade-off between consumer benefits and merchant acceptance when the interchange fee is increased because some merchants will not accept cards. He concludes that the interchange fee that maximizes overall social welfare may be higher or lower than the interchange fee that maximizes the number of transactions. In

¹² Richard Schmalensee, "Payment Systems and Interchange Fees," *Journal of Industrial Economics* 50(2): 103–122 (June 2002).

¹³ Jean-Charles Rochet and Jean Tirole, "Cooperation among Competitors: Some Economics of Payment Card Associations," *RAND Journal of Economics* 33(4): 549–570 (Winter 2002).

¹⁴ Julian Wright, "The Determinants of Optimal Interchange Fees in Payment Systems," *Journal of Industrial Economics* 52(1): 1–26 (2004).



particular, restricting the total number of transactions by setting higher interchange fees raises total welfare if the gain in surplus of the marginal card user who now *starts* using his card along with all those merchants who accept his card exceeds the loss in surplus of the inframarginal merchant who now *stops* accepting cards along with all those card users who can no longer use their cards for purchases at her store.

Models with price differentiation at the point of sale

The models discussed so far have largely ignored the ability of merchants to pass on a part or all of their payment cost to consumers—whether in the form of higher prices to their card-based consumers or as a higher uniform price to all consumers. In some cases, merchants are not allowed to add a surcharge for payment card transactions because of legal or contractual restrictions, but they may be allowed to give cash discounts. However, in jurisdictions where merchants are free to set higher prices for purchases made with payment cards, they usually do not.¹⁵ Even if differential pricing based on the payment instrument used is not common, the possibility to do so may enhance the merchants' bargaining power in negotiating their fees. If merchants charged different prices, cash-paying consumers would not be paying or they would be paying less for card-paying customers.

Wright extends Rochet and Tirole to consider the effects of no-surcharge rules.¹⁶ He finds that no-surcharge rules generate higher welfare than when monopolist merchants are allowed to set prices based on the payment instrument used. He argues that merchants are able to extract consumers' surplus *ex post* from payment card users, while cash users are unaffected. Wright only considers equilibria where merchants will continue to sell the same quantity of goods to cash users at the same price. When merchants are allowed to surcharge, they extract "too much" surplus *ex post* from customers who use payment cards because merchants set higher prices for card purchases.

Schwartz and Vincent study the distributional effects among cash and card users with and without no-surcharge rules.¹⁷ They find that the absence of pricing based on the payment instrument used increases network profit and harms cash users and merchants. The payment network prefers to limit the merchant's ability to separate card and cash users by forcing merchants to charge a uniform price to all of its

¹⁵ IMA Market Development AB, "Study Regarding the effects of the abolition of the Nondiscrimination Rule in Sweden," report, Lerum, Sweden, February 29, 2000; Wilko Bolt, Nicole Jonker, and Corry Van Renselaar, "Incentives at the Counter: An Empirical Analysis of Surcharging Card Payments and Payment Behavior in the Netherlands," *Journal of Banking and Finance* (forthcoming 2009).

¹⁶ Julian Wright, "Optimal Card Payment Systems," *European Economic Review* 47(4): 587–612 (August 2003).

¹⁷ Marius Schwartz and Daniel R. Vincent, "The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network," *Review of Network Economics* 5(1): 72–102 (March 2006).



customers. When feasible, the payment network prefers rebates (negative per transaction fees) given to card users. Granting such rebates to card users boosts their demand, while simultaneously forcing merchants to absorb part of the corresponding rise in the merchant fee, because any resulting increase in the uniform good's price must apply equally to cash users. In this way, the network uses rebates to indirectly extract surplus from cash-paying customers in the form of higher prices. If rebates are feasible, card users are always better off. Overall welfare rises if the ratio of cash users to card users is sufficiently large and merchants' net benefits from card acceptance are sufficiently high.

Gans and King argue that, as long as there is "payment separation," the interchange fee is neutral regardless of the market power of merchants, issuers, and acquirers.¹⁸ The interchange fee is said to be neutral if a change in the interchange fee does not change the quantity of consumer purchases and the profit level of merchants and banks. When surcharging is costless, merchants will implement pricing based on the payment instrument used, taking away the potential for cross-subsidization across payment instruments and removing the interchange fee's role in balancing the demands of consumers and merchants. In effect, the cost pass-through is such that lower consumer card fees (due to higher interchange fees) are exactly offset by higher goods prices from merchants. Payment separation can occur if one of the following is satisfied: There are competitive merchants, and they separate into cash-accepting or card-accepting categories, in which each merchant only serves one type of customer and is prevented from charging different prices; or merchants are able to fully separate customers who use cash from those who use cards by charging different prices. Therefore, they argue that policymakers should remove any merchant pricing restrictions, such as no-surcharge rules.

Models with competition between networks

We have not yet considered models where competition among payment networks is explored. Economic theory suggests that competition generally reduces prices, increases output, and improves welfare. However, with two-sided markets, competition may yield an inefficient price structure. A key aspect of network competition is the ability of end-users to participate in more than one network. When end-users participate in more than one network, they are said to be "multihoming." If they connect only to one network, they are said to be "singlehoming." As a general finding, competing networks try to attract end-users who tend to singlehome, since attracting them determines which network has the greater volume of business. Accordingly, the price structure is tilted in favor of end-users who singlehome.

¹⁸ Joshua S. Gans and Stephen P. King, "The Neutrality of Interchange Fees in Payment Systems," *Topics in Economic Analysis & Policy* 3(1), article 1 (2003), available at www.bepress.com/bejeap/topics/vol3/iss1/art1.



Rochet and Tirole extend their previous work by considering network competition.¹⁹ Their primary focus is on the price structure or balance between consumers and merchants in a three-party network. They do not explicitly model the interchange fee but study the impact of competition on the structure of prices. Under a set of plausible assumptions they find that the price structures for a monopoly network and competing platforms are the same, and if the sellers' demand is linear, this price structure in the two environments generates the highest welfare under a balanced budget condition.

Guthrie and Wright extend Rochet and Tirole by assuming that consumers are able to hold one or both payment cards and that merchants are motivated by "business stealing" in deciding whether to accept payment cards in a four-party network.²⁰ They only consider networks that provide identical payment services, and they find that network competition results in higher interchange fees than those that would be socially optimal. In this model, competition results in both networks charging the same interchange fee because both networks offer identical payment products.

Chakravorti and Roson extend Rochet and Tirole by considering the effects of network competition on total price and on price structure where networks offer differentiated products.²¹ They only allow consumers to participate in one card network, whereas merchants may choose to participate in more than one network. They compare welfare properties when these two networks operate as competitors and as a cartel where each network retains demand for its products from end-users. Like Rochet and Tirole and Guthrie and Wright, they find that competition does not necessarily improve or worsen the balance of consumer and merchant fees from the socially optimal one. There are other fee structures for a given sum of consumer and merchant fees that would improve consumer and merchant welfare. However, they find that the welfare gain from the drop in the sum of the fees from competition is generally larger than the potential decrease in welfare from less efficient fee structures.

Models accounting for the role of credit

So far, we have considered models that ignore the extension of credit as a benefit to consumers and merchants. Given the high level of antitrust scrutiny targeted toward credit card networks, we find this omission in most of the academic literature surprising. In the long run, aggregate consumption over

¹⁹ Jean-Charles Rochet and Jean Tirole, "Platform Competition in Two-sided Markets," *Journal of the European Economic Association* 1(4): 990–1029 (2003).

²⁰ Graeme Guthrie and Julian Wright, "Competing Payment Schemes," *Journal of Industrial Economics* 55(1): 37–67 (2007).

²¹ Sujit Chakravorti and Roberto Roson, "Platform Competition in Two-sided Markets: The Case of Payment Networks," *Review of Network Economics* 5(1): 118–143 (March 2006).



consumers' lives may not differ because of access to credit, but such access may enable consumption smoothing that increases consumers' utility. From a merchant's perspective, extension of credit may lead to intertemporal business stealing. In other words, merchants attract consumers who do not have funds today by accepting credit cards, resulting in merchants tomorrow being unable to make sales to consumers who bought today on credit. In addition to extracting surplus from consumers and merchants, banks have an additional source of surplus—liquidity-constrained consumers. How much surplus can be extracted depends on how much liquidity-constrained consumers discount tomorrow's consumption.

Chakravorti and Emmons consider the costs and benefits of consumer credit in a four-party network where consumers are subject to income shocks after making their credit card purchases and some are unable to pay their credit card debt.²² To our knowledge, they are the first to link the insurance aspect of credit cards to their payment component. Observing that over 75% of U.S. card issuer revenue is derived from cash-constrained consumers,²³ they consider the viability of the credit card system if it were completely funded by these types of consumers. They also study the convenience use of credit cards—that is, the usage by those who do not need credit to make purchases and its impact on social welfare.

Chakravorti and Emmons derive three main results. First, if consumers sufficiently discount future consumption, liquidity-constrained consumers who do not default would be willing to pay all credit card network costs *ex ante*, resulting in all consumers being better off. The key assumption is that at least a certain number of consumers face binding liquidity constraints and do not default. Second, if merchants charge a single price for a good regardless of how consumers pay, and if there are no side payments made by issuers to convenience users, card-accepting merchants who charge a single price for all purchases will attract only liquidity-constrained consumers because some merchants charge a lower price and only serve cash-paying customers. Note, only those consumers who are liquidity-constrained use credit cards, and there is no convenience use. Third, if card issuers extend rebates to convenience users, a merchant can, under certain conditions, attract all types of consumers—including consumers who carry a credit card balance month to month and those who do not—when a single price is charged.

Chakravorti and To consider a scenario with monopolist merchants and a monopolist bank that serves both consumers and merchants where the merchants absorb all credit and payment costs in a two-period

²² Sujit Chakravorti and William R. Emmons, "Who Pays for Credit Cards?" *Journal of Consumer Affairs* 37(2): 208–230 (2003).

²³ Jeffrey Green, "Exclusive Bankcard Profitability Study and Annual Report 2008," *Card and Payments* (May 1, 2008): 36–38.



dynamic model.²⁴ Their model yields the following results. First, if merchants earn a sufficiently high profit margin and the cost of funds is sufficiently low, the economy is able to support credit cards. In other words, the benefits to consumers and merchants must be greater than the cost to support the credit card network. Second, the fee that merchants are willing to pay their banks increases as the number of credit-constrained consumers increases. Third, a prisoner's dilemma situation may arise: Each merchant chooses to accept credit cards, but by doing so, each merchant's discounted two-period profit is lower. In other words, there exists intertemporal business stealing among merchants across different industries, potentially resulting in all merchants being worse off.

Models with competition among payment instruments

Most of the literature ignores competition between payment instruments, with one payment provider offering multiple payment options to its customers and setting prices to maximize profits. Moreover, most economic models of payment cards generally do not consider price incentives offered by merchants to steer consumers to a specific type of payment card.

In Bolt and Chakravorti, we study the ability of banks and merchants to influence the consumer's payment instrument choice when they have access to three payment forms—cash, debit card, and credit card. To our knowledge, this model is the first to analyze payment network competition by combining elements of models that stress price balance with those that consider liquidity constraints and safety concerns of consumers.

In our model, consumers participate in payment card networks to insure themselves from three types of shocks—uncertain income flows, theft, and the merchant they are matched to. Our key results can be summarized as follows. With sufficiently low processing costs relative to theft and default risk, the social planner sets a zero merchant fee to completely internalize the card acceptance externality. The bank may also set zero merchant fees but only if merchants are able to sufficiently pass on payment costs to their consumers or if payment costs are zero. If payment card costs are too high, the social planner sets a higher merchant fee than the bank so as to contract card acceptance in favor of cash. We find that bank profit increases when merchants are unable to pass on payment costs to consumers due to lower goods prices and greater ability to extract merchant surplus. The relative costs of providing debit and credit cards determine whether the bank will provide both or only one type of payment card. Finally, uniform price policies increase bank profits when the bank supplies both types of payment cards than when merchants

²⁴ Sujit Chakravorti and Ted To, "A Theory of Credit Cards," *International Journal of Industrial Organization* 25(3): 583–595 (June 2007).



adopt pricing based on the payment instrument used. However, consumers and merchants are worse off when consumers without liquidity constraints use credit cards because they do not receive the proper price incentives, resulting in use of a less efficient payment instrument.

Conclusion

In summarizing the payment card literature, we find that no one model is able to capture all the essential elements of the market for payment services. It is a complex market with many participants engaging in a series of interrelated bilateral transactions. Moreover, appropriate pricing arrangements for payment instruments is difficult, since payment networks are subject to large economies of scale and give rise to strong usage and adoption externalities. Much of the debate over various payment card fees is concerned with the allocation of the surpluses from consumers, merchants, and banks, as well as who is able to extract surpluses from whom.

We are able to draw the following conclusions. First, a side payment between the issuer and the acquirer may be required to get both sides on board. There is no consensus among policymakers or economists on what constitutes an efficient fee structure for card payments. Second, while consumers generally react to price incentives at the point of sale, merchants are reluctant to charge higher prices to consumers who benefit from card use. Third, network competition may not improve the price structure but may significantly reduce the total price paid by consumers and merchants. Fourth, consumers and merchants both value credit extended by credit card issuers (along with other benefits such as security), and consumers and merchants are willing to pay for it.

Sound public policy regarding payment fees is difficult. The central question is whether the specific circumstances of payment markets are such that intervention by public authorities can be expected to improve economic welfare. Efficiency of payment systems is measured not only by the costs of resources used, but also by the social benefits generated by them. While the theoretical literature on payment cards is growing, there are still too few empirical investigations to guide policymakers. We hope that recent regulatory changes in different parts of the world will generate rich sets of data that can be exploited by economists to test how well the theories fit the data.



What Does the Credit Card Market Have In Common with a Peacock?

Joshua M. Frank[†]

The credit card industry is highly concentrated. Ninety percent of the lending market is controlled by the top 10 issuers, yet there still appears to be fierce competition for balances and customers among these top issuers. However, there is a widespread belief that pricing is unfair despite this competition, resulting in reforms implemented by the Federal Reserve and passage of the May 2009 Credit CARD Act. The question explored here is how a market like the credit card market can fail to function properly despite heavy competition. This article offers a new conceptualization of a market structure, defined as a “peacock market,” characterized by a disproportionate mix of form over substance. It then shows how practices within the credit card industry can be better understood through the new structure.

Defining a Peacock Market

Charles Darwin first wondered how traits that clearly negatively impact survival rates can continue to exist in an evolutionary process. One classic example he used is the elaborate feather displays of male peacocks. Generating and maintaining these feathers takes considerable energy away from other purposes and adds the cost of slowing the males down if they attempt to escape from a predator. To explain this phenomenon, Darwin created the theory of sexual selection. While male peacock displays are inefficient in terms of personal survival, they have great signal value, which attracts mates and increases reproductive success. Early peacocks initially showed a modest display of relatively low cost that became self-perpetuating, and eventually grew to absurd proportions (from a survival standpoint). The peacock remains a classic example in evolutionary biology of how signal value can displace efficiency and be evolutionarily self-perpetuating despite intense competition.

[†] Josh Frank is Senior Researcher for the Center for Responsible Lending. Josh received his Ph.D. in Economics from Rensselaer Polytechnic Institute, his MBA from Texas Tech, and his BA from the University of California at San Diego. At CRL, he conducts research primarily on credit card issues as well as on auto loans. Before joining the Center, Josh was a professor of economics at SUNY, an economist for the State of New York Assembly’s Ways and Means Committee, and executive director of FIREPAW, a research and education-oriented animal welfare non-profit organization. Josh has also spent a number of years in the credit card industry for a top ten issuer. At different times he managed the Strategic Planning, Financial Planning, and various Information Systems functions for the company.



