A Chip Off the Old Block or a New Direction for Payment Card Security?

Chips, PINs, and the Law and Economics of Payment Card Fraud

James C. Cooper and
Todd J. Zywicki

MERCATUS WORKING PAPER

All studies in the Mercatus Working Paper series have followed a rigorous process of academic evaluation, including (except where otherwise noted) at least one double-blind peer review. Working Papers present an author's provisional findings, which, upon further consideration and revision, are likely to be republished in an academic journal. The opinions expressed in Mercatus Working Papers are the authors' and do not represent official positions of the Mercatus Center or George Mason University.

MERCATUS CENTER
George Mason University

3434 Washington Blvd., 4th Floor, Arlington, Virginia 22201
www.mercatus.org

James C. Cooper and Todd J. Zywicki. "A Chip Off the Old Block or a New Direction for Payment Card Security? Chips, PINs, and the Law and Economics of Payment Card Fraud." Mercatus Working Paper, Mercatus Center at George Mason University, Arlington, VA, 2017.

Abstract

The issue of consumer payments and data security has reached a high level of public and regulatory interest as a result of a number of recent high-profile data breaches that compromised consumer payment cards. In addition, the ecosystem of consumer payment security has changed dramatically in recent years as a result of the introduction and rapid spread of contactless payment technologies. In response to growing concerns about payment fraud, payment card networks in the United States have moved toward the rapid replacement of traditional magnetic-stripe payment card technology to new EMV (Europay, Mastercard, and Visa) computer chip–based technology. Notably, however, US card issuers and networks have chosen not to adopt the personal identification number (PIN) method of customer verification that has been standard in the United Kingdom and much of Europe for the past decade or so but instead have chosen signature verification as the preferred method. This article conducts an economic analysis of the regulation of consumer payment cards and payment card fraud. We examine the marginal benefits and costs from heightened levels of payment card security. We examine the dynamic evolution of payment card anti-fraud technology over time and suggest that there is little evidence of market failure in the provision of payment security by card networks and issuers and little reason to believe that mandating one exclusive, decades-old, static verification technology (namely, chip and PIN) would be likely to improve overall consumer welfare and economic efficiency today. We conclude that rather than blindly adopting the particular verification technology that Europe put into place many years ago, US regulators should be alert to the evolving and contemporary nature of consumer payments and the fluid nature of threats to data privacy and thus should not freeze or hamper the adaptability of the payment system.

JEL codes: D18, E50, K20, K23

Keywords: credit cards, cardholder verification, chip technology, consumer payment, CVM, EMV, fraud, law & economics, payment card security, PIN, joint care, data security

Author Affiliation and Contact Information

James C. Cooper
Director, Program on Economics and Privacy
Associate Professor of Law, Antonin Scalia Law School, George Mason University

Todd J. Zywicki
Executive Director, Law and Economics Center, George Mason University
Professor of Law, Antonin Scalia Law School, George Mason University

Financial support for this research was provided by the Mercatus Center.

This paper can be accessed at https://www.mercatus.org/publications/payment-card-security-chips-PINs

A Chip Off the Old Block or a New Direction for Payment Card Security?

Chips, PINs, and the Law and Economics of Payment Card Fraud

James Cooper and Todd Zywicki

I. Introduction

In October 2015, the consumer payment system in the United States underwent a dramatic change. Merchants began to use machines that could accept payment cards with a computer chip, replacing the traditional magnetic-stripe card. The new chips, known as *EMV chips* (for Europay, Mastercard, and Visa), are expected to dramatically reduce point-of-sale (POS) fraud for credit cards and debit cards. Today, consumers increasingly are being required to "dip" their cards (insert the cards in the machine and wait for authentication) rather than to swipe the cards' magnetic stripe.

The transition to EMV is expensive. According to one estimate, chip-enabled cards cost approximately $2 each to manufacture, compared with "pennies" for magnetic-stripe cards.[1] Large card issuers may have tens of millions of cards outstanding at any given time because many consumers have multiple bank-issued credit cards, in addition to debit cards and certain store credit cards. Thus, issuing new cards alone is likely to end up costing issuers at least tens of millions of dollars. To avoid liability for fraud, merchants are also required to buy new payment terminals. It is estimated that a new EMV sales terminal costs roughly $500 to $1,000, a nontrivial cost for a very small business.[2] For a larger business with more than one checkout

[1] *See* Olga Kharif & Blanca Vázquez Toness, *Target Breach Spurs Retail Rush to Accept Tougher Credit Cards*, BLOOMBERG.COM (Apr. 12, 2014), http://www.bloomberg.com/news/articles/2014-04-10/target-breach-spurs-retail-rush-to-accept-tougher-credit-cards.

[2] *Id.* Other estimates say the range is as wide as $100 to $1,500 per terminal. *See How Much Will Chip/PIN Cost to Implement?*, BLUEPAY BLOG (Feb. 23, 2015), https://www.bluepay.com/blog/how-much-will-chippin-cost-implement/.

register, the investment in new equipment could add up to several thousand dollars—and potentially millions of dollars for the largest chains.

Understanding the constantly evolving framework of payment security requires analysis on multiple levels. In this article, we model the payment card security ecosystem as a joint-care problem, in which merchants, networks, and financial institutions can take precautions to reduce fraudulent transactions. Unlike the tort model of bilateral care, in which legal rules (e.g., strict liability or negligence) coordinate the actions of injurers and victims, participation in the payment card network is contingent on fulfilling certain contractual requirements. Thus, to a first approximation, we assume that the network will act as the hypothetical social planner would, setting the optimal mix of POS security and network security to maximize network value and enforcing that mix with contracts. We then use this framework to understand the current landscape of the payment card industry—in particular, the transition in the United States to the EMV standard and the current debate over requiring PINs in addition to chips as a means of authentication.