Similarly, defined here a peacock market is a marketplace that has evolved to become dominated by elaborate signals (such as price or quality signals to consumers) no longer connected in a meaningful way to price, quality, or efficiency. As in biological evolution, peacock markets can become self-perpetuating because the signal retains value in creating demand despite being dysfunctional. The signals are designed to take maximum advantage of information shortages, limited attention, hard-wired behavioral biases, and socially-reinforced beliefs.¹

Just as male peacocks are rewarded in the natural selection process for their signals, peacock markets will reward firms that structure products and services around powerful but distorted signals that tend to suggest prices are low or quality is high. Firms will then ensure they are compensated for the prominently signaled pricing through complex, deferred, and other pricing mechanisms generally underweighted by consumers. These pricing mechanisms may be buried deep into contractual terms and conditions. They may be hidden in practice due to the complexity of the pricing structure and the scarce attention of consumers. They also may take advantage of well-known consumer biases. The known tendency for people to discount the distant future too heavily can result in prices that are backloaded, with the initial price being artificially low. Biases towards excessive optimism and overconfidence can result in prices that exploit contingencies that consumers underweight in likelihood (for example, the probability of being charged a late fee or receiving a penalty interest rate). Generally, firms in a peacock market focus their marketing and product design on signaling strongly in those dimensions that appeal to these biases, while utilizing mechanisms to maximize profitability in underweighted areas, rendering consumers overconfident and misinformed.

Why Peacock Markets Form

Certain consumer purchases are less inherently transparent both in quality and in pricing. These are the markets more likely to evolve into peacock markets. When the underlying desirable qualities of a product are not transparent and when pricing is inherently complex, cues are more easily converted to a signal that has little relation to true quality or price.

Complex ongoing services, including many within the financial marketplace, may be more likely to morph into peacock markets. Important characteristics of quality are often difficult to measure, making selection among alternatives difficult. It is easy to add multidimensionality to pricing due to the complicated nature

¹ “Shrouded attributes, consumer myopia, and information suppression in competitive markets,” describes how biases and information shortages may be exploited and how it may not be in any firm’s interest to reveal additional information. Xavier Gabaix and David Laibson, working paper 05-18, MIT Department of Economics, April 11, 2005.



of the service. In addition, multiple contingencies often exist that require additional service or can be used as an excuse to change the cost of a service. The book *Gotcha Capitalism* warns consumers about a variety of markets where hidden pricing is common, and tends to be dominated by multidimensional services received over an extended period of time such as cellular phone service, cable television, credit cards, and checking accounts.²

There may also be an arbitrary element in the determination of which markets become peacock markets, with minor historical events shaping the industry's future in important ways. It is now well-established that both the use of technology in the economy and institutions related to the economy are subject to path-dependence, where the path of future change and innovation depends heavily on past events.³ Often, dimensions along which progress is defined can get "locked-in" to a particular path based on prior events and choices. In the case of peacock markets, the dependence on inefficient signals can become self-perpetuating for a number of reasons. First, the technology for creating future signals can be cheaper and more efficiently produced when signals with a shared technological basis have been used in the past. For example, one "innovation" in the credit card industry has been the ability to price different account balances separately. This technology could then be expanded from two separate balances to three and higher. It also spawned a variety of complex pricing strategies that depended and built on this technology. Peacock markets can also be self-perpetuating because the ability to shroud new terms increases based on the number of terms already in existence. Behavioral economics suggests that people have limits on how many dimensions of price or quality they can consider at once when evaluating goods or services. For example, if a credit card has a single annual percentage rate (APR), adding a single fee that must be clearly disclosed is likely to receive more consumer attention when evaluating the product than if there are a half dozen potential fees and multiple APRs on the account already.

Peacock markets are also self-perpetuating from a cultural/institutional perspective of the firm. While individual tactics may vary, over time the strategic philosophy of a firm and even an industry can become deeply entrenched. If industry leaders have seen their historical success driven by price reduction through increased efficiency or core product improvements through innovation, then future strategies will tend to

² Bob Sullivan, *Gotcha Capitalism* (New York: Ballantine Books, 2007).

³ To some readers, this point may seem obvious. However, much of economic theory is based on assumptions of path independence, where markets always evolve towards the most efficient configuration. See W. Brian Arthur, "Competing Technologies, Increasing Returns, and Lock-in by Historical Events," *Economic Journal* (1989): 99, 116-131 and Paul A. David, "Clio and the Economics of QWERTY," *American Economic Review (Papers and Proceedings)* (1985): 75, 332-37 for the case for technology. See Douglass North, *Institutions, Institutional Change and Economic Performance* (Cambridge: Cambridge University (1990) for the case regarding institutions.



gravitate towards these themes. For example, historically innovation in the computer industry focused on increasing core product improvements (i.e., processing power, memory, lighter laptops, flat screens, CD and portable memories instead of floppy disks, wireless capacity, etc.). In contrast, if the leading firms have grown in profit and size by producing deceptive signals that suggest low prices while actually charging higher prices, this strategic philosophy will become just as deeply ingrained and can become self-perpetuating. For example, in the mortgage industry preceding the financial crisis, deceptive pricing practices grew in prevalence and were reflected in product features such as “exploding” adjustable rate mortgages with large built-in price increases.

For all of these reasons, markets that have peacock-market-tendencies will likely tend to become more and more differentiated into predominantly peacock markets over time.

A Field Guide to Spotting Peacocks

Since some level of distortion may occur in many markets, it is important to have applied indicators that differentiate a peacock market from other markets. Signs of a peacock market include:

- **Signals are no longer strongly correlated to their underlying intended purpose.** Often in traditional signaling theory, an intended signal is accompanied by random distortion.⁴ However, in a peacock market the distortion is intentional. In fact, a price signal may yield little information about the overall price and may even be inversely correlated with it.
- **Increasing disparity *over time* between the signaled information and the underlying factor it is intended to represent.** A peacock market is an evolving market. Therefore, over time, the signals may get further and further removed from the information they are intended to signal. For example, there may be growth in the disparity between the average signal price and non-signal price over an extended period.
- **The signal comes to dominate and drive product design.** Signals in peacock markets are no longer merely marketing to enhance a core product, but become the primary driver of demand and profitability. Therefore, products evolve to be designed around the signal. This is a clear sign of market inefficiency since innovation is no longer focused on making fundamentally better products.
- **The evolution of “fragmentation.”** Fragmentation occurs when firms intentionally break price or quality into smaller pieces for the purpose of creating signals. The traditional economic literature

⁴ See, e.g., A. Michael Spence, “Signaling in Retrospect and the Informational Structure of Markets,” (Nobel Prize Lecture, December 8, 2001).



on signaling theory indicates the signal is generally used to improve information in an uncertain environment.⁵ Fragmentation works in the opposite direction, where a segment of price (or quality) is intentionally broken off from the whole to create an artificial signal. Therefore, fragmentation is actually an information reduction strategy.

- **“Shrouded” costs (or reductions in benefits) are prevalent.** These shrouded attributes are used to compensate for losses due to extreme signals and to generally enhance profits. Shrouded costs can be naturally underweighted due to behavioral biases, or hidden through complexity of contracts or complexity of the mechanism itself. A shrouded reduction in benefits could, for example, be an obscure but important exception to an insurance policy that reduces firm costs and consumer value.
- **Product and market efficiency is reduced as a result of the above.** Inefficiency is often difficult to conclusively prove in economics, but intuition suggests that certain pricing and product structures are likely to be inefficient. If a snack food manufacturer decides, instead of raising prices per unit, to reduce the contents by weight of each box sold while maintaining the same box size, this seems to likely be inefficient packaging designed to distort consumer perceptions.

Applying the “Peacock Market” Concept to Credit Cards

A Federal Reserve paper described the evolution of the industry as follows: “[T]he relatively straightforward pricing model of a single APR, an annual fee, and modest penalty fees has been replaced by a model with a complex set of APRs, new and increased fee structures, and sophisticated finance charge computation techniques.”⁶ Much of price complexity in credit cards is related to an attempt by firms to create the illusion of low prices as well as an attempt to raise demand through differentiation of a product that is a fairly uniform commodity in reality. In a June 2009 speech, President Obama described problematic credit card issuers as companies that “compete not by offering better products, but more complicated ones, with more fine print and more hidden terms.”⁷

Some of the practices discussed here may be limited in scope or prohibited by the Credit CARD Act passed in May 2009 as well by Federal Reserve Unfair and Deceptive Acts or Practices (UDAP) Rules announced in

⁵ Ibid.

⁶ Mark Furletti, “Credit Card Pricing Developments and Their Disclosure,” discussion paper, Philadelphia Federal Reserve Payment Card Center, January 2003.

⁷ Barack Obama, “Remarks by the President on 21st Century Financial Regulation Reform,” June 17, 2009, available at http://www.whitehouse.gov/the_press_office/Remarks-of-the-President-on-Regulatory-Reform/



December 2008. However the majority of these changes have not yet taken effect. In addition, the behavior of companies prior to implementation of these rules lends insight into issuer behavior and market structure in general. As previously noted, even if specific practices change because Congress and the Federal Reserve have determined that many issuer pricing policies are unfair and deceptive, the strategic philosophy and corporate culture at firms in this market climate may be less easily altered.

Signals No Longer Are Strongly Correlated to Their Underlying Intended Purpose

The most obvious credit card price signal is the low introductory “teaser” rate. Often this price will be displayed very prominently in large font on the first page of the card offer, as well as several other places, and sometimes even on the envelope. Sometimes “no annual fee” will also be used as a prominent price signal.

Both of these prove to be poor signals for overall product price. The presence of a low teaser rate is not a useful indicator of an overall low cost of credit. In fact, in 2009 there was no significant correlation between the teaser APR and the regular purchase APR in credit card solicitations.⁸ Likewise, “no annual fee” says little to nothing about the size of other fees or total price paid.

As one should expect in a peacock market, these signals are also distorted towards known biases. There is an emphasis on short-term teasers that takes advantage of excessive discounting of future costs and an overly optimistic view of how much the consumer might borrow in the future. The annual fee, which consumers cannot avoid, is typically avoided in recent product offerings, while back-end contingent fees and APR changes that consumers underestimate in likelihood of occurrence are prevalent. Credit card terms also take advantage of cognitive limits and limited attention by adding many dimensions to price that are mathematically complicated (such as payment allocation methods) or simply go unnoticed.

Increasing Disparity Over Time between the Signaled Information and the Underlying Factor It Is Intended to Represent

⁸ Results are based on analysis of Mintel Comperemedia credit card solicitations through July 2009. Mintel Comperemedia is a searchable competitive database tracking direct mail and print advertising in the United States and Canada, as well as e-mail in the U.S. Mintel Comperemedia tracks information to analyze eight vertical markets: Banking, Credit Card, Investments, Insurance, Mortgage and Loan, Telecom, Travel and Leisure, and Automotive. The regression coefficient between the two variables was slightly negative, suggesting if anything an inverse relationship. However, the relationship was not statistically significant. It should be noted that the purchase rate is merely another signal of cost rather than the true cost of credit, which may vary considerably from the purchase rate due to default APRs, fees, and different APRs for other balances.



There is strong evidence of a growing disparity between the signal price and other dimensions of credit card industry costs. The number and magnitude of penalty and miscellaneous service fees have increased over time. When the use of penalty fees in the credit card industry is studied, their use was shown to be increasing over time both in its appearance in the product contract (Figure 1), and in terms of how many people were at a penalty rate (Figure 2).⁹

Figure 1: Share of Credit Card Solicitations with a Penalty Rate

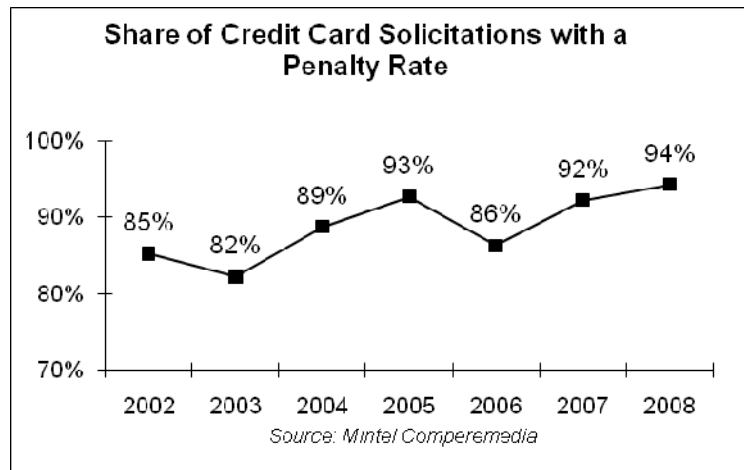
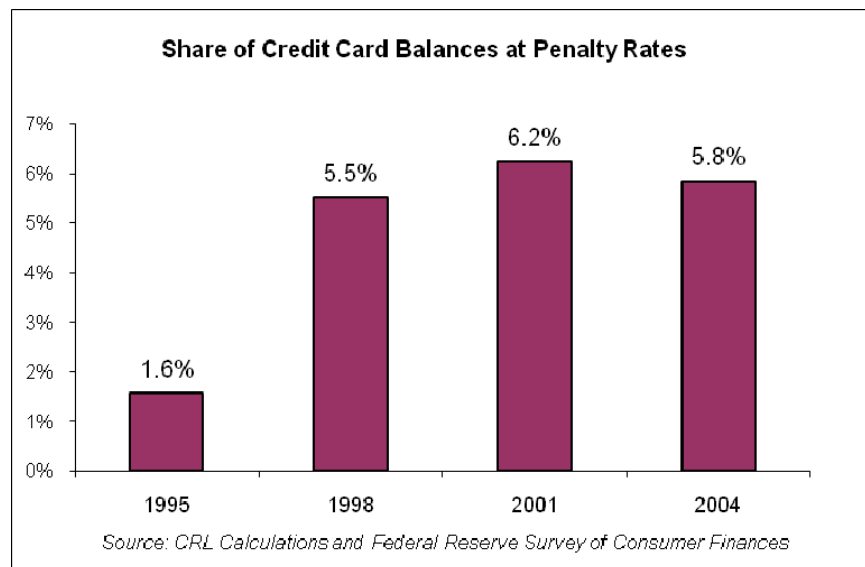


Figure 2: Share of Credit Card Balances at Penalty Rates

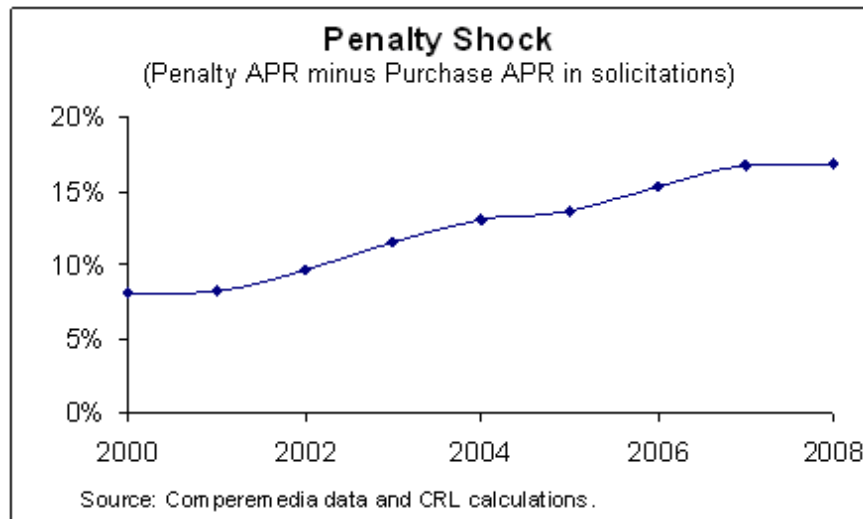


⁹ Joshua M. Frank, "Priceless or Just Expensive? The Use of Penalty Rates in the Credit Card Industry," Center for Responsible Lending, December 16, 2008, available at: <http://www.responsiblelending.org/credit-cards/research-analysis/priceless-or-just-expensive-the-use-of-penalty-rates-in-the-credit-card-industry.html>



There was also a growing disparity between the penalty rate and the regular purchase APR, something we referred to as “penalty shock” (see Figure 3). In addition, it turns out that the penalty rate is an underlying price dimension that remains hidden, with the Survey of Consumer Finances suggesting most people (54%) were unaware they were being charged this dramatically higher rate.¹⁰

Figure 3: Penalty Shock (Penalty APR minus Purchase APR in Solicitations)



The Signal Comes to Dominate and Drive Product Design

While there are certainly signal distortions in the purchase of a new car, basic car design still is centered on useful attributes such as space, power, fuel efficiency, reliability, etc. The majority of innovations likewise focus on improving these product features. For credit cards, however, much of what is called product innovation comes in the form of new pricing structures that create perceived value without creating underlying benefits (or conversely that maintain perceived cost while actually increasing the underlying cost). An example of a product designed around creating perceived value through a distorted signal is a “0% for life” offer on certain credit card account balances. The offer is only profitable due to repricing mechanisms such as penalty APRs and payment allocation rules that create higher prices for the majority of consumers than what the offer would seem to imply.