We show that because there is a tradeoff between the friction in the payment system (i.e., speed, convenience, and reliability) on one hand and security on the other, the payment card system should not strive to attain zero fraud. Such an approach would be cost prohibitive in practice, dramatically reducing the value and usefulness of payment cards generally and making consumers worse off. Further, we demonstrate that the optimal allocation of the cost and responsibility for payment security between network and POS measures will vary across societies and over time, which implies that no particular security technology (such as Chip and PIN) is likely to be universally efficient across different economies or even within the same economy over time. We identify as a particularly important factor the cost and reliability of telecommunications over time in various countries. Accordingly, the late adoption of the EMV standard by the United States is best understood as an efficient response to the costs and benefits of POS precautions, not as a market failure. Further, we find that using a liability shift rather than a government (or private) mandate to move to EMV is efficient when there is heterogeneity in the benefits from adopting certain security measures. Thus, by adopting a rule that allows firms to opt in to a security standard only if doing so reduces total fraud losses and costs of care, the United States would harness private information and maximize network value.

Finally, we assess recent efforts by merchant interests to require the adoption of one particular payment security technology—Chip and PIN—through either legislative mandates or litigation. For example, Kroger, Home Depot, and a class involving small merchants have each lodged antitrust suits against the major payment card networks for the networks' refusal to allow chip-based debit transactions to be routed through PIN networks.[3] When examined in an optimal care framework, the PIN requirement does not appear to be cost justified, as it does little to prevent counterfeit fraud—the primary type of payment card fraud—and is likely to cause nontrivial increases in the marginal costs of payment card transactions. As such, our analysis suggests that these lobbying efforts for government intervention by merchants likely are not a response to a market failure but rather a reflection of political economy considerations—in particular, an effort by merchants to try to steer consumers from their traditional preference for signature debit to greater use of PIN debit, which provides a financial benefit to larger merchants.

The remainder of the article is as follows. In part II, we lay out the basic economic tradeoff between security and payment friction. Part III introduces our model of bilateral precautions and derives some comparative static results that help explain why certain jurisdictions may rely on relatively higher levels of POS or network security. Part IV examines

[3] *See* Home Depot, Inc. v. Visa, Inc. (N.D. Ga. June 13, 2016); Kroger Co. v. Visa Inc. (S.D. Ohio June 27, 2016); B&R Supermarket, Inc. v. Visa, Inc. (N.D. Ca. Mar. 8, 2016).

recent US and European Union (EU) experiences with EMV through the lens of our model, and part V examines the political economy of the ongoing debate over whether PINs should be adopted as a cardholder verification method (CVM). The final section summarizes the article and offers some conclusions.

II. Payment Friction and Payment Security: The Economic Tradeoff

Assessing the optimal set of rules and institutions governing the payment card system is extremely complex.[4] The global payment card system is one of the most complex and efficient financial institutions in the history of the world: a 24-hour, secure, globally interconnected, instantaneous network of consumers, card networks, issuers, and merchants that reaches to the farthest corners of the world. Merchants gain access to near-instantaneous payments without the risk, delay, and cost associated with checks and cash. Consumers gain the flexibility and safety of not having to carry cash, thereby avoiding the risk of theft or loss as well as the cost and inconvenience of acquiring cash from an automated teller machine (ATM) or bank teller. In addition, in higher interest rate conditions, the process enables consumers to keep their funds in interest-bearing accounts instead of carrying depreciating cash in their wallets. Governments gain from the widespread use of electronic payments by not having to print cash, and they also benefit from the reduction of crime and tax evasion as consumers transition away from cash payments. Worldwide, billions of payment card transactions occur every day, with an astonishing degree of accuracy, speed, and security. In addition, for many consumers in the United States, access to this system has been virtually free, as for several decades most consumers have been able to acquire credit cards with no annual fee and no interest charge if the

[4] *See* Todd J. Zywicki, *The Economics of Payment Card Interchange Fees and the Limits of Regulation* (George Mason Law & Economics Research Paper No. 10-26, 2010), http://papers.ssrn.com/sol3/papers.cfm?abstract_id =1624002.

cardholder pays his or her bill in full every month. Indeed, once cash-back and other rewards are considered, some consumers may actually be paying a negative price for the ownership and use of their cards.[5]

When cardholders interact with the payment card system, they experience a near-seamless and simple experience. A consumer swipes or inserts a card, and within seconds the transaction is approved and the consumer is on his or her way. But like the proverbial tip of the iceberg, the simplicity of the consumer experience obscures the massively complicated system that lies beneath. In particular, behind this simple consumer interface rests a series of tradeoffs that crucially determine the efficiency of the payment card system.

From an economic perspective, at the most fundamental and overarching level the efficiency of the payment card system rests on a tradeoff between the speed and flexibility of the system (often called the *friction* of using the system) on one hand and the security of payment card use on the other.

On one hand, consumers, merchants, card issuers, and card networks seek a payment experience that is as frictionless as possible—that is, the fastest possible speed and convenience of payments. This minimization of friction has many elements, but they all rest on the basic observation that no one goes to Macy's, Starbucks, or Amazon.com to partake of the payment experience. The payment part of a transaction is the prototype of what economists refer to as *transaction costs*—namely, the necessary costs of accomplishing the parties' central goal, which is to buy and sell goods and services.

Payment friction takes several basic forms. First is the speed of payments (how quickly they can be authenticated) and the final decision whether to approve or decline a transaction.

[5] This phenomenon, in which consumers pay a subsidized, zero, or even negative price, is common in two-sided markets such as payment cards, newspapers, Internet search engines, and the like. *See id.*

Second, friction increases when there are higher levels of incorrect declinations of legitimate transactions (for example, when consumers incorrectly enters their PIN numbers or the card network incorrectly rejects a transaction as fraudulent). A third form of payment friction is the direct cost to the consumer and merchant—for example, the cost to consumers of transacting business (such as the costs of carrying a card or replacing a lost or damaged card) and the cost to the merchant of maintaining payment-processing equipment. The merchant's cost includes not only the direct costs of acquiring and maintaining certain equipment and dealing with repairs to broken equipment, but also the costs associated with the location of terminals in stores and the payment experience of consumers and merchants as part of a transaction.

In this section, we first examine the types of security risks attendant to payment card use. We then consider the frictions introduced by some security measures.

A. Security Risks

Payment card fraud broadly can be defined as any improper charge to an account made without the cardholder's awareness and consent. The channels through which payment card fraud occur vary. Ultimately, though, they all involve an unauthorized user having access to sufficient account information to pose as an authorized user. Such information may include the credit card number, the expiration date, and the customer verification number on the back of a card. Fraudsters can get this information through a variety of channels. First, a card may be lost or stolen. Second, credit card information may be compromised without loss or theft of the physical card. This form of access can occur through physical interaction (e.g., a waiter or clerk writing down a credit card number) or through more technologically sophisticated means. For example, "skimming" occurs when a thief places a small device at an ATM or a merchant's card reader that collects the information on cards' magnetic strips. The thief later returns to retrieve the device.

Similarly, large databases of credit card information held by merchants increasingly have become the target of identity thieves, as was the case in the widely publicized breaches at Michaels, Home Depot, and (probably most prominently) Target. Each of those breaches came about as a result of inadequate security precautions by the retailers. With respect to Michaels, for example, the attack was remarkably low tech: it has been reported that the criminals physically replaced devices at cashier checkout lanes at 80 Michaels locations in 19 states.[6] The terminals were infected with malware that collected the card numbers and expiration dates of approximately 2.6 million cards over an eight-month period before the breach was detected.[7]

The Target breach, by contrast, was much more elaborate. Hackers tapped into the computer network of one of Target's heating, ventilation, and air conditioning (HVAC) vendors, stealing the vendor's credentials and installing malware on its system.[8] The hackers then used the vendor's credentials to gain access to an area of Target's computer network, where they installed malware on Target's system. Because Target lacked adequate firewalls and other security devices between vendor operations and the consumer sections of Target's system that held consumer data, the hackers were able to install malware initially only on Target's vendor system but then were able to use that point of entry to obtain consumer data. The hackers then sent the malware through Target's computer system to cashier stations in all domestic Target stores. Soon, credit card numbers started flowing out of the registers and into several servers in the

[6] *See* Tracy Kitten, *Michaels Breach: What We've Learned*, BANKINFOSECURITY: THE FRAUD BLOG (Aug. 4, 2015), http://www.bankinfosecurity.com/blogs/-p-1910.

[7] *See* Mathew J. Schwartz, *Michaels Data Breach Response: 7 Facts*, DARKREADING.COM (Apr. 22, 2014), http://www.darkreading.com/attacks-breaches/michaels-data-breach-response-7-facts/d/d-id/1204630.

[8] *See* Michael Riley, Benjamin Elgin, Dune Lawrence & Carol Matlack, *Missed Alarms and 40 Million Stolen Credit Card Numbers: How Target Blew It*, BLOOMBERG.COM (Mar. 17, 2014), http://www.bloomberg.com/news/articles/2014-03-13/target-missed-warnings-in-epic-hack-of-credit-card-data.

United States before they were apparently routed to Moscow. The outflow of card numbers continued for several days despite alarms within Target's system that a breach had occurred. In the end, the Target data breach resulted in the theft of approximately 40 million credit card numbers. The breach affected all 1,797 of Target's US stores.

Home Depot's breach was similar to Target's, in that its network was compromised by gaining access through a third-party vendor's stolen credentials.[9] Once the hackers gained access to the system, they were able to install "unique, custom-built malware" on self-checkout systems in the United States and Canada. They used that malware to steal information on approximately 56 million credit and debit cards and to steal email addresses for another 53 million consumers.[10] Home Depot did not confirm that the breach had occurred until a week after credit card data linked to its customers went up for sale on the black-market website Rescator.cc.[11] The breach continued for months and occurred despite the fact that Home Depot had software that could have encrypted consumer data and thereby reduced the risk of the theft.[12] In addition, just months before the major breach, the company had suffered two minor breaches yet still chose not to deploy software that could have prevented the consumer data from being stolen. It has also been reported that Home Depot was using outdated antivirus software in its stores.

When data are skimmed or breached, there is likely to be a longer lag time between theft and discovery than for stolen cards. It will almost always take the card owner less time to discover that a physical card is missing than to discover fraudulent charges, which may not be

[9] *See* Jai Vijayan, *New Details of Home Depot's Attack Reminiscent of Target's Breach*, DARKREADING.COM (Nov. 7, 2104), http://www.darkreading.com/attacks-breaches/new-details-of-home-depot-attack-reminiscent-of-targets-breach/d/d-id/1317323.

[10] *See* Michael Winter, *Home Depot Hackers Used Vendor Log-On to Steal Data, E-mails*, USATODAY.COM (Nov. 7, 2014), http://www.usatoday.com/story/money/business/2014/11/06/home-depot-hackers-stolen-data/18613167/.

[11] *See* Benjamin Elgin, Michael Riley & Dune Lawrence, *Home Depot Hacked After Months of Security Warnings*, BLOOMBERG.COM (Sept. 18, 2014), http://www.bloomberg.com/news/articles/2014-09-18/home-depot-hacked-after-months-of-security-warnings.

[12] *Id.*

evident until the bill is viewed. As discussed later in this article, fraud detection techniques are helping to close this gap.