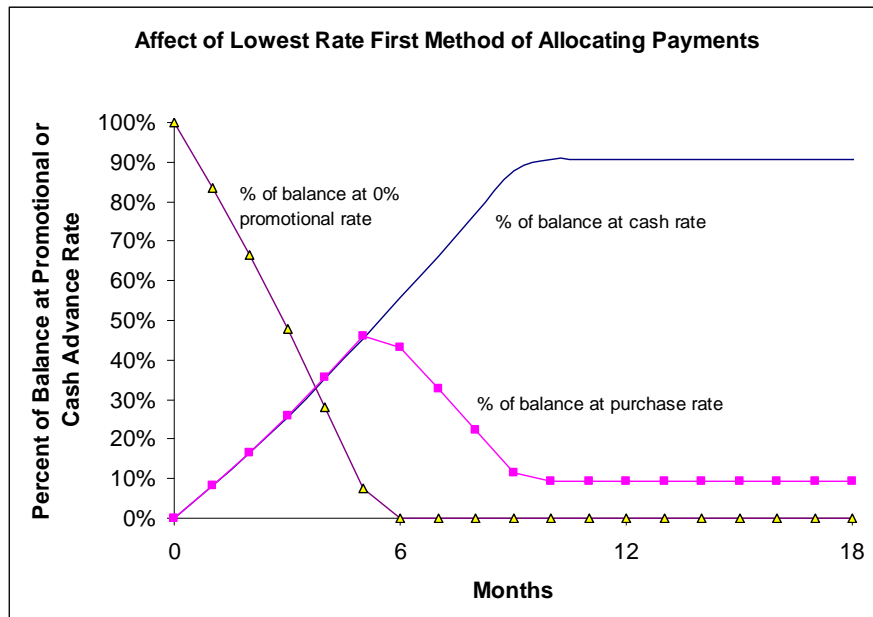
The pervasive use of teasers also drives the design of credit card products, other dimensions of pricing, and the technology used to support those products. Other examples of products designed around the price structure include “fee harvester” cards that are targeted to the subprime market and often feature a

¹⁰ *Id.* at 6.



moderate APR, but have a low line of credit and numerous front-end fees.¹¹ These fees, including application fees, monthly maintenance fees, and similar recurring charges often consume the majority of the consumer's credit line. Even card products targeted to the prime market may let fee income considerations influence credit lines. Some issuers might offer low lines while issuing multiple accounts to the same household, a strategy that seems unlikely to be driven by risk (given the aggregate line extended to the household) but instead may arguably be an attempt to collect higher fee income.

Figure 4: Affect of Lowest Rate First Method of Allocating Payments



One of the most effective shrouded pricing mechanisms issuers have used has been structuring the allocation of payments to maximize revenue without changing the stated APR for any balance category. This mechanism has greatly influenced product design, resulting in the proliferation of multiple balance categories on a single account as well as teaser rates and other price gimmicks. When analyzed, payment allocation practices are found to be powerful, poorly understood by consumers, and inefficient.¹² Figure 4 shows just how powerful the issuer practice of allocating payments to the lowest rate balance first can be using a particular scenario, where a consumer starts with a teaser rate balance and makes regular

¹¹ For further discussion of fee harvester cards, see Rick Jurgens and Chi Chi Wu, "Fee-Harvesters: Low-Credit High-Cost Cards Bleed Consumers," National Consumer Law Center, Nov. 2007, available at www.consumerlaw.org/issues/credit_cards/.../FEE-HarvesterFinal.pdf.

¹² Joshua M. Frank, "What's Draining Your Wallet? The Real Cost of Credit Card Cash Advances," Center for Responsible Lending, December 16, 2008, available at: <http://www.responsiblelending.org/credit-cards/research-analysis/what-s-draining-your-wallet-the-real-cost-of-credit-card-cash-advances.html>



payments, purchases, and cash advances.¹³ In this example, the promotional rate balance disappears in six months, rendering it useless at that point even if the offer gave the promotion for a full year. After month six, the purchase balance rapidly declines leaving almost all of the customer's balance at the cash advance rate which is typically much higher.

The Evolution of "Fragmentation"

With fragmentation, a uniform price or quality is broken into smaller pieces for the purpose of creating signals. Unlike the traditional literature on signaling where information is naturally fragmented and the signal is used to improve information,¹⁴ in a peacock market fragmentation is actually an information reduction strategy. A firm manipulates a product design to shroud underlying pricing or quality while presenting prominently the factors optimized to draw in consumers.

Credit card pricing has become dominated by prominent (but misleading) price signals combined with shrouded revenue enhancement mechanisms to compensate for the cost of these signals. While a uniform price — or at least not highly disparate price — over time is more informative and more consistent with product cost, the price of credit cards over time is intentionally fragmented into two general components. The first is a highly prominent price signal that consumers are known to disproportionately attend to, while the second is a less obvious but larger component of price that is adjusted to make up for losses in the signal price.

Perhaps the archetypical example of fragmentation for any industry could be the credit card teaser rate. In a perfect market with rational consumers using a fixed, reasonable discount rate, there would be no reason to price products at 0% for the first 6 months or year, and then a much higher interest rate thereafter. It makes little sense based on the cost of lending or any other economic rationale. APRs on credit cards were at one time uniform. The primary purpose of disaggregating the near term interest rate and charging a separate price appears to be to create a powerful price signal. But this is not an information-enhancing signal as the traditional economic literature suggests. Instead, unnecessary complexity is introduced to create a signal that is lower than the long-term cost of credit. Due to cognitive constraints and limited attention, the information level digested by the consumer is probably lower due to the creation of the signal. This is what the signal is intended to do in peacock markets. It creates a compelling impression of low prices despite the fact that the total price paid by the typical consumer is much higher. In some cases

¹³ In this scenario, the consumer starts with a \$1,000 promotional balance and makes purchases and cash advances of \$100 a month each, while making payments in an amount equal to this monthly activity (\$200 a month).

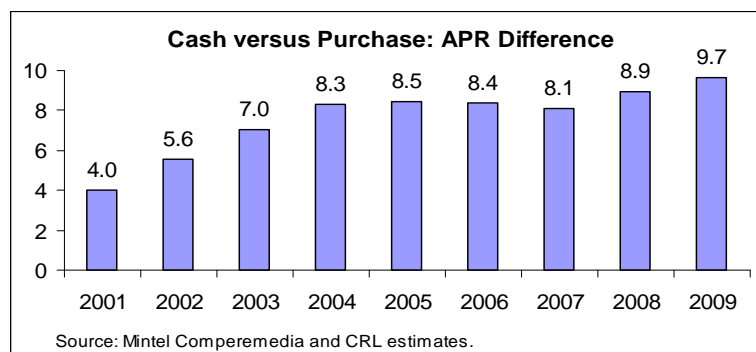
¹⁴ See, e.g., Spence, "Signaling in Retrospect."



these signals will be constrained by their cost or by reduced effectiveness beyond a certain point. However, in the case of teaser rates, the signaled price evolved over time to become lower and lower until a natural limit is reached; currently the most common teaser rate is 0%.¹⁵

The penalty shock from penalty rates (shown in Figure 3 above) is evidence of growing fragmentation. Penalty rates create a false dichotomy in pricing by typically charging much higher rates to customers with only very minor violations (such as being a single day late). The triggering mechanisms of these penalty rates have grown broader over time to include less risky consumers while at the same time the disparity between regular and penalty rates has increased. This suggests that penalty rates have probably risen due to the fact that they are underweighted in consumer decisions, creating a higher level of fragmentation. There is also a growing disparity between the regular purchase and cash advance rates, with the difference between these two rates more than doubling between 2001 and 2009 (see Figure 5). While there is a risk difference between cash advance and purchase balances, it is highly unlikely that the probable loss rate on cash advances is 10 percentage points higher than the loss rate on purchase balances.¹⁶ It is also highly unlikely that this risk differential has more than doubled in the span of a few years. The more likely explanation is that this is an example of fragmentation that takes advantage of which price signals are overweighted and underweighted. While the most highly-weighted dimension of APR is probably the teaser rate, the regular purchase rate likely receives the second most weight. Cash advance APRs and other conditional APRs probably receive little attention given the number of fees and other terms consumers must simultaneously digest. Previous research backs this up, finding that half of people, even when they are cash advance users, do not know that they are charged a different rate for these balances.¹⁷

Figure 5: Cash versus Purchase: APR Differences



¹⁵ Conclusion based on analysis of data from Mintel Comperemedia.

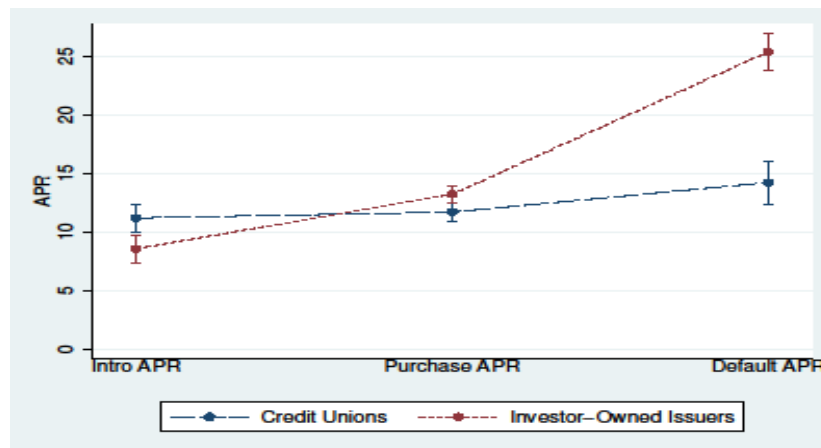
¹⁶ In addition, the price difference between these two balances is even higher due both to cash advance fees and to payment allocation methods that cause the cash advance rate to persist longer than it normally would.

¹⁷ Frank, "What's Draining Your Wallet?"



Further evidence of fragmentation can be observed in a study by Ryan Bubb and Alex Kaufman.¹⁸ The authors compare the credit card practices of banks to credit unions and find substantial differences. As shown in Figure 6, credit unions have much more uniform pricing, while bank pricing is highly fragmented. This is not surprising since the mission of credit unions is to act in the interest of their members, who are also their borrowers while the mission of for-profit banks is to maximize profits. Banks have an incentive to fragment prices in a way that sends powerful signals while raising prices in underweighted areas. Credit unions may not have as strong an incentive to do so because such fragmentation would in effect be deceiving their own members.¹⁹

Figure 6



Source: Bubb & Kaufman, 2009. Reprinted with permission.

“Shrouded” Costs (or Reductions in Benefits) are Prevalent

Credit cards clearly include numerous shrouded costs. Two that have already been mentioned include penalty APRs and payment allocation mechanisms. The payment allocation policy utilized by issuers was found to be a shrouded attribute of the product. Consumer understanding was extremely low. When surveyed, only 3.4% of the public got three basic questions right that were necessary for understanding

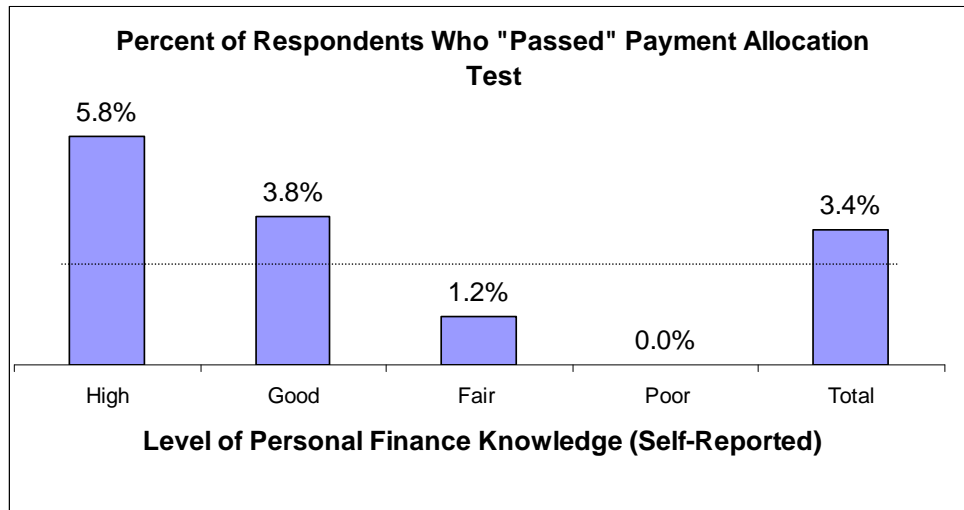
¹⁸ Ryan Bubb and Alex Kaufman, “Consumer Biases and Firm Ownership,” working paper, Harvard University, September 2009, available at http://www.people.fas.harvard.edu/~ryanbubb/papers/BubbKaufman_ConsumerBiasesandFirmOwnership.pdf

¹⁹ While credit unions have much less fragmented prices, this practice is not universal. Some credit unions have offered credit cards in partnership with some of the issuers with the most aggressive pricing (possibly without being aware of the full implications of their practices). Likewise, in other areas (such as overdraft fees), a number of credit unions use the same policies as the large banks.



payment allocation’s impact.²⁰ These were multiple questions with only 2-4 answers. Randomly guessing would have resulted in answering all the questions correctly 6.25% of the time. Even respondents with a high self-reported level of personal finance knowledge scored worse than if they had guessed at random.

Figure 7: Percent of Respondents Who “Passed” Payment Allocation Test



There are also numerous fees that raise the total cost for credit card services. Some of these late fees and over-limit fees are “penalties” for violating an agreement. However, these have evolved from being used as a tool for stopping undesirable behavior to becoming an important revenue stream. In effect, card issuers now want, and in some cases have been found to actually encourage, behavior that causes a cardholder to pay late or go over their credit limit.²¹ Other non-penalty fees are incurred for a wide range of other reasons such as engaging in foreign transactions, taking a cash advance, or account inactivity.

Shrouded reductions in benefits also frequently occur in the credit card industry. These frequently occur in rebate programs. One dimension of rebate benefit reductions may come in the form of restrictions on cashing in rebates. Even more shrouded may be the ability to cancel rebate programs arbitrarily (as some issuers have done recently) or rebate points due to a consumer’s late payment.