Thieves use stolen payment card information in various ways. Criminals commit so-called card-not-present (CNP) fraud, which occurs when card information is used to purchase goods or services online, over the phone, or in other circumstances in which the seller doesn't need access to the physical card. When only the card information is compromised, such as through skimming or data breaches, the information is often sold in bulk on the so-called dark web.[13]

Stolen numbers also can be encoded onto counterfeit cards with easily obtainable technology. According to a report by the Aite Group, counterfeit fraud is the largest category of credit card fraud, accounting for 45 percent of losses, followed by CNP fraud, which accounts for 38 percent of losses.[14] Lost and stolen cards account for only 9 percent of losses. According to the Federal Reserve's analysis of fraud losses on debit cards, in 2015 lost and stolen fraud losses accounted for about 1.0–1.5 basis points as a share of transaction value for PIN and signature debit.[15] By contrast, "the majority of fraud losses for single-message debit transactions [i.e, PIN debit] was attributed to counterfeit fraud." Overall, fraud losses from counterfeit cards were 3.1 basis points per transaction value for PIN debit and 5.4 basis points for signature debit.

[13] The marketplace for this type of information is saturated such that a single account sells for no more than $10. *See* JFC, *The Life of a Stolen Credit Card*, DEEPDOTWEB.COM (June 27, 2016), https://www.deepdotweb.com/2016/06/27/life-stolen-credit-card/. One survey of dark-web credit card sites in 2015 claimed that more than 1.4 million US credit cards were available for sale at that time. Although an in-depth analysis found that claim to be exaggerated, it did verify that at least 50,000 card numbers were available for sale at that time. *See* Joseph Cox, *We've Never Seen a Stolen Credit Card Market as Slick as This*, MOTHERBOARD BLOG (Nov. 9, 2015), http://motherboard.vice.com/read/weve-never-seen-a-stolen-credit-card-market-as-slick-as-this.

[14] THAD PETERSON & JULIE CONROY, *Chip Cards in the United States: The PIN, PINless, Debit, Credit Conundrum* at 12, Fig. 2, AITE GROUP, LLC (July 2016).

[15] BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM, 2015 INTERCHANGE FEE REVENUE, COVERED ISSUER COSTS, AND COVERED ISSUER AND MERCHANT FRAUD LOSSES RELATED TO DEBIT CARD TRANSACTIONS 20 (Nov. 30, 2016).

For signature debit, by contrast, 56 percent of fraud losses were from card-not-present fraud, amounting to roughly 7 basis points per transaction value.[16]

B. Types of Security

Payment card networks use a variety of means to reduce fraud, some occurring at the merchant POS and others through the network. Broadly, POS methods focus on verifying the identity of the card presenter, whereas network security focuses on whether the card itself is valid or whether the transaction suggests fraud.

1. Point of sale. When a card is presented for payment at a merchant terminal, the merchant can use several non–mutually exclusive techniques to verify that the user of the card is authorized. For example, the merchant can check the ID of the person presenting the card or examine the signature on the receipt or device capture to see if it matches the signature on the back of the card. In some cases (primarily for debit cards), the presenter also may have to enter a PIN or other identifying information, such as a zip code. Biometric identifiers, such as fingerprints, retina scans, and facial recognition, increasingly are being used as identifiers as well.

An additional dimension of CP security involves securing the data transmitted from the card to the terminal at the time of the transaction. As previously noted, fraud is primarily from card information captured during transmission or stolen from databases. As will be discussed in more detail later, EMV is a POS security method that reduces the fraudsters' ability to complete

[16] Overall, fraud rates for signature debit are higher than for PIN debit. But this may be only partially or slightly attributable to PIN's being a more secure system than signature debit. Differential fraud rates between PIN and signature also reflect the reality that many higher-fraud transaction settings—such as online shopping—accept only signature debit or non-PIN credit cards. As a result, the higher rate of fraud for signature cards reflects that signature is accepted much more widely, including in contexts that have higher baseline fraud rates unrelated to the particular verification method.

a transaction with a counterfeit card by transmitting a transaction-specific number rather than a static account number. Mobile devices reduce the ability of thieves to capture account information by encrypting it during transmission.

2. Network. Security is also performed at the network (or issuer) level. For example, the issuing bank will deny a card that has been reported as lost or stolen, or if there is evidence that it has been compromised. Further, algorithms are used to determine whether a transaction is inconsistent with normal use (for example, because the card is being used in a different area or for a very large purchase).

C. Frictions from Security

Security is necessary to deter fraudsters, but it comes at a cost. Obviously, there are direct fixed costs to employing security, such as building (or upgrading) network infrastructures and purchasing EMV terminals. However, there are also marginal costs—precaution costs per transaction—that have important implications for determining the optimal level of security. Broadly, these costs are associated with frictions introduced into the payment system, and they fall into two bins: (a) reductions in speed and convenience, and (b) an increase in false positives.

1. Payment speed and convenience. Consumers and merchants seek a speedy, convenient, and low-cost method of making payments. Speed is of particular importance for many merchants as they seek to maximize the throughput of their customer experience and minimize the store's labor costs of dealing with the transaction of processing payments. Consider a simple intuitive example: assume that it takes 10 seconds longer for a merchant to process a payment using a slower payment device, A, (say, a check) than a faster one, B, (a credit card). Even at this small marginal difference in time, if there are six people in a checkout line, this delay will increase the checkout time for the sixth person in line by one minute, and so on. From the perspective of the merchant, however, the effect is even larger: for a large merchant who conducts hundreds or thousands of transactions a day, these small increments could add up to hundreds or thousands of dollars of additional labor costs each year as employees simply wait for transactions to clear. The increments may also require a retailer to maintain additional registers and may lead to some abandoned sales.

Over time, the coevolution of information technology, telecommunications infrastructure, and consumer and merchant demand for faster payment times has dramatically reduced the friction associated with the consumer payments system. In the United States, for example, average transaction time to make a payment of less than $25 at a quick-service restaurant is only 4–5 seconds for a payment card, which is substantially faster than even cash (8–10 seconds).[17] For payments at discount stores or grocery stores, a recent estimate was that the average time was approximately 17 seconds for a cash payment, 17–19 seconds for a debit card transaction, and 57 seconds for a check.[18] This reduction in processing time has contributed to the increased ubiquity in the acceptance of payment cards. For example, in 2003 McDonald's made the decision to accept payment cards. As a result of that decision, the value of McDonald's stock increased by 2.7 percent.[19] Other quick-service restaurants have also benefited from accepting payment cards, both because of reduced costs of handling cash and also because of faster

[17] Anne Layne-Farrar, Are Debit Cards Really More Costly for Merchants? Assessing Retailers' Costs and Benefits of Payment Instrument Acceptance 51 (Charles River Associates working paper, Sept. 9, 2011), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1924925.

[18] *Id.*

[19] *Id* at 8.

throughput and so-called ticket lift from increased size in the average sale.[20] These effects can be significant; according to one estimate, each 10-second increment that can be cut from the average drive-through time is worth approximately $1,000 in revenue for a typical restaurant.[21]

This desire to reduce the friction of payments and the costs to the consumer and merchant explains many important retail trends of recent years. For example, consider the development of self-checkout lines at many stores (e.g., grocery stores, drugstores, and hardware stores) or transactions with unmanned kiosks at locations such as gas stations, train stations, and vending machines. In all these locations, the ubiquity of electronic payments has enabled some consumers to forgo an interaction with a sales clerk, thereby speeding the checkout process, enabling the merchant to reduce the number of employees assigned to the routine work of ringing up consumers, and freeing up employees for other, more important tasks. Paying at the pump at gas stations, for example, (a) saves the consumer the time and effort of walking to and from the cash register (usually twice in the case of a payment card transaction), (b) saves time and reduces lines at the checkout counter (especially during busy times), and (c) allows the station to reduce the number of employees.[22] Self-checkout at grocery stores also speeds up checkout time, permits reductions in employee staffing, and even takes up less space than traditional checkout lanes.[23]

The switch to EMV illustrates the tradeoff between security and friction. It was expected that as consumers and merchants became more familiar with EMV payments, average checkout times would not be much longer than when using traditional magnetic-stripe technology. Yet

[20] *Id.* at 14.

[21] Linda Punch & Jeffrey Green, *Fast Food Meets Fast Payment*, 15 CREDIT CARD MANAGEMENT, no.11, Jan. 2003, at 18.

[22] *See* DOUGLAS F. ALDRICH, MASTERING THE DIGITAL MARKETPLACE: PRACTICAL STRATEGIES FOR COMPETITIVENESS IN THE NEW ECONOMY 37–39 (1999) (describing time and cost savings from adoption of pay-at-the-pump technology at Mobil gas stations and subsequent improvements).

[23] *See* Nick Mann, *The Pros and Cons of Using Self-Checkouts*, BUSINESS BEE, http://www.businessbee.com /resources/profitability/the-pros-and-cons-of-using-self-checkouts/ (last visited Aug. 12, 2016).

according to an article in the *Wall Street Journal* in August 2016, it still took twice as long to pay with a chip card than with a swipe or mobile payment—on average, 13 seconds versus 6 seconds; over the span of a year, a consumer could spend 85 extra minutes standing in line to pay. [24] But note—that is just the extra time it takes for *one* person to pay. If there are, say, five people in line, the person at the end of the line could wait more than half a minute longer in line just because of the delay in payment times. According to one estimate, the average consumer will spend five and a half hours per year waiting for EMV transactions to go through, and businesses will experience 116 million hours of additional checkout time as a result of EMV.[25]

A similar economic tradeoff applies to analyzing the rapidly growing world of e-commerce and online shopping. Consider the decision of whether to store one's credit card number with Amazon.com, iTunes, or some other online merchant. The costs of such a decision are obvious: it is possible that the merchant's website might get hacked and one's payment card information might be compromised. On the other hand, the benefits of permitting Amazon.com to store your payment card information are sizable: access to Amazon's "1-Click Ordering" feature and the ability to make purchases without having to reenter one's payment card number for each transaction. Many consumers are willing to accept the slight risk of a possible compromise of their credit card number to capture the efficiency and convenience of storing one's credit card information online, as long as they feel that the merchant is credible and committed to security.

On the other hand, at the same time that these innovations have reduced payment friction and enabled additional efficiencies related to payments, they have also raised novel problems of

[24] *See* Joanna Stern, *Chip Card Nightmares? Help Is on the Way*, WALL ST. J. (Aug. 2, 2016), http://www.wsj.com/articles/chip-card-nightmares-help-is-on-the-way-1470163865.

[25] *See* Beth Braverman, *Consumers Spend 5 1/2 Hours a Year Waiting for Chip-Card Transactions*, BUS. INSIDER (Sept. 6, 2016), http://www.businessinsider.com/customers-spend-5-and-a-half-hours-a-year-waiting-for-chip-cards-2016-9.

fraud. For example, when a credit card is stolen, often the first place the thief tries to use it is at a self-service gas station or subway ticket kiosk. Why? Because the impersonal nature of the interface enables the thief to verify whether the card is still active without risking a confrontation with a sales clerk if the transaction is declined. Thus, although these sorts of innovations present huge benefits to consumers and merchants in terms of reducing payment friction, this reduction in friction for legitimate transactions also can come at a cost of increasing the potential for illegitimate transactions.

This tradeoff between the costs and benefits of reducing payment friction at the risk of some higher incidence of fraud is also reflected in the decision by payment card networks to adopt policies that permit many merchants to waive CVM requirements for transactions below a certain size to speed the checkout process. Granted, elimination of this authentication requirement would be expected to increase the incidence of payment card fraud overall. However, apparently the payment networks and merchants who choose to forgo CVM have implicitly decided that the costs of increased fraud with respect to some small-dollar transactions are outweighed by the benefits of faster throughput at the register and the small size of the transactions. Moreover, as detailed later in this article, the elimination of the signature requirement for some small-dollar transactions does not mean an absence of any security protocols whatsoever—instead, merchants are just eliminating the *marginal* cost and *marginal* benefit of requiring a signature authentication. At the same time, the network retains its full apparatus of authorization and authentication protocols. Moreover, the application of these waivers and exceptions to ordinary procedures is highly calibrated and is tied to specific merchants, industries, geographic locations, and the like, all of which affect the tradeoffs between reducing payment friction at the margin and the marginal impact on payment security.