If a cardholder benefit is liquidity in case of an emergency, a shrouded benefit reduction also occurs when issuers change terms arbitrarily. Many issuers have recently reduced lines and closed accounts due to

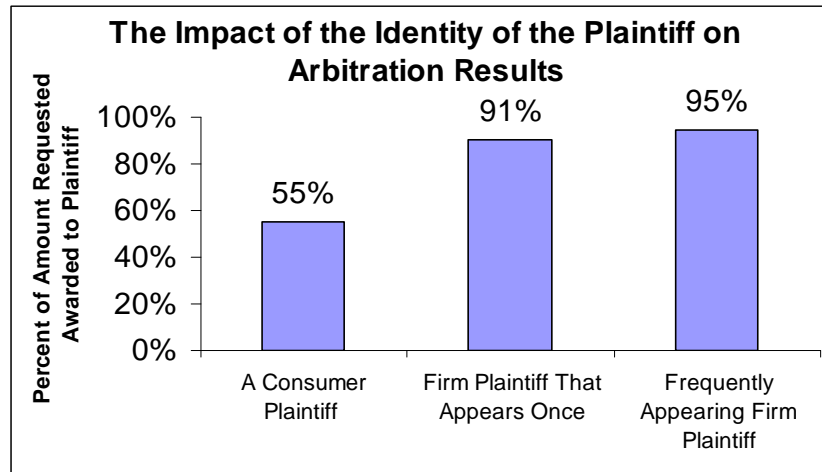
²⁰ Frank, “What’s Draining Your Wallet?”

²¹ Credit card issuers have even sometimes intentionally not credited payments in a timely manner in order to cause consumers to be late. Linda Punch, “Getting Tough?” *Credit Card Management* (Feb. 1, 2005).



current economic conditions. This leaves consumers without emergency liquidity that they relied on when they are most vulnerable.

Figure 8: The Impact of the Identity of the Plaintiff on Arbitration Results

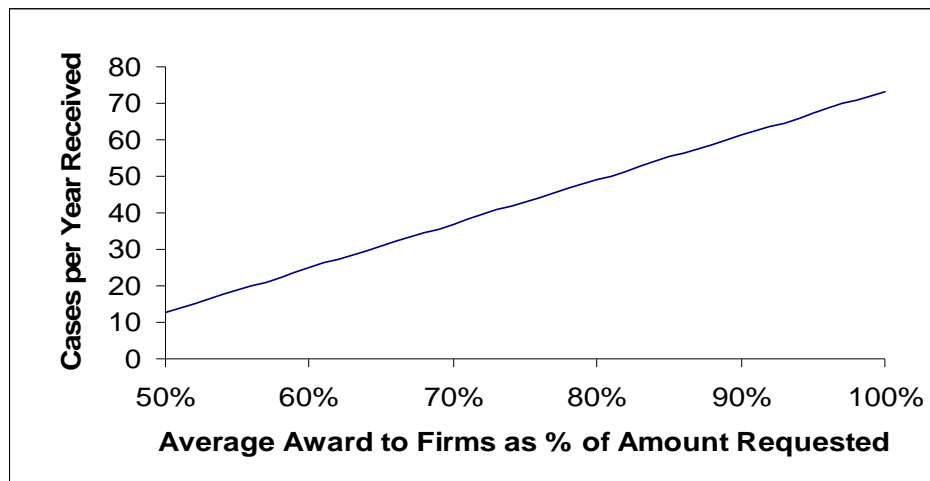


The use of arbitration clauses is an important shrouded benefit reduction. Many borrowers are unaware of the arbitration clause hidden in almost all credit card terms and conditions, or its potential impact on their ability to settle disputes. Looking at a database of arbitration disputes compiled by Public Citizen and consisting primarily of credit card cases, we found evidence of bias against consumers. All types of firm plaintiffs got better results than consumer plaintiffs, but firms that appeared frequently got the best results (see Figure 8). In addition, arbitrators had a clear incentive to be biased, with arbitrators that sided with firms more receiving a larger number of cases in later periods (see Figure 9). The bias in arbitration combined with lack of access to the court system can result in a shrouded benefit reduction in a number of ways. For example, credit card issuers often tout the limited fraud liability consumers have when they use a credit card. However, that right depends on a credit card issuer not ignoring claims of fraud. And in some cases when issuers persisted in pursuing a debt despite claims that it was incurred through fraud, arbitrators have sided with the issuers and ignored the fraud claims.²²

²² John O'Donnell, "The Arbitration Trap: How Credit Card Companies Ensnare Consumers," Public Citizen, 2007, available at <http://www.citizen.org/documents/ArbitrationTrap.pdf>.



Figure 9: Average Award to Firms as % of Amount Requested



Product and Market Efficiency is Reduced as a Result of the Above

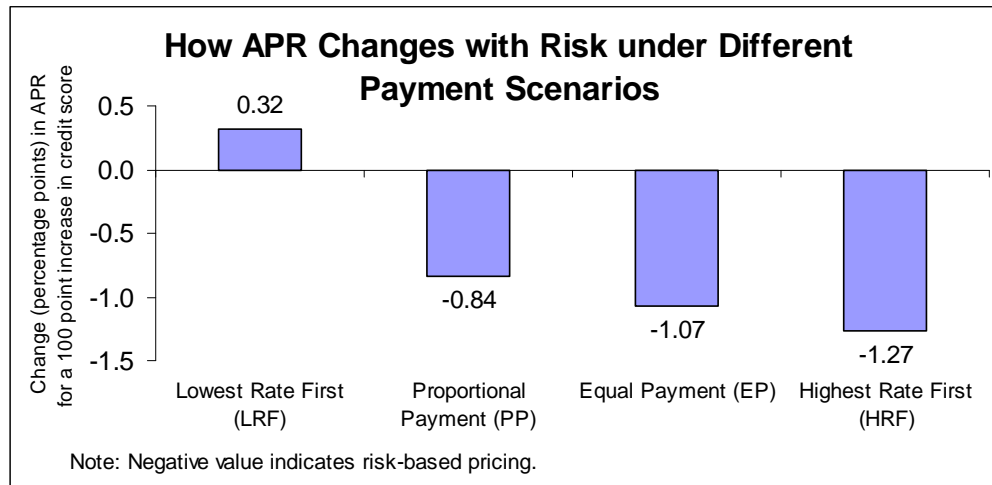
According to economic theory, sending the wrong price signals generally will cause market inefficiency. Teaser rates, for example, cause inefficiency by creating the wrong incentives for borrowing funds and encourage excessive short-term borrowing. Prices that are too high can also send the wrong signals to the extent that consumers are aware of them.

Previous research has also found other evidence of inefficiency. Figure 10 shows the impact of various methods of allocating payments made by consumers on risk-based pricing. Although the practice will soon be outlawed by the enactment of the Credit CARD Act, issuers have almost universally selected the lowest rate first (LRF) method of allocating payment (i.e. the teaser balance is paid first, before the purchase balance, which is paid before the cash advance balance). We found that issuers choose this method despite the fact that it is the least efficient payment allocation mechanism. In fact, this method of allocating payments resulted in inverse risk-based pricing, where the least risky consumers are charged the highest rate.²³

²³ Frank, "What's Draining Your Wallet?"



Figure 10: How APR Changes with Risk under Different Payment Scenarios



Considering penalty repricing, it was found that issuers typically did not notify consumers of the price change. (As mentioned previously, a related finding is that most borrowers did not know when they were at a penalty rate.²⁴) While legally permissible right now, this policy counters arguments that penalty repricing reduces moral hazard. Minimizing consumer knowledge certainly creates inefficiency, especially if it perpetuates moral hazard.²⁵

Conclusion

Credit cards have evolved from a product with one price to a product with several interest rates for different types of activity that can change for a number of reasons related to timing, index rates, triggering events, or simply the issuer’s choice. Credit cards also have a wide range of fees and add-on products. Much of price complexity in credit cards as a lending product is related to an attempt by firms to create misleading signals that cater to biases. This mix of prominent price signals in overweighted domains combined with revenue enhancement mechanisms in underweighted domains describes an archetypical peacock market.

Since this market structure is inherently inefficient, market intervention has been and continues to be appropriate. Attempts to deceive consumers and exploit biases should not be costless. Issuers need to fear some consequences for these actions other than that they will eventually be painlessly phased out. For

²⁴ Frank, “Priceless or Just Expensive?”

²⁵ If a borrower knew they were at a penalty rate and further knew that certain behavior (such as paying on time for six months) would cause their account to return to the regular rate, this would encourage responsible behavior.



example, the Credit CARD Act had a nine-month delay for implementation of most provisions and the Federal Reserve rules allowed a full year and a half for issuer transition. The goal (particularly with the Federal Reserve rules) appears to have been to make the transition as easy and costless for issuers as possible. But this leaves the incentives in place for issuers who have a peacock-market strategic philosophy to try new tricks without fear of adverse consequences.

Furthermore, while very valuable to consumers, the improvements from the Credit CARD Act are static and are not a long-term solution to a rapidly evolving peacock market. Ongoing monitoring and intervention (as necessary), such as with the proposed Consumer Financial Protection Agency (CFPA), could offer a long-term solution.



Marketing Credit Cards in Postcommunist Russia

Alya Guseva[†]

One of the fundamental questions faced by nearly every retailer trying to promote a new product is how to build demand. In most cases, new products are competing with similar products already in use, and their promoters have to underscore how new products are comparable to the old ones, but are at the same time better – faster, smaller, or cheaper. In some cases, however, the products that are offered are entirely new in their class. These are the products that revolutionize the ways in which we communicate (such as telephones when they were first introduced), transport (automobiles) or pay (credit cards). The particular challenge of building demand for these products is the novelty factor.

The novelty of credit cards is two-fold. First, cards offer a new means of payment to people who historically relied on cash only. Second, they are a source of quick general-purpose credit. While some limited forms of lending to consumers existed in socialist economies, by and large, daily experience of shortages on most consumer goods resulted in the so-called “monetary overhang” – too much money held by households, and not enough products to buy. Most saved upfront and paid upfront rather than relying on credit. So for cultural and historic reasons, one should not expect that postcommunist consumers and merchants would eagerly embrace cards, especially since they are expected to pay for the privilege and convenience.

Besides the novelty factor, an additional challenge for credit card promoters stems from the fact that credit card markets are what economists call “two-sided markets.” These are the markets that through an intermediary (in this case, a bank that issues cards) attempt to connect two segments, merchants and consumers; and each of the segments is sensitive to how well the intermediary performs in the other segment. Cardholders and merchants are said to be mutually *complementary*, as one group cannot function without the other, and the growth in each group makes joining the other one more attractive. The more cardholders there are, the more merchants would be willing to accept cards, and vice versa. The reverse is

[†] Alya Guseva is Associate Professor of Sociology at Boston University. Her interests include market emergence and consumer credit, especially in the context of the East European postcommunist transition. She is the author of *Into the Red: The Birth of the Credit Card Market in Postcommunist Russia* (Stanford: Stanford University Press, 2008). For this project, she conducted a series of interviews with representatives of Russia’s banks in 1998-1999 and 2003-2005. Professor Guseva is currently working on a comparative institutional analysis of credit card markets in the postcommunist world, this time focusing on Hungary, Poland, Czech Republic, Russia, Ukraine, Bulgaria, China and Vietnam (in collaboration with Akos Rona-Tas of UCSD). You can direct your questions and comments to aguseva@bu.edu.



also true: no cardholders – no merchants. Unless the vicious circle is broken, the market simply would not take off. So where should card issuers start?

American credit card issuers that initiated their first mass programs in the late 1950s and 1960s initially targeted consumers. As told in the award-winning book by Joe Nocera, *A Piece of the Action*, Bank of America – home to what will later become known as Visa – decided that the best strategy was to create cardholders by “dropping” cards onto a large number of consumers at once. “That’s a word they liked to use in the credit card business to characterize a mass mailing of cards: “a drop,” and it is an unwittingly apt description. There had been no outward yearning among the residents of Fresno, [California] for such a device, not even the dimmest awareness that such a thing was in the works. It simply arrived one day, with no advance warning, as if it had dropped from the sky.”¹ This is how the first 60,000 BankAmericard cards (precursors of Visa cards) were mailed to residents of Fresno in 1958. By the end of next year, Bank of America mailed 2 million more cards to unsuspecting residents of several bigger Californian towns and cities, including “untrustworthy” Los Angeles, a home to “the fast Hollywood crowd, the blue suede shoe boys” (in the words of a Bank of America executive).² Eight years later, during the pre-Christmas sale of 1966, several Chicago banks mailed another 5 million unsolicited cards.

What is so remarkable about these mailings? After all, millions of Americans today routinely receive “pre-approved” offers for credit cards in the mail. What was different then was that individuals received *actual* cards, ready to be used in merchants’ establishments, without applying for them, without being screened, and apparently, even without their names being verified. So in some cases, cards were mailed to prison inmates, persons long deceased, infants, and even dogs. For example, Shepherdson reports that “[a] dachshund named Alice Griffin was sent not one but four cards, one of which arrived with the promise that Alice would be welcomed as a “preferred customer” at many of Chicago’s finest restaurants.”³

As a result of issuing cards indiscriminately and entirely forgoing preliminary screening, after the first 15 months of Bank of America’s credit card program, its official losses amounted to \$8.8 million – a huge sum of money for a bank in that era. Instead of the expected 4% of delinquent accounts (average for loans), they comprised 22%. Fraud was rampant and collections – another risk management mechanism – were problematic, as the bank never even established a special collections department, so confident it had

¹ Joseph Nocera, *A Piece of the Action: How the Middle Class Joined the Money Class* (New York: Simon and Schuster, 1994), 15.

² *Ibid.*, 29

³ Nancy Shepherdson, “Credit Card America,” *American Heritage* 42 (1991): 128.



always been in its clients. The situation was especially difficult in Los Angeles. There, Bank of America faced a major moral hazard problem, as credit cards seemed to corrupt even once-good clients. New cardholders perceived credit lines as free money, and ran up bills without any intention of paying back.

The Chicago experiment with unsolicited mailing was even more disastrous: credit cards corrupted merchants and postal workers. Merchants worked together with criminals to supply fraudulent slips to banks, while postal workers were discovered "carting off bags of unmailed credit cards to sell on the black market."⁴ Some cardholders disputed purchases they had actually made, and banks preferred to pay rather than argue, given an already negative reputation credit cards had earned. The situation in Chicago went completely out of control when the press reported a story of a bank clerk who slipped and fell while carrying several boxes of unprocessed merchant slips that were blown away by the wind. Another time a truck loaded with new credit cards flipped over and newspapers published photographs of people grabbing handfuls of cards for future use. Total losses of the Chicagoan banks were estimated at \$6 million, though some analysts believed it to had been no less than \$25 million.

The magnitude of the losses must have caught issuing banks by surprise. Yet, had they been more careful, we may never have lived to see or use credit cards. The losses were a price to pay for solving the complementarity problem. At the time that these decisions were made, bank executives believed this was their only chance to signal to merchants that there would be enough cardholders to make it worthwhile for them to accept cards. History proved that they were right. In addition to three hundred merchants who signed up in advance, the mass-mailing in Fresno drew another eight hundred more merchants in the next five months. By the end of 1959, about the same time it posted multi-million dollar losses, the BankAmericard program attracted more than 20,000 merchants to accept its credit cards. Eight years later, more than 800 banks were involved in the card business, and 32 million cards were issued in that year alone, the majority through unsolicited mailing. Unsolicited mailings of cards were eventually outlawed by President Nixon in 1970, but by that time, the market was firmly on its feet. Twelve years after the Fresno "drop," BankAmericard cards circulated in 44 American states, Master Charge cards in 49 states, and around 29 million people had used cards at least once, which was more than 20% of the adult population at that time.

Thus, unsolicited mailing of cards helped American banks solve the initial problem of complementarity; as a result of signing up a lot of cardholders, banks managed to attract merchants, which prompted more cardholders to join. The price that banks had to pay for this in the short run was steep: as applicants'

⁴ Nocera, *A Piece of the Action*, 58



screening was foregone in the haste of getting cards to as many people as possible in the shortest time possible, banks incurred serious financial losses. Were the banks that undertook card programs not as big and stable, and had they not had as much organizational slack, they would not have likely recovered financially. Yet, the market was successfully born, and the staggering losses accompanying its early development were soon forgotten by all but business historians.

Parts of Nocera's book were translated into Russian and published in the mid-1990s by one of Russia's leading magazines covering the plastic card industry. So it is likely that many of the Russian bankers engaged in promoting cards knew of the story. But they undoubtedly took it as an example of how *not* to market cards. They appreciated the challenge of developing a two-sided market, but would never dare "dropping" cards the "American way" because in the context of the transitional Russia it could have been a suicidal tactic. As much as they did not trust their prospective customers, they were also suspicious of mail being intercepted or stolen. In fact, in the 1990s, Russian banks did not even mail monthly statements, citing considerations of confidentiality and security, and requiring cardholders to pick them up in person at their local branch. At any rate, most of the banks were small, undercapitalized, and lacked organizational slack that helped Bank of America to withstand the blow of losses and persist with its credit card program.

What the Russian card issuers did instead was to issue cards after an exhaustive analysis of each case to *insiders* – applicants who had a direct or indirect social tie to the bank, were visible politically or socially, or were anchored in some other way that made the bank nearly confident they would not default.⁵ For example, cards were issued to bank's own employees, to family and friends of the bank's top executives, to top executives and other employees of banks' corporate clients, or to politicians, singers, and TV personalities. Here the borrower-creditor relationship was intermingled with workplace ties or close social bonds. For example, issuing a card to a friend of a bank manager was quite safe in the eyes of banks administration. The manager's position in the bank – as well as the manager's relations with the friend – were at stake, and it was expected that the manager would do anything to ensure that the friend did not violate trust placed in him or her. Faced with uncertainty and unable to predict the risk of future defaults, banks turned to various strategies of controlling cardholders' future behavior through social networks and organizational power.

⁵ The majority of cards issued during this period were secured debit cards (exceptions were made for particularly well-connected and trusted cardholders) that pre-supposed both an account to pay for regular charges, and a security deposit account to serve as protection measure for the bank. Because of a predominant use of off-line authorizations, imprinters, and slow information exchange, these cards allowed for "technical" (unauthorized) overdraft despite being positioned as debit cards. Banks recognized that such cards were quite risky for the bank to issue, and required security deposits to protect themselves from intentional or unintentional overspending or fraud. In the early 1990s, security deposits on Gold cards were sometimes as high as \$10,000.



While this strategy left bank fairly confident in its cardholders, it led to only limited market expansion. And very soon, Russian banks realized, just like Bank of America did 35 years earlier, that slow and careful recruitment of cardholders was like swimming against a strong current. It was not bringing them closer to their destination – a market where the numbers of cardholders and merchants are increasing in a complementary manner, and where growth is generated by demand externalities. They needed an equivalent of a steam ship. And a ship they did build. In the subsequent decade-and-a-half, Russian banks pursued two strategies of mass-issuing cards to millions of Russians. However their success in tackling complementarity and uncertainty problems simultaneously remained limited.

At first, Russian banks turned towards employing organizations peddling cards to whole working collectives through *salary projects*. A typical salary project involved an agreement between a bank and an enterprise to issue cards to all of the employees, from the top manager to the janitor, while their salaries were directly deposited to the bank.⁶ Banks' possible losses from unauthorized overdraft or misconduct in the use of the card would be paid by the next month's salary, and the enterprise would be obliged to notify the bank as soon as the employee was fired or quit. Popularity of salary projects Russia in the 1990s has been mirrored in other postcommunist countries as well.

Salary projects were attractive to enterprises because they allowed them to save on transporting, securing, and dispensing cash. But from the bank's perspective, salary projects were an ingenious strategy of quickly mass-issuing cards without any investment into developing an elaborate network of branches and with little investment in advertising. Prospective cardholders did not have to be enticed. If the management of their company decided on a salary project with a particular bank, workers had no choice but to carry a card of that particular bank. Each salary agreement allowed banks to sign up hundreds or thousands of new cardholders. Banks hoped this growth in the number of issued cards would also make cards popular among retailers. Thus, in a way similar to American unsolicited mailings, Russian banks also engaged in creating cardholders rather than recruiting them. But they went even further. Not only did employing organizations enable banks' access to prospective cardholders, but they also helped them avoid large initial losses that plagued the American credit card market. To the extent that companies exercised control over their employees, they ensured Russian banks' ability to monitor cardholders, and helped them to control uncertainty. The dual role played by cardholders – as customers of the bank and employees of the enterprise – limited their flexibility, reduced their ability to exit the bank-client relationship (which would mean also quitting their job), and made them available for negotiation and sanctioning by the bank. In such

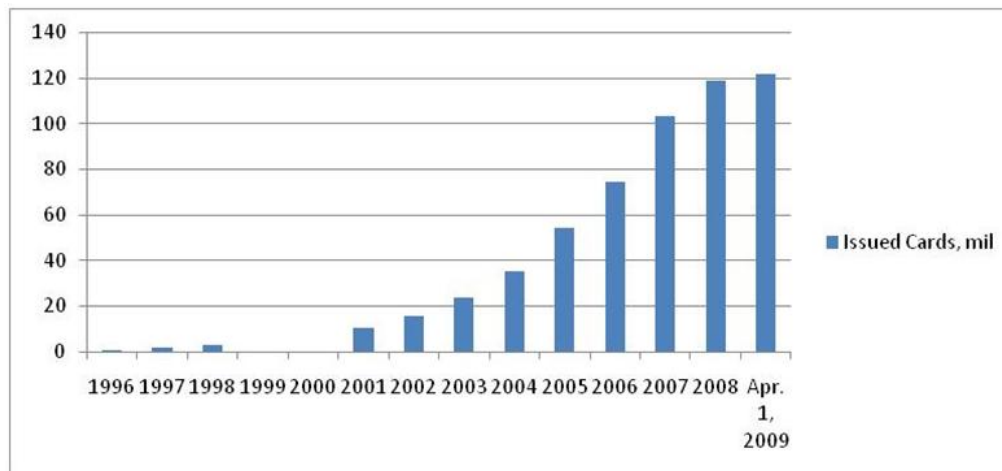
⁶ Before the introduction of salary cards, salaries were paid in cash. Initially, salary cards were debit cards, but eventually, banks started to offer overdrafts for up to 75% of one's monthly salary.



an arrangement, employment became a type of social collateral: as long as it was at stake, cardholders could not easily disappear. Thus, compared to the previous strategy of issuing cards to elites, salary projects were much more successful in propelling the development of the Russian card market forward because they disseminated cards beyond exclusive inner circles to wider masses without compromising on security of card issuing.

The Russian state played a particularly important role in this coercive dissemination of cards. The state was eager to transfer salary payments of its employees (doctors, teachers, members of the police force and the military, etc.) into the banks' hands. Moreover, the idea of salary cards was picked up by municipal governments across Russia, which issued cards to millions of retirees and recipients of other social benefits. For instance, in 2002, the Moscow municipal government through the collaboration between Bank of Moscow and Visa issued Visa Electron Moscow Social Cards to 2.5 million Muscovites who receive some 350 types of subsidies from 60 agencies directly deposited to their card accounts. In addition to bank branches, ATMs, and Visa merchants, the cards can be used to pay for public transit, health and medical insurance, and to make government-subsidized purchases at participating stores.

Figure 1: Number of Cards Issued by Russian Banks, 1996-2009



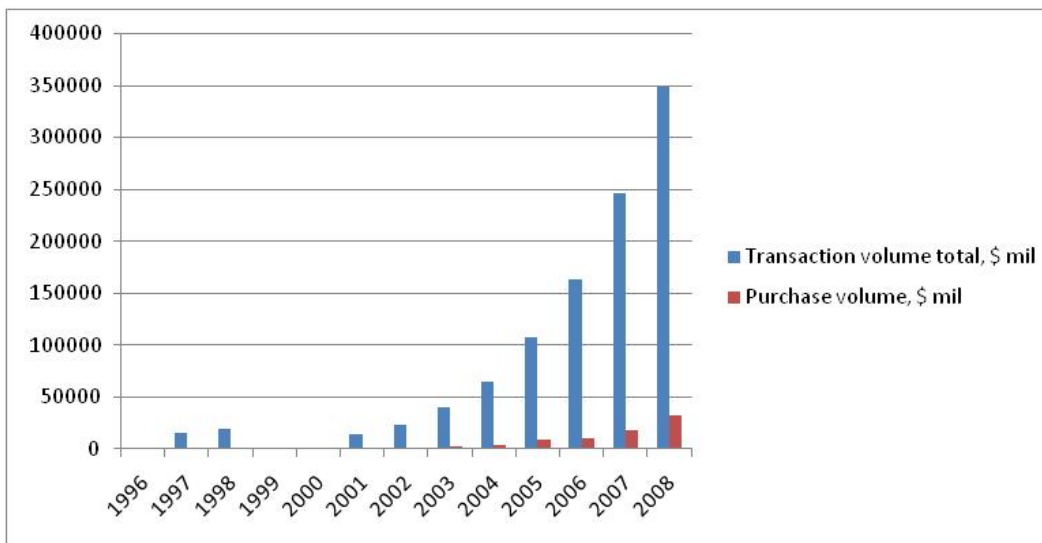
Source: Data from Platezhi. Sistemy. Kartochki (1998, p. 10; 1999, p. 6-7) and Central Bank of Russia (www.cbr.ru). Data are missing for 1999 and 2000.

Have Russian card issuers been as successful as American issuers in attracting merchants and therefore in solving the problem of complementarity? Not really. While the number of issued cards and the volume of transactions were steadily increasing (Figures 1 and 2), the overwhelming portion of that transaction volume was cash withdrawals rather than non-cash purchases (Figure 2). Holders of salary cards went to



ATMs to withdraw cash, and then, armed with cash, to the stores. Whatever spectacular growth the credit card market was undergoing, it was not involving the merchants, and therefore instead of a classic two-sided market development, the Russian card market was developing in a one-sided (really, lop-sided) fashion. Through salary projects, banks managed to coerce individuals to have cards, but they could not make them use cards in lieu of cash.

Figure 2: Volume of Total Card Transactions and Purchase Volume Only, 1996-2008

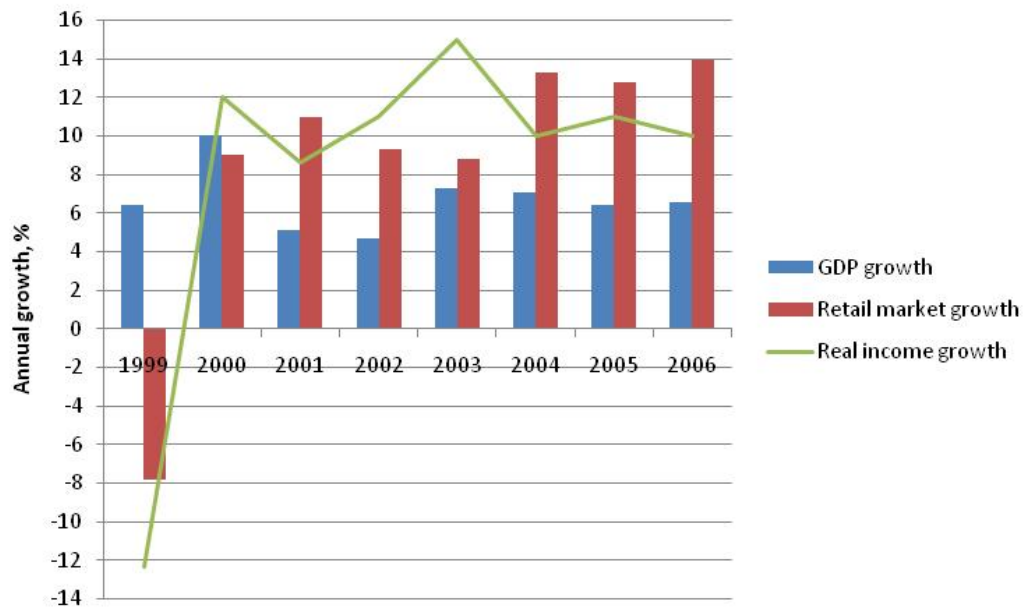


Source: Data from Platezhi. Sistemy. Kartochki (1998, p. 10; 1999, p. 6-7) and Central Bank of Russia (www.cbr.ru). Data are missing for 1999 and 2000.

While salary projects dominated the Russian card market in the 1990s, in the past several years Russian issuers have refocused their energy on mass-issuing cards to Russian consumers in a rapidly growing retail sector. Following a decade-long recession and a devastating financial crisis of 1998, Russia experienced a steady growth of the economy, increase in incomes, and increase in consumer purchasing power. GDP grew 6.4% in 1999, and since then has averaged 6.7% annually (Figure 3). Real incomes have been rising by an average of 11% annually, and the consumer sector as a whole has doubled in size over the last decade. The growth in retail was accompanied by the growth in consumer lending (Figure 4). With consumption and consumer credit taking center stage, banks turned their attention away from employees and towards consumers as the preferred group for marketing cards, and from salary cards to cards issued in connection with *express loans*. For the first time Russian banks have mass-issued cards that come with credit lines, the possibility of revolving the portion of the credit line that was unused or repaid, and sometimes even with grace periods.

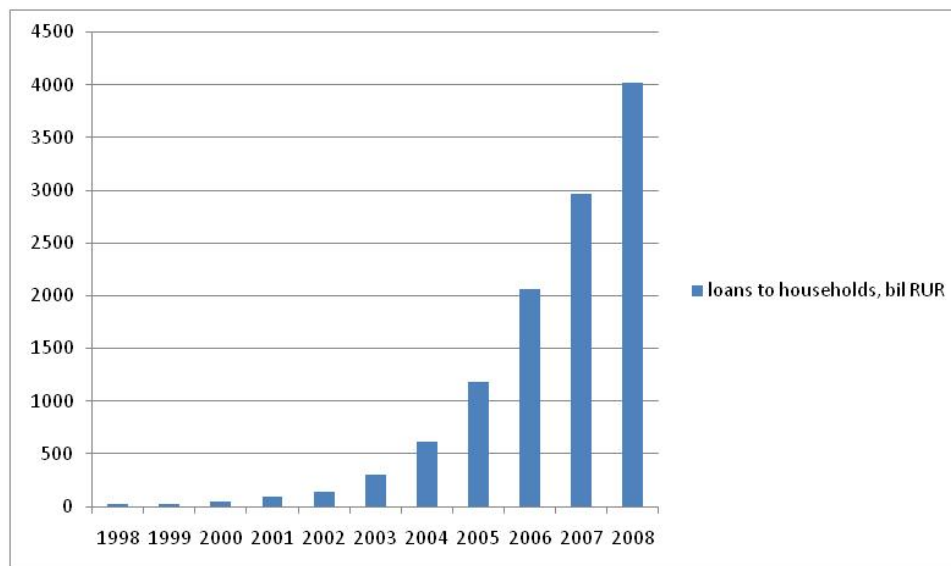


Figure 3: GDP, Retail Market and Real Income Growth Rates in Russia, 1999-2006⁷



Source: Data from Russian Federal Statistical Service, available at www.gks.ru; Alya Guseva, *Into the Red: The Birth of the Credit Card Market in Postcommunist Russia* (Stanford: Stanford University Press, 2008), 113.