Finally, consumers seem to understand the tradeoff as well. Although consumers express support for EMV as a means to increase data security, they also have expressed frustration with it, mainly from increased friction in transactions and longer checkout times. According to one analysis conducted soon after the liability shift occurred, the time needed to pay using a chip card was on average 7–10 seconds, as compared with 2–3 seconds using a magnetic-stripe card.[26] The survey also found that 20 percent of users said that EMV payments "take too long." In addition, after having had experience with chip cards, "nearly four times as many survey respondents [were] worried about speedy processing times over chip card security or availability of EMV terminals." Consumers also have had to deal with extended hassles and delay from removing the card from the reader prematurely and having the transaction canceled, resulting in further delay and frustration.[27] A survey by the Mercator Advisory Group in November 2015 found that 28 percent of EMV cardholders were bothered or confused by the EMV card or tried to avoid shopping at stores that required them to use it.[28] A September 2016 survey by Square found even higher levels of discontent, reporting that 91 percent of debit card users and 87 percent of credit card users are "frustrated" with EMV cards, primarily because the cards increase checkout time.

2. Accurate authentication of payment card transactions. Minimizing payment friction also includes accurately processing and approving transactions. In particular, as a first approximation this means that the payment system must approve all legitimate transactions the first time they

[26] *See Harbortouch Survey: 20 Percent of Users Say EMV Payments Take Too Long*, GREENSHEET.COM (Nov. 16, 2015), http://www.greensheet.com/newswire.php?flag=display_story&id=40303.

[27] *EMV Rollout Coming with a Few Expected Glitches, and One Unexpected Recommendation*, CUTODAY.INFO (Oct. 19, 2015), http://www.cutoday.info/THE-feature/EMV-Rollout-Coming-With-A-Few-Expected-Glitches-And-One-Unexpected-Recommendation.

[28] *Wary About Credit Card Security, Consumers Want EMV Cards but Find Using Them Frustrating*, STREETINSIDER.COM (Nov. 24, 2015), http://www.streetinsider.com/Press+Releases/Wary+about+Credit+Card+Security,+Consumers+Want+EMV+Cards+but+Find+Using+Them+Frustrating/11103873.html.

are attempted. If a legitimate transaction is incorrectly declined and must be attempted a second time, that increases the costs and friction of the system. Again, at an intuitive level, consider a card transaction that is improperly declined, thereby leading the consumer to have to pull another card from his or her wallet and reattempt the transaction. Having to repeat the transaction increases the transaction costs of making the payment and the attendant costs in terms of inconvenience to customers as well as labor and other costs to the merchant.

The costs of payment friction, especially for inaccurate declinations of legitimate transactions, can be especially high in some contexts. For the average consumer, for example, the cost of a declination of an attempted transaction using a debit card is higher than that of a declination using a credit card. This is because although many consumers carry more than one credit card (and thus can simply pull an alternative card from their wallet), few consumers carry more than one debit card. In addition, many households (especially younger and lower-income households) do not have a credit card and therefore rely almost entirely on using their debit card to conduct electronic transactions. In that situation, as a result, an improper transaction declination can have high costs in terms of wasted time and energy for both the consumer and the merchant.

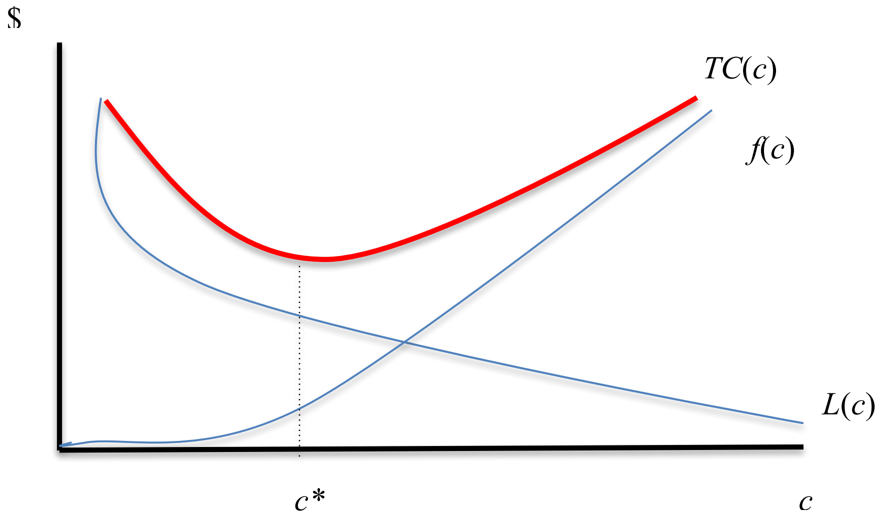
Approval of payment card transactions thus presents a classic tradeoff between type I and type II errors—that is, false positives and false negatives. One can easily see that when a thief uses a stolen card to make an improper payment, there is a cost to the payments system that must be allocated in some fashion among the consumer, merchant, issuer, and card network. Yet it should be recognized that there is also a cost when a legitimate payment is declined. Most trivially, there is a cost in terms of the time needed to try the transaction again using the same card or a different card. But in some instances there may be a larger cost—the cost of not being able to conduct the transaction at all if the consumer has no other payment device available. For example, if the card is being used to buy baby formula, or medicine, or gasoline to get to work in the morning, there can be a substantial cost to a consumer if that payment is incorrectly declined as fraudulent.

Thus, a substantial part of the cost of the payment system is the development of the complex network of computer systems and complicated algorithms that payment networks, card issuers, and merchant acquirers use to more accurately distinguish legitimate from illegitimate transactions—that is, to minimize the joint costs of each transaction in terms of reducing the costs of false positives (incorrect declinations) and false negatives (approving improper transactions). As should be readily apparent, the more vigilant the card networks are about trying to prevent unauthorized transactions, the more likely they will also be to inadvertently block valid transactions.