Figure 4: Volume of Loans to Russian Households, 1998-2008



Source: Central Bank of Russia, available at www.crb.ru.



Express loans are small short-term loans extended in stores and shopping malls to finance consumer purchases (electronics, furniture, computer equipment). To reach consumers, banks open up makeshift booths with one or two representatives to accept applications. Consumers are approved while they wait and can walk home with goods they have chosen. According to Russian pollster VTsIOM's 2006 opinion poll, 59% of all borrowers obtained their loans in stores, while only 38% in bank branches. Some banks extend these loans in the form of revolving credit cards. DeltaBank, recently purchased by GE Consumer Finance, issues Visa Electron Instant Issue cards that can be used for cash advances up to the specified limit or for retail purchases anywhere Visa is accepted. Others, like Russkiy Standart, the current leader of the Russian credit card market, established in 1999 and in 2004 named one of the most profitable banks in the world by *The Banker* magazine, issues credit cards to those who have successfully repaid their express loans. (Such borrowers become automatically eligible for a credit card). The majority of its cards are revolving credit cards, and the bank considers express loans the main way of increasing their credit card holder clientele. Of the 3.2 million cards Russkiy Standart issued by 2005, only about 100,000 (less than 3%) were issued through bank branches. To attract potential customers, Russkiy Standart works with large stores, such as household electronics retailer M-Video, which reports that 30% to 50% of their sales are now done on credit.

This strategy leaves banks vulnerable, however. Compared to salary projects, the ability of banks to control cardholders has disappeared. Yet screening is at best rudimentary. A few card issuers started to routinely use standard decision-making tools, but the lack of key data (such as verified incomes) and the ways in which these models were constructed (by quantifying applicants' characteristics based on experts' opinions and common sense rather than as a result of statistical analysis) did not bring banks closer to calculating statistical probability of default. Decisions on express loans are made on the spot, usually within 30-40 minutes, and bank officers typically have little to rely upon besides the information supplied by the applicant. In some cases, all that is needed to successfully obtain an express loan is one's passport. Surprisingly, although 76% of banks consider credit risk their biggest problem, according to the information from Russia's Central Bank, they express profound disregard for screening. Several of the banks I interviewed claimed they were much more concerned with attracting new customers and increasing their market share, preferring to worry about uncertainty and developing screening techniques later.

In summary, card issuers in a developing credit card market are faced with twin problems of complementarity and uncertainty. Unlike Bank of America, which in 1959 rushed to saturate Californian towns with credit cards in an effort to solve the problem of complementarity while completely disregarding



uncertainty, Russian card issuers initially decided to proceed with caution. They issued cards to national elites and others connected to the bank. As a result, in the early- to mid-1990s, the Russian card market was small, and except for intentional fraud, essentially risk-free. Salary projects were instrumental in helping the market to expand and to democratize access to cards without exposing banks to greater uncertainty. Both elite issuing and salary projects illustrated the idea that in the absence of means to predict cardholders' future behavior,⁸ existing social ties and organizational structure can help banks control cardholders by putting pressure on them to pay through social ties or precluding their exit by using their employment as a form of collateral.

Russian card issuers' most recent strategy of targeting shoppers in malls and large stores is successful in recruiting a significant number of new cardholders and spurring card use. It solves the problem of complementarity, but leaves banks no tools to deal with uncertainty. Cards are issued to unaffiliated strangers, and verification of information provided by the applicant is quick and often superfluous. The future success of developing credit card markets in Russia and in other emerging economies depends on whether they manage to find an effective solution to the problem of uncertainty. This involves organizing credit bureaus that would accumulate information about borrowers and assist card issuers in pre-screening the applicants.

⁸ First credit bureaus were not founded until 2005.



Who Owns Mobile Money?

Ari Hyytinen[†] and Tuomas Takalo[§]

It has long been predicted that the digital revolution will transform the way people pay at their point-of-sale transactions. The end of the cash era could indeed have dramatic consequences, not least since it might eventually even lead to the obsolescence of central banks.

While many digital payment media innovations have failed to take off and the paper form of cash has turned out to be surprisingly resilient, many commentators argue that mobile phones and other similar portable electronic devices finally enable digital cash to solve its chicken-and-egg problem.¹ For example, mobile phones can currently be easily equipped with payment features, and since almost all consumers in the developed countries carry mobile phones, shopkeepers are willing to install systems to handle mobile money. The prospects of mobile money are even greater in poorer countries where it often constitutes the only feasible digital payment medium. Moreover, mobile phones' display and ability to act as a mobile ATM add to mobile money's utility. But in contrast to the supply of paper cash, which is the monopoly of central banks, it is not clear who owns the property rights over mobile money.

Since the mid-1990s it has been clear that the emergence of new mobile technologies represents an untapped business opportunity. But who has tried to seize the opportunity and invested in the

[†] Dr. Hyytinen works currently as Professor of Economics at the School of Business and Economics at the University of Jyväskylä, Finland. Before joining the university, he worked at the Financial Markets and Statistics Department of the Bank of Finland and at the Research Institute of the Finnish Economy (ETLA). Dr. Hyytinen is a Member of the Scientific Economic Council of the Finnish Ministry of Finance, an Associate Research Fellow of the Research Institute of the Finnish Economy (ETLA), a Docent (Adjunct Professor) of Helsinki School of Economics, and a Member of the Financial Sector Committee of the Central Finland Chamber of Commerce. Dr. Hyytinen's research interests include small business finance and banking, economics of entrepreneurship and industrial economics.

[§] Dr. Takalo works currently as Research Supervisor at the Monetary Policy and Research Department of the Bank of Finland and as Professor of Economics at the School of Business and Economics at the University of Jyväskylä, Finland. Before joining the Bank, he worked as the Director of the Finnish Doctoral Programme in Economics and the Editor-in-Chief of the *Finnish Economic Papers*. Dr. Takalo is a Member of the Scientific Committee at the IPR University Center and a Docent (Adjunct Professor) of Helsinki School of Economics, Turku School of Economics and the University of Oulu. Takalo's research primarily deals with innovation, intellectual property, and financial economics.

¹ See, e.g., our own work (Ari Hyytinen and Tuomas Takalo, "Consumer Awareness and the Use of Payment Media: Evidence from Young Finnish Consumers," *Review of Network Economics* 8 (June 2009): 164-188), and the references therein.



development of mobile payment technologies? What kinds of firms have tried to enter into this new, promising market? Where do these entrants come from?

At least two distinct industries should have had an incentive to innovate in and compete for the market for mobile payments. On the one hand, traditional incumbents in the market for payment media are financial institutions, such as banks and payment card companies, who obviously had a head start in capturing the emerging new market. On the other hand, ICT firms, equipment makers, and telecommunications operators, for example, are specialists of mobile communications technology and have been vying for new applications and revenue sources for their devices and services. Moreover, the economics of payment media markets is quite similar to that of the communications industry: both are two-sided markets characterized by network externalities and platform competition.

It is, of course, equally possible that new entrants could come outside these two main candidate industries. For example, Internet marketplaces and service providers, software and operating system producers (such as Microsoft and its rivals and collaborators), or completely new research-based entrants with unconventional business models could be interested in developing mobile payment media applications and entering the market.

The new mobile and digital payment media markets are emerging. While their eventual landscapes are yet unknown, patent statistics provide a window to the future and potential entry, characterizing innovative activities and intellectual property strategies of potential entrants at the birth of new mobile payment media.

Based on our own preliminary data analyses and Hall, Thoma, and Torrissi's research, some clear and interesting patterns seem to emerge from the patenting data.²

First, as Table 1 shows, patenting activity picked up late 1990s and accelerated after the millennium. This pattern emerges from the world-wide patent data on mobile payment technologies to which we have had access (covering publications from 42 patent offices and coming from Derwent World Patent Index - database). The same pattern can also be extracted from the European Patent Office (EPO) data on European financial patents compiled by Hall, Thoma, and Torrissi. While their study aims at covering a broad range of financial patenting, their data appear to cover a lot of payment methods and technologies. Hence their results could be seen as providing an upper bound for payment innovation patenting in Europe.

² For details, see Bronwyn Hall, Grid Thoma, and Salvatore Torrissi, "Financial Patenting in Europe," NBER working paper no. 14714 (2009); Bronwyn Hall, Grid Thoma, and Salvatore Torrissi, "Financial Patenting in Europe," *European Management Review* 6 (2009): 45-63.



Table 1: Average Number of Patent Applications per Year

Time period	Mobile payment patents*	EPO Financial patents**
Before 1990	1	16
From 1990 to 1994	9	49
From 1995 to 1999	45	257
From 2000 to 2002/2005	258	342

* = Based on the authors' calculations using data from Derwent World Patent Data Index -database and keyword search ("mobile" OR "cellular AND "payment" OR "billing" OR "transaction" OR "paying" OR "electronic money OR "commerce" OR "cash"). Data ends to 2002.

** = Based on the authors' calculations using data from B. Hall, G. Thoma and S. Torrissi, 2009, "Financial patenting in Europe" NBER Working Paper No 14714, Table B.1. Data ends to 2005.

Second, U.S. firms are the most important source of mobile payment patent applications. In Hall, Thoma, and Torrissi's research, it is reported that 49.9% of EPO financial patents originate from the U.S. In our technologically targeted but geographically broad data, the U.S. innovators account for 31% of the (global) patent applications.

Third, the involvement of incumbents (financial institutions) is modest. Some payment card platforms figure in the statistics, but banks and other financial intermediaries are nearly absent. In our patent data on mobile payment technologies, traditional financial institutions are virtually entirely absent; among the top 20 patentees, none is coming from the traditional financial sector. From Hall, Thoma, and Torrissi's research, we can infer that out of 52 top financial patentees in EPO, nine (17%) come from the financial and insurance sector. These nine include four major payment card platforms and service providers (First Data, Mastercard, Visa, and American Express) but only few financial intermediaries.

In contrast, established ICT firms, device manufacturers, and operators in particular, seem to account for most of the mobile payment and European financial patents. In Hall, Thoma, and Torrissi's data, top five patentees include IMB, Citicorp, NCR, Fujitsu, and Siemens. In our (unweighted) data, they are Ericsson, Siemens, Nokia, Motorola and IBM. This suggests that ICT firms, which are entrants into the financial service sector, have begun to create and compete for a new market.

Interestingly, the patterns of patenting of mobile payment technologies are in stark contrast to some other areas of financial innovation, such as financial exchange systems and infrastructures. In this area, most



innovations (according to the patent statistics) arise from the traditional incumbents, such as investment banks and financial exchanges.³

There are a couple of potential explanations for the absence of financial institutions and dominance of the ICT firms in the patenting of (mobile) payment technologies.

The first potential explanation is that the different industry backgrounds of the most likely entrants have repercussions for their innovative strategies. Investments in R&D and IPR management have long been the core competitive strategies in telecommunications industry, whereas financial institutions have hardly bothered to document their R&D investments. This view suggests that both incumbents and entrants may innovate equally but use different intellectual property strategies. It is possible that financial institutions waive patent protection, resorting to their traditional appropriability strategies (e.g., lead time and secrecy) to protect their mobile payment innovations, whereas ICT firms just follow their patent-based intellectual property management systems.

However, it is well documented that the Court of Appeals for the Federal Circuit's landmark 1998 decision in *State Street Bank & Trust Co. v. Signature Financial Group, Inc*⁴ has raised the awareness of intellectual property issues not only in the U.S. but also in the global financial services sector. In some cases the *State Street* decision is also known to have drastically changed the management of financial innovations in financial services, prompting a large scale use of patents as an appropriability strategy. For instance, the active innovators and patentees in the field of financial exchange systems and infrastructures have been investment banks and exchanges themselves. It therefore seems a bit unlikely that different levels of awareness and interest would account for the difference in the patenting patterns over the past ten years, i.e., over the period when the most of mobile payment patent applications have been filed. Indeed, if one only looked at the overall patenting of payment inventions over time and across regions, one could easily make the misleading conclusion that the *State Street* decision prompted the U.S financial institutions to patent their payment method and technology innovations, too.

The *State Street* decision also implies that the research and innovations of financial institutions are hardly unsuitable for patenting. Moreover, many of the (mobile) payment innovations are inherently technological

³ See Mari Komulainen and Tuomas Takalo, "Does State Street Lead to Europe? The Case of Financial Exchange Innovations." Bank of Finland discussion papers 22/2009 (2009).

⁴ 149 F.3d 1368.



and as such have been always patentable. For example, according to the USPTO, one of the first U.S. patents was granted on March 19, 1799, for an invention used for "Detecting *Counterfeit Notes*".⁵

An alternative explanation for the lack of patenting by financial institutions in the area of mobile payment technologies is that they do not innovate. This fits well with the classic argument, associated with Kenneth Arrow, according to which incumbents generally may have weak incentives to innovate since they recognize that the new innovations cannibalize the revenues from their existing products.⁶ More recently, Raghuram Rajan and Luigi Zingales emphasize that incumbents may prefer to conspire with the politicians to preserve the status quo and to prevent entry rather than to engage in innovative activities and competition for new markets. This could account for the cross-country variation in the pace of financial development over time.⁷

Whether some firms can eventually create and conquer the market for mobile payments and enforce effective property rights over the mobile money so created is a question on which existing data or research has little to say. Since this is an industry in which network externalities are crucial, it would be tempting to predict that the market will eventually "tip" towards a dominant solution. But politics and regulation creates a particular source of uncertainty, as the global financial industry and therefore a large part of the payment media markets (e.g., deposit services) still operate under extensive regulation. The limited prospects for further deregulation in this area suggest that potential entrants with the most radical mobile payment media innovations face a rocky road ahead.

If we had to bet a euro now on the future of mobile money, we would bet it on a rather fragmented market outcome, with different technological solutions and business models co-existing in different geographical and business areas.

⁵ Patentability of payment media innovations is discussed further in Robert M. Hunt, Samuli Simojoki, and Tuomas Takalo, "Intellectual Property Rights and Standard Setting in Financial Services: The Case of the Single European Payment Area," in *Financial Innovation in Retail and Corporate Banking*, eds. L. Anderloni, D.T. Llewellyn, and R.H. Schmidt (Edward Elgar: Cheltenham, UK, 2009).

⁶ Kenneth J. Arrow, "Economic Welfare and the Allocation of Resources for Inventions", in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, ed. R.R. Nelson (Princeton University Press: Princeton, 1962).

⁷ Raghuram Rajan and Luigi Zingales, *Saving Capitalism From the Capitalists* (Princeton University Press: Princeton, 2003). As also Rajan and Zingales point out, the notion that in particular (financial) intermediaries have an incentive to lobby for restrictions in competition goes back at least to Adam Smith (see Adam Smith, *The Wealth of Nations*, Book I, Chapter XI, ed. E. Cannan (1776; Chicago University Press: Chicago, 1976, p. 278).



Two-Sided Bank Card Payment Networks and Public Policy

By David VanHoose[†]

Introduction

The provision of payment-clearing services is a big business for banks and other payment intermediaries. About \$75 in U.S. payments is processed each year per \$1 of U.S. gross domestic product, which translates into more than 90 billion payments valued in excess of \$1,000 trillion per year.¹ Estimates indicate that intermediating payment transactions accounts for about 38% of the largest U.S. banking companies' total operating revenues.² In light of banks' significant role as payment intermediaries, Lacker and Weinberg suggest that more work should be done on "payment economics."³ Most of the work surveyed by Lacker and Weinberg examines large-value, or wholesale, payment systems—such as the U.S. Clearing House Interbank Payment System or the Fedwire system—from central-planning perspectives of monetary economists. Thus, much payment-economics work focuses on funds-market and various liquidity-, credit-, and systemic-risk issues germane to monetary theory and policy. This orientation perhaps helps to explain the employment of overlapping-generations models in the study of payment flows in overnight and

[†] David VanHoose earned his B.A. in economics at Indiana University and his Ph.D. in economics at the University of North Carolina at Chapel Hill. He is currently the Herman W. Lay Professor of Private Enterprise at the Hankamer School of Business at Baylor University. VanHoose has published more than fifty articles in professional economics and finance journals and has authored or co-authored several textbooks. He serves as a co-editor at the *North American Journal of Economics and Finance* and on the editorial boards of several journals. His book *The Industrial Organization of Banking* will be published by Springer in December 2009. He is grateful for comments received from Jack Tatom and for financial support provided by Networks Financial Institute.

¹ Charles Kahn and William Roberds, "Why Pay? An Introduction to Payments Economics," *Journal of Financial Intermediation* 18 (2009): 1-23.

² Lawrence Radecki, "Banks' Payment-driven Revenues," *Federal Reserve Bank of New York Economic Policy Review* 5 (1999): 53-70.

³ Jeffrey Lacker and John Weinberg, "Payment Economics: Studying the Economics of Exchange," *Journal of Monetary Economics* 50 (2): 381-387 (2003).



intraday funds markets.⁴ Only a handful of studies relating to payment issues of acute concern to central banks employ an industrial organization perspective.⁵

Nevertheless, industrial organization research is more relevant to regulation and antitrust issues confronted in retail payment systems, such as bank card payment networks, which over the years have been subjected to a variety of lawsuits, legal regulations, and antitrust actions. On the legal and regulatory front, most attention has been placed on interchange fees. In 1979, National Bancard Association unsuccessfully challenged in U.S. courts the existence of interchange fees, which a payment network charges an issuing bank on behalf of an acquiring bank. A few years ago, the Reserve Bank of Australia banned no-surcharge rules and began regulating interchange fees. In Denmark, interchange fees on domestic card payments are prohibited; in Sweden and the Netherlands, no-surcharge rules requiring retailers accepting a payment network's cards not to provide discounts for cash use are illegal; and in the European Union, the European Commission recently intervened to bring down interchange fees. On the antitrust front, U.S. payment card networks have faced Department of Justice challenges to restrictions on issuers and to a proposed merger between acquirers. Payment card networks have also confronted private lawsuits challenging the legality of "Honor All Cards" rules requiring retailers that honor a network's credit cards to accept its debit cards as well.

This article reviews and evaluates the public policy implications of recent studies that have sought to apply analysis of network externalities and the theory of two-sided markets to card payment networks. These studies reach mixed conclusions about the efficacy of regulating the fee structure of card payment networks. They provide support, however, for antitrust challenges of certain restrictive rules that some card payment networks have attempted to impose.

Two-Sided Markets and Network Externalities

Most analyses of credit card networks are based on the concept of two-sided markets.⁶ As shown in Figure 1, in a two-sided market, one or more *platforms* facilitate dealings between *end users*. In the case of a card-

⁴ See, e.g., Charles Kahn and William Roberds, "Real-time Gross Settlement and the Costs of Immediacy," *Journal of Monetary Economics* 47 (2001): 299-319; Antoine Martin, "Optimal Pricing of Intraday Liquidity," *Journal of Monetary Economics* 51 (2004): 401-424; Joydeep Bhattacharya, Joseph Haslag, and Antoine Martin, Federal Reserve Bank of New York Staff Report No. 281, April 2007; and Cyril Monnet and William Roberds, "Optimal Pricing of Payment Services," *Journal of Monetary Economics* 55 (2008): 1428-1440.

⁵ David VanHoose, "Central Bank Policy Making in Competing Payment Systems," *Atlantic Economic Journal* 28 (2000): 117-139; Cornelia Holthausen and Jean-Charles Rochet, "Efficient Pricing of Large Value Interbank Payment Systems," *Journal of Money, Credit, and Banking* 38 (2006): 1797-1818.



payment network, the network itself—for instance, Visa, MasterCard, Discover, or American Express—constitutes the platform. Network owners, which may or may not include banks, are *platform providers*. End users are retailers and cardholders.

Figure 1: A Two-Sided Card Payment Network

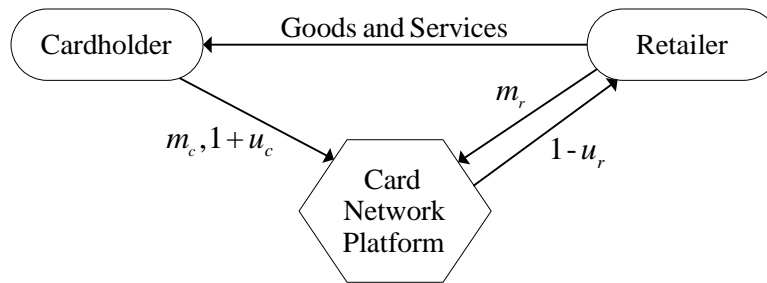


Figure 1 depicts a card network that may charge end users usage fees (u_c and u_r) and membership fees (m_c and m_r). Let's suppose for the time being—see discussion of membership fees below—that the card network platform assesses usage fees only, in which case a cardholder pays a per-dollar fee u_c for each dollar of payments the network transmits to a retailer on behalf of the cardholder. From each dollar transmitted to the retailer, the network also deducts a usage fee equal to u_r . The total per-unit usage charge to end users, therefore, is $u = u_c + u_r$.

If the overall transactions volume handled by the card network platform were to depend solely on this total price—that is, if changes in u_c and u_r with u unchanged had no effect on transactions volumes—then the market would only be one-sided.⁷ In this instance, only the overall price would influence end users. The network platform providers could vary that price without generating unbalanced reactions on the part of end users. In a two-sided market, however, variations in u_c and u_r bring about a change in the aggregate transactions volume even if the overall price u remains unchanged. Thus, changing the relative price structure alters total network activity. The network platform providers must specify a price structure that balances this asymmetry of price effects on the two sides of the market.

⁶ See Sujit Chakravorti, "Theory of Credit Card Networks: A Survey of the Literature," *Review of Network Economics* 2 (2003): 50-68.

⁷ See, e.g., Jean-Charles Rochet and Jean Tirole, "Two-sided Markets: A Progress Report," *Rand Journal of Economics* 37 (2006): 645-667.



In the bank payment card industry, users' valuations of access to a given payment network vary with the number of other end users. For example, when more people opt to use cards issued by a particular network, an incumbent cardholder benefits from the fact that the expanded number of cardholders raises the likelihood that retailers will choose to accept the network's card. Hence, payment services provided by card networks are subject to *network externalities*, or third-party spillover effects.

Hayashi suggests that a key reason that retailers accept payment cards is transactional benefits in the form of sales that otherwise would not take place.⁸ In an analysis of retailers' motivations in accepting credit card payments, Chakravorti and To discuss an additional incentive, which is that accepting credit cards permits merchants to close current sales with illiquid consumers rather than confronting uncertainty regarding future sales.⁹ Rochet and Tirole argue that merchants' recognition of actual and anticipated sales expansions arising from network benefits to customers also internalizes those benefits, which expose merchants as well to network externalities.¹⁰ Merchants accept a payment platform's cards as long as fees are less than the internalized benefits.

Network externalities impinge on the balancing act that card payment networks must perform in developing their price structures. As discussed by Rochet and Tirole, the Coase theorem indicates that parties should be able use contracts internalizing usage externalities to attain a social optimum if property rights are clearly defined and if there is neither transaction cost nor asymmetric information.¹¹ Transaction costs and constraints established by networks or policymakers that impede such welfare-improving contractual agreements, they suggest, accounts for the two-sidedness of the price structure of card payment networks.

Why do networks often charge membership fees as well as usage fees? It is well known from the literature on price discriminating two-part tariffs that including access fees can efficiently enable the recouping of fixed costs, which for issuers include marketing and advertising expenses and for acquirers include expenses in developing and maintaining payment-clearing systems. Membership fees such as those depicted in Figure 1 reflect the application of two-part tariffs within the overall pricing structure of card

⁸ Fumiko Hayashi, "A Puzzle of Card Payment Pricing: Why Are Merchants Still Accepting Card payments?" *Review of Network Economics* 5 (2006): 144-174.

⁹ Sujit Chakravorti and Ted To, "A Theory of Credit Cards," *International Journal of Industrial Organization* 25 (2007): 583-595.

¹⁰ Jean-Charles Rochet and Jean Tirole, "Externalities and Regulation in Card Payment Systems," *Review of Network Economics* 5 (2006): 1-25.

¹¹ Rochet and Tirole, "Two-sided Markets."

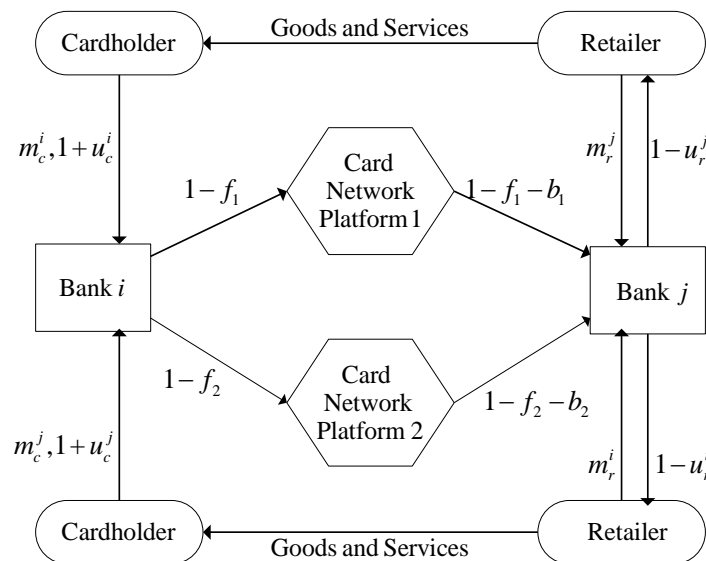


payment networks. In addition, Rochet and Tirole suggest that if cross-subsidization is part of an optimal pricing structure, portions of end-user surpluses can be captured via access prices such as membership fees.¹² Recently, Armstrong has also provided an analysis of two-sided markets in which flat membership charges can assist in balancing externalities confronting end users.¹³

Multiple Levels of Participation and Pricing Structures in Card Payment Networks

As explained by Schmalensee, fee structures of a card payment network depend on whether the network is *proprietary*—that is, operated as a profit-maximizing “closed network,” as in the case of the American Express network—or *cooperative*—that is, operated by an association of coordinating institutions, such as Visa’s “open network.”¹⁴ The basic two-sided structure depicted in Figure 1 is most applicable to a *unitary* proprietary network that engages in all issuing and acquiring activity on its own behalf. An alternative proprietary system is a *non-unitary* system in which the network contracts with other parties, such as selected institutions that do some of the issuing and/or acquiring. In contrast, in a cooperative card payment network, independent institutions establish a contractual arrangement for coordinating card payments via the network.

Figure 2: Competing Card Payment Networks



¹² Ibid.

¹³ Mark Armstrong, “Competition in Two-Sided Markets,” *Rand Journal of Economics* 37 (2006): 668-691.

¹⁴ Richard Schmalensee, “Payment Systems and Interchange Fees,” *Journal of Industrial Economics* 50 (2002): 103-122.



Figure 2 provides a stylized depiction of co-existing card-payment networks. The figure displays two institutions, Bank i and Bank j , which are members of two networks. For simplicity, the figure shows Bank i as an issuer with respect to both card network platforms and displays, and Bank j as a payment acquirer with respect to the two platforms. Some banks perform both roles in cooperative networks, although increasingly banks that operate as acquirers contract with third parties to process payments.

As in Figure 1, banks potentially charge membership and usage fees to cardholders and retailers. The two networks depicted in Figure 2 also may charge banks the fees b_1 and b_2 , which denote membership and/or usage fees and the fees f_1 and f_2 , which denote interbank fees—the interchange fees that the network collects from issuing banks on behalf of acquiring banks. Baxter¹⁵ provided the first formal analysis of interchange fees within a bank payment-clearing association, which he showed must take into account the two-sided nature of the payment transactions. In Baxter’s model with perfectly competitive net-issuing and net-acquiring banks, the association is indifferent about levels of the interchange fee—a result sometimes called “interchange-fee neutrality.”

Schmalensee and Rochet and Tirole point out that imperfect competition yields considerably more scope for variation in pricing structures across card payment networks.¹⁶ Studies of networks’ pricing structures often presume low search costs and little product differentiation on the acquiring side, which researchers often offer as support for assuming perfectly competitive behavior among acquirers, even though Evans and Schmalensee document increased concentration among acquirers. In contrast, such studies presume that search costs and the scope for product differentiation are greater on the issuing side, thereby resulting in some market power for issuers.¹⁷

Some studies assume that rival card payment networks process a volume of transactions that is either arbitrarily fixed or essentially predetermined via an exogenously specified distribution of consumer transaction choices.¹⁸ In the context of a model in which issuers have no incentive to operate as acquirers,

¹⁵ William Baxter, “Bank Interchanges of Transactional Paper: Legal and Economic Perspectives,” *Journal of Law and Economics* 26 (1983): 541-588.

¹⁶ Schmalensee, “Payment Systems”; Jean-Charles Rochet and Jean Tirole, “Cooperation among Competitors: Some Economics of Payment Card Associations,” *Rand Journal of Economics* 33 (2002): 549-570.

¹⁷ David Evans and Richard Schmalensee, “Paying with Plastic: The Digital Revolution in Buying and Borrowing,” (Cambridge, Massachusetts: MIT Press, 1999).

¹⁸ Rochet and Tirole, “Cooperation among Competitors”; Wilko Bolt and Alexander Tieman, “Social Welfare and Cost Recovery in Two-sided Markets,” *Review of Network Economics* 5 (2006): 103-117; Julian Wright, “Optimal Card Payment Systems,” *European Economic Review* 47 (2003): 587-612; Julian Wright, “The Determinants of Optimal Interchange Fees in Payment Systems,” *Journal of Industrial Economics* 52 (2004): 1-26.



Rochet and Tirole conclude that the interchange fee that maximizes overall social welfare is never higher than the profit-maximizing level in the absence of other pricing complications such as surcharges for purchases with cash instead of cards.¹⁹ They find that in the absence of such surcharges, the profit- and output-maximizing interchange fees coincide, a situation in which a socially inefficient *over*provision of payment cards can occur. When card surcharges are allowed, Rochet and Tirole's framework indicates that there can be an *under*provision of payment cards, resulting in ambiguous welfare effects.

In the context of a framework in which retailers' benefits from network participation are heterogeneous, Wright shows that the welfare-maximizing interchange fee must balance the differing externalities faced by cardholders and retailers, with the profit-maximizing aggregate usage level typically lying below the socially optimal level.²⁰ Wright separately finds that the welfare effects of imposing a "one-price policy" prohibiting surcharges for card purchases depends on the degree of market power possessed by retailers.²¹ Retailers possessing market power and the capability to impose surcharges will set prices to extract surplus from cardholders, thereby reducing revenues accruing to issuers. In the absence of retailer market power, however, a one-price policy results in a bifurcation of the retailing market into cash-only and card-only segments.