This tradeoff is illustrated in figure 1. The payment system can take additional care, *c*, to avoid fraud, which is measured on the horizontal axis. As it takes more care, fraud losses, *L*(*c*), decrease. At the same time, however, as efforts to avoid care increase, so do costs from increased frictions, *f*(*c*).

The goal of the system, therefore, is not to minimize fraud. Instead, it is to minimize the sum of fraud and friction costs, *TC*(*c*), which in the case of figure 1 occurs at *c**. In the next section, we explore more deeply how a payment system allocates care between the network and merchants, which is at the heard of the movement to the EMV standard and the chip-versus-PIN debate.

Figure 1. Optimal Level of Fraud Precaution



III. Understanding Optimal Network Security: A Model of Joint Care

A primary goal of the payment card network is to maximize its value to consumers. Every time a consumer uses a payment card, there is a risk that the information will be stolen and used to make illicit purchases. Although consumers generally are not directly responsible for fraudulent charges, those charges are a cost to the system that ultimately gets passed on in a competitive market.[29] Therefore, a payment card network has an incentive to minimize the total costs from fraud—both the direct costs of illicit transactions and the costs of preventing fraud. Broadly,

[29] According to the Federal Reserve, for example, in 2015 consumers absorbed only 3 percent of the losses from debit card fraud, whereas issuers absorbed 58 percent and merchants 39 percent (mainly from CNP fraud). *See* FEDERAL RESERVE, *supra* note 15, at 22. Although consumers are not directly responsible for fraud losses, they can experience costs in terms of inconvenience and indirect loss (such as changing credit card numbers and more closely monitoring against unauthorized charges), which suggests that their effective cost from card fraud is nonzero. In addition, consumers pay indirectly in higher card fees or higher prices for goods and services from fraud losses.

one can imagine that networks have two leverage points to combat fraud: at the POS or through the network.[30]

This problem can be couched in a stylized joint-care model, in which the payment care industry would like to avoid losses from fraudulent transactions, L, which can be reduced by action at both the point of sale, P, and through the network, N.[31] These actions have marginal costs ϕ and θ respectively. A consumer's marginal willingness to pay for a payment card transaction is u, and his or her net value from using the payment card network is $u - L(P, N) - \phi P - \theta N$, under the assumption that the marginal cost of network $(L(P, N) + \phi P + \theta N)$ is the price paid by the consumer in the form of monetary and time costs.[32] Rearranging the conditions for optimality (shown in the appendix) gives rise to the following expression, which provides insight into the substitution between network and POS security measures:

$$-\frac{L_P}{L_N} = \frac{\phi}{\theta}.$$

This equality states that the ratio of the marginal reduction in fraud losses from POS and network care is equal to the ratio of each method's marginal cost. This relationship implies that as the relative marginal cost of POS verification rises, payment card networks will choose greater reliance on network authentication, and vice versa. To see this, suppose that the marginal cost of POS precaution rises. To maintain optimality, L_P must also rise. Because of diminishing marginal returns to increased precaution, a reduction in the use of POS services will lead to an

[30] For purposes of simplification, we largely ignore the potential role of consumers in preventing fraud. The basic model of joint care that we develop could be generalized to create a three-way system of allocation of fraud prevention and insurance costs among issuers/networks, merchants, and consumers. But the underlying analysis is largely identical; therefore, little of use is gained through that additional complexity. In addition, many of the actions that consumers can take are largely captured in the costs incurred by merchants and overall friction costs.

[31] P and N are a decomposition of c (care) shown in figure 1.

[32] Nonsecurity marginal costs are normalized to zero.

increase in L_P, while substitution to network care simultaneously will reduce L_N until the equality of the ratios is reestablished.

The graphical solution to the joint-care problem can be represented in two dimensions in figure 2.[33] $\bar{L}(P,N)$ is an iso-loss curve, representing the minimum achievable loss.[34] Point A, along a 45° line, represents an equal use of POS and network care. The slopes of the tangent lines represent the relative costs of network and POS authentication, with the steeper curve representing relatively more expensive network costs and the flatter curve representing relatively cheaper network costs. The optimal mix of network and POS service occurs at the tangency point, which is where the slope of the iso-loss line—which represents the technical ability to substitute POS for network authentication—equals the ratio of network and POS costs or, more technically, where $-\frac{L_P}{L_N} = \frac{\phi}{\theta}$.

As one can readily see, the solutions to the cost minimization problem are intuitive: if POS and network costs are equal, then the solution is at point A, where the tangent bisects the 45° line, meaning that both security measures are used equally. Systems with relatively higher network costs, point C, rely more on POS authentication, and vice versa for systems with relatively more expensive POS costs (point B). Importantly, not only do higher network costs lead to less reliance on network methods of authentication, but they also lead to higher overall losses.