Schmalensee also studies the functioning of a single card payment platform in the context of a basic usage-pricing framework.²² He finds that a single unitary proprietary platform maximizing its own private value establishes a set of fees that generate the smallest transaction volume. Essentially, a unitary proprietary platform acts as a profit-maximizing monopoly and establishes a fee structure that restrains output of transaction services to the profit-maximizing level. A cooperative system, in contrast, aims to attain an output objective while covering operating costs; hence, it spreads payment volumes across the separate net-issuing and net-acquiring groups utilizing the payment network.

Within Schmalensee's single-network framework, which abstracts from banks' interactions with cardholders and presumes fixed weights in network decision-making by issuing or acquiring banks, the optimal interchange fee can be decomposed into two parts. One part depends on the difference between demand elasticities across cardholders and retailers. The other part depends on the difference in costs faced by issuers and acquirers. Schmalensee concludes that a cooperative platform typically will set a

¹⁹ Rochet and Tirole, "Cooperation among Competitors."

²⁰ Wright, "Optimal Card."

²¹ Wright, "Determinants."

²² Shmalensee, "Payment Systems."



pricing structure that yields, relative to unitary and non-unitary proprietary platforms, the highest volume of payment transactions. Double-marginalization (or double-monopoly-markup) problems confront both cooperative and non-unitary networks, with the latter producing a volume of payment transactions that is intermediate between the volumes forthcoming from cooperative and unitary proprietary networks.

Eisenmann et al. contemplate governance issues faced by two-sided platforms including providers responsible for serving as users' primary point of contact, as sponsors determining who may participate in a platform-intermediated network, *and* as developers of its technological workings—much as banks do in card payment networks such as Visa and MasterCard.²³ Eisenmann et al. argue that over time, two-sided networks tend to evolve into hybrids of proprietary and cooperative structures. In the long run, they conclude, such networks naturally gravitate toward centralized control over platform technology via closed sponsorship of a platform and shared responsibility for serving end users via open platform services. This could help explain why the American Express and Discover networks have recently sought to encourage participation by banks.

Bolt and Tieman study a model of a monopoly two-sided platform such as that depicted in Figure 1, and thus ignore interchange fees. In addition, they abstract from membership fees.²⁴ In this simplified setting, they show that under the socially optimal pricing structure, the monopoly platform fails to cover its operating costs. As a consequence, they conclude, network externalities faced by the card payment platform create a second-best pricing problem analogous to that faced by a natural monopoly, in which first-best, allocatively efficient pricing generates negative economic profits. A possible consequence, Bolt and Tieman suggest, is a card payment network pricing structure that includes cross-subsidization schemes, higher interchange fees, and no-surcharge rules preventing retailer participants from giving discounts to customers who pay with cash.

When platform providers compete for end users, as in Figure 2, the range of pricing-structure complexities broadens considerably. Nearly all work analyzing such competitive, two-sided-network settings has appeared very recently. In these settings, end users can “multihome”: Cardholders can participate in more than one network by utilizing more than one payment card, or retailers can accept cards from more than one network.

²³ Thomas Eisenmann, Geoffrey Parker, and Marshall Van Alstyne, “Opening Platforms: How, When, and Why?” in *Platforms, Markets, and Innovation*, ed. Anabelle Bawer (Cheltenham, UK: Edward Elgar, 2008).

²⁴ Bolt and Tieman, “Social Welfare.”



Rochet and Tirole briefly touch on the impacts of cardholder multihoming in competing payment systems.²⁵ They contend that if one network seeks to undercut another by setting a slightly lower interchange fee, then retailers have an incentive to respond by accepting the former system's card and rejecting the card of the other system. The result is a potential reduction in welfare compared with the monopoly case.

In a more general framework, Rochet and Tirole examine a model that extends the setting in Figure 1 to the case of more than one platform competing for the same end users, in which "buyers" (cardholders in the payment-network context) utilize the services of differentiated platforms that either are profit-maximizing firms or non-profit associations.²⁶ Rochet and Tirole include buyers and sellers of different types, with a network platform competing to obtain "marquee buyers"—cardholders that generate a particularly high surplus to retailers—from another platform while taking into account "captive buyers" that remain loyal to their initial platform choice. Other things being equal, the presence of marquee buyers naturally raises the seller price, but the presence of capital buyers skews the pricing structure to the benefit of sellers. Naturally, an increase in predisposition to multihoming on the part of buyers also is beneficial to sellers. These results could be sensitive, however, to Rochet and Tirole's assumption of no fixed membership fees, which Armstrong suggests could complicate joint determination of pricing structures in competing two-sided markets.²⁷ Rochet and Tirole find that under some circumstances interchange fees could be lower and total market surplus greater in a monopoly setting than when two networks compete.

Cabral suggests, however, that Rochet and Tirole's ambiguous welfare results regarding monopoly versus competition among two-sided networks does not necessarily rule out the existence of other more-than-offsetting welfare losses that might arise from the existence of a monopoly network.²⁸ A recent analysis by Chakravorti and Roson studies a variety of interactions between payment networks, including duopoly and cartel behavior with or without symmetry of the networks' payment instruments (such as credit cards offered by one platform and debit cards on another).²⁹ Within their framework, which includes some simplifying assumptions such as monopolistic retailers, Chakravorti and Roson find that competition

²⁵ Rochet and Tirole, "Cooperation among Competitors."

²⁶ Jean-Charles Rochet and Jean Tirole, "Platform Competition in Two-sided Markets," *Journal of the European Economic Association* 1 (2003): 990-1029.

²⁷ Armstrong, "Competition in Two-Sided Markets."

²⁸ Luis Cabral, "Market Power and Efficiency in Card Payment Systems: A Comment," *Review of Network Economics* 5 (2005): 15-25.

²⁹ Sujit Chakravorti and Roberto Roson, "Platform Competition in Two-sided Markets: The Case of Payment Networks," *Review of Network Economics* 5 (2006): 118-140.



unambiguously raises the welfare of end users, although not generally uniformly, even in the case of symmetric competition.

Guthrie and Wright seek to classify the wide array of equilibrium pricing structures that can emerge from competition between two card identical payment platforms facing potentially multihoming cardholders and retailers.³⁰ Guthrie and Wright find that either both networks set the same structure of fees with at least one side of end users (cardholders or retailers) multihoming, or only one network survives and attracts the exclusive participation of all end users. Guthrie and Wright conclude that in a setting with homogeneous retailers, competition between networks cannot boost fees charged to retailers. They argue that if retailers are more realistically heterogeneous, platform competition can yield higher fees for retailers, thus leading to higher interchange fees. The fee structure can be further biased against retailers if consumers choose a preferred card to hold and thereby benefit from platform competition that focuses on attracting cardholders rather than retailers. In contrast, if retailers know that cardholders always hold multiple cards, retailers can “steer” cardholders to their preferred network, resulting in platform competition focusing on attracting retailers, to the retailers’ benefit.

Ambrus and Argenziano analyze a model of two-sided network competition.³¹ They allow for heterogeneous consumer valuations of external benefits, which results in equilibrium outcomes that include multiple networks. Equilibrium network configurations entail differentiation of network products and asymmetric pricing structures across networks. In the context of payment card networks, these conclusions suggest that sufficient heterogeneities across cardholder and merchant end users could yield a stable outcome with card networks offering differentiated payment services—as observed at present.

Regulatory and Antitrust Issues in Card Payment Networks

Is the common European practice of regulating interchange fees and no-surcharge rules appropriate? Should card-payment-network rules be subjected to antitrust scrutiny? There has been some initial work evaluating these regulatory and antitrust issues. Schmalensee, in his analysis of a single card network platform, finds that interchange fees under either cooperative or non-unitary proprietary systems are driven primarily by different weights within the network of net-issuing versus net-acquiring banks.³² He

³⁰ Graeme Guthrie and Julian Wright, “Competing Payment Schemes,” *Journal of Industrial Economics* 55 (2007): 37-67.

³¹ Attila Ambrus and Rossella Argenziano, “Asymmetric Networks in Two-sided Markets,” *American Economic Journal: Microeconomics* 1 (2009): 17-52.

³² Schmalensee, “Payment Systems.”



also finds that the proprietary, private-value maximizing interchange fee may be higher or lower than the output-maximizing fee of cooperative network platform providers. Furthermore, his analysis suggests that a policy action reducing the interchange fee potentially could depress network transactions output. Policy intrusions into the fee-setting process, Schmalensee concludes, could place cooperative networks at a competitive disadvantage with respect to proprietary networks, leading institutions to abandon higher-output cooperative networks in favor of lower-output proprietary networks. Thus, there is no clear economic argument in favor of antitrust policy that interferes in the setting of interchange fees.

Regulators often argue that fees should be purely cost-based. As Rochet and Tirole point out, however, applying this logic would yield the conclusion that governmental fee regulations should be applied to the television industry, newspapers, videogame platform providers, and other two-sided industries.³³ On one hand, Rochet and Tirole question whether retailers actually experience harm as a result of interchange fees and suggest that some perceived harm may result from the fact that heterogeneous retailers internalize customer benefits at different levels.³⁴ On the other hand, Rochet and Wang analyze a three-stage game with four parties—acquirers, issuers, consumers, and merchants—interacting within a single card network.³⁵ They examine the role of intensity of issuer competition and conclude that consolidation among issuers could drive up interchange fees to the benefit of issuers and the detriment of merchants and consumers—a theoretical implication that they find receives support from U.S. data. Their model suggests that regulating the interchange fee to a lower level generates a revised pricing structure that, on net, engenders entry of new issuers and boosts the welfare of merchants and consumers. Not addressed by their single-platform framework is whether this result would hold true for competing networks.

Schwartz and Vincent³⁶ utilize a model with variable transaction volumes to examine the implications of no-surcharge rules, such as those already imposed in some nations and that some observers argue should be adopted in the United States.³⁷ Schwartz and Vincent suggest that such rules imbalance the fee structure

³³ Jean-Charles Rochet and Jean Tirole, “Competition Policy in Two-sided Markets, with a Special Emphasis on Payment Cards,” in *Handbook of Antitrust Economics*, ed. Paulo Buccirossi (Cambridge, Massachusetts: MIT Press, 2008), 543-582

³⁴ Rochet and Tirole, “Externalities and Regulation.”

³⁵ Jean-Charles Rochet and Zhu Wang, “Industry Concentration and Credit Card Pricing Puzzles,” unpublished manuscript, Toulouse University and Federal Reserve Bank of Kansas City, 2009.

³⁶ Marius Schwartz and Daniel Vincent, “The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network,” *Review of Network Economics* 5 (2006): 72-102.

³⁷ See, e.g., Nicholas Economides, “Competition Policy Issues in the Consumer Payments Industry,” in *Moving Money: The Future of Consumer Payments*, eds. Robert Litan and Martin Neil Baily (Washington, DC: Brookings Institution Press, 2009), 113-126.



between end users, resulting in harm to retailers.³⁸ Networks respond by reducing cardholder fees and granting rebates to card users if possible, with cardholder welfare potentially declining if rebates are not feasible. Network profits increase, and overall welfare increases only if there are a sufficiently large number of cash users.

Rochet and Tirole examine effects of an honor-all-cards rule.³⁹ In their benchmark model based on their earlier study,⁴⁰ they find that such a rule not only benefits the multi-card network platform providers that impose it but also allows providers to optimally rebalance externalities across end users. The network then processes larger volumes of both credit and debit payments, resulting in greater social welfare, which implies that the 2003 U.S. court judgment (ruling the application of an honor-all-cards rule to multiple types of cards to be illegal) could be welfare-reducing. Generalizing their framework to allow for heterogeneous retailers, differentiated platforms, and varying substitutability between credit and debit transactions yields the same rebalancing effect on the multi-card network's pricing structure. Welfare implications become ambiguous in this more general setting.

Emch and Thompson contemplate the application of the Department of Justice merger guidelines to card payment networks.⁴¹ Emch and Thompson show that it is possible that monopolization of a two-sided market considerably raises the prices charged to cardholders but not retailers, or *vice versa*. They derive price markup formulas for fees charged to end users and propose applying the guidelines' standard test for a "small but significant and non-transitory increase in price" that might result from a proposed merger to the sum of the prices the network charges to both groups of end users. White notes, however, that the pricing relationships derived by Emch and Thompson hinge their assumption of fixed proportions.⁴² He also concludes that their results likely depend on the simple two-sided framework they utilize, which more readily apply to proprietary networks than to open card associations.

³⁸ Schwartz and Vincent, "No Surcharge Rule."

³⁹ Jean-Charles Rochet and Jean Tirole, "Tying in Two-sided Markets and the Honor All Cards Rule," working paper, Toulouse University, February 9, 2006.

⁴⁰ Rochet and Tirole, "Cooperation among Competitors."

⁴¹ Eric Emch and T. Scott Thompson, "Market Definition and Market Power in Payment Card Networks," *Review of Network Economics* 5 (2006): 45-60. For a detailed discussion of these guidelines, see David VanHoose, "Policy Implications of Endogenous Sunk Fixed Costs in Banking: Has U.S. Antitrust Policy Been on the Wrong Track?" Networks Financial Institute policy brief PB-2008-06, December 2008.

⁴² Lawrence White, "Market Definition and Market Power in Payment Card Networks: Some Comments and Considerations," *Review of Network Economics* 5 (2006): 61-71.



Sun and Tse (2007) seek to examine the impacts of multihoming by applying a differential game analysis to both monopolistic and competing payment networks.⁴³ They conclude that when end users engage in multihoming, networks can co-exist. The steady-state market shares of competing networks can, however, diverge considerably depending on participants' multihoming propensities. Sun and Tse also argue that network distributors, such as card issuers, play a crucial role in maintaining and expanding a network. Sun and Tse's analysis provides some measure of support for the Department of Justice's actions charging Visa and MasterCard with antitrust violations for their rules forbidding banks from becoming issuers of American Express and Discover cards.

Conclusion

Whereas monetary-theoretic frameworks of analysis emphasize how the nature of payments can complicate a monetary equilibrium, the industrial organization approach takes as given the monetary system in which payment networks are imbedded. Furthermore, in contrast to monetary-theoretic models of payment systems, the modern industrial organization approach focuses on the two-sided nature of bank card payment network platforms. This approach highlights the importance of two-sided pricing structures constructed with an aim to account for network externalities that differentially affect cardholders and retailers.

So far, this industrial organization approach has yielded several conclusions and policy implications. First, privately and socially optimal pricing structures of two-sided card payment networks entail a complicated balancing of incentives faced by end users confronting network externalities.

Second, platform providers typically will charge cardholders and retailers unbalanced fees. Optimal interchange fees must reflect this asymmetrical pricing structure, so regulation of interchange fees will not necessarily improve overall social welfare, nor will restrictions on payment networks' rules governing surcharges. Fee and rule regulations could well be harmful, although regulations may be more likely to boost the well-being of merchants and consumers when issuers possess market power.

Third, some recent work on decomposing the pricing structure of card payment networks suggests that it may be possible to develop methods of applying current antitrust policy guidelines to card payment networks. Nevertheless, this conclusion must be regarded as tentative given that this initial work is predicated on a number of simplifying assumptions.

⁴³ Mingchun Sun and Edison Tse, "When Does the Winner Take All in Two-sided Markets?" *Review of Network Economics* 6 (2007): 16-40.



Fourth, payment-platform rules requiring retailers to honor debit cards if they accept the network's credit cards are not necessarily anti-competitive. Such rules may be consistent with an aim to effectively balance interests of all end users within a coherent pricing structure. Nevertheless, more work must be done to establish conditions under which honor-all-cards rules are welfare improving rather than welfare reducing.

Finally, rules restricting participation of banks in competing networks may be anti-competitive. Efforts to steer banks away from a competing platform reduce its ability to establish a critical mass of distributors. Rules restricting bank participation in competing networks reduce the likelihood of a multihoming outcome involving multiple rival card payment networks, thereby potentially reducing social welfare.