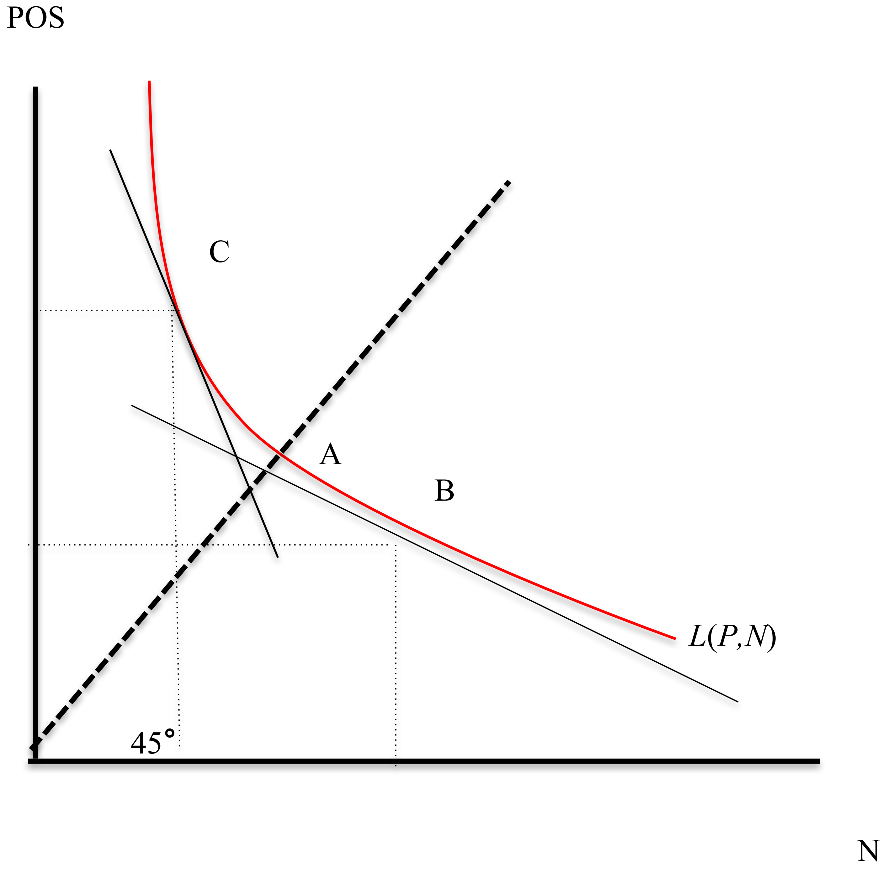
[33] The three-dimensional solution is shown in figure A1 in the appendix.

[34] Slopes of isoloss curve come from total differentiation of the loss function holding loss constant:

$$L(P,N)_P dP + L(P,N)_N dN = 0$$
$$\frac{d(P)}{d(N)} = -\frac{L_P}{L_N}.$$

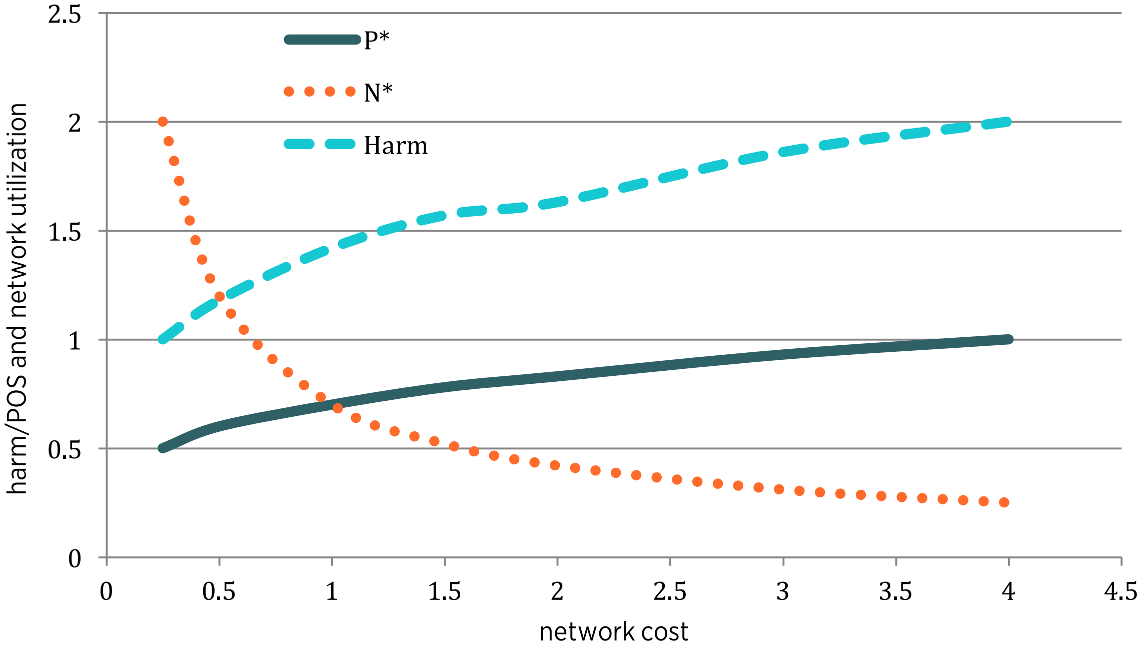
In reality, as the relative costs of POS and network change, the level of total loss at the new P* and N* will rise unless P and N are perfect substitutes. In this way, figure 1 captures the pure substitution effect of changes in relative costs.

Figure 2. Optimal Mix of POS and Network Care



The solution shown in figure 2 captures only the substitution effect from relative price changes. There is also a second-order impact on harm: because the cost of POS services does not fall to compensate for higher network costs, the payments system is devoting fewer resources overall to fraud prevention, which leads to higher losses. To illustrate the full effect of more expensive network care, we conducted a simulation by parameterizing the joint-care model to derive equilibrium levels of care and fraud. POS costs are held constant at 1, with network costs ranging from 0.25 to 3. Results are shown in figure 3.

Figure 3. Simulation Results



When the POS and network costs of care are equal, they are employed in equal amounts to combat fraud. When the marginal cost of network authentication is one-fourth the cost of POS, network use is four times that of POS use. As network costs are increased, not only is there a marked substitution to POS authentication, but total harm also rises because substitution is imperfect. Although the payment system can shift from network to POS authentication, the marginal POS precautions are not as productive as the lost network precautions because of diminishing marginal returns.

In the next section, we examine the model's predictive capability against the EU and US experiences with payment card security, focusing primarily on the recent transition to EMV. In brief, we find that this model of payment security is consistent with the way in which payment security has developed in the United States and around the world. The model explains the peculiar status of the United States as a historical outlier with respect to payment security, and particularly the divergent paths in payments taken by Europe (which adopted EMV Chip and PIN technology in the early 2000s) and the United States (which remained standardized on magnetic-stripe technology until 2016 and even then adopted a new standard using Chip and Signature technology instead of Chip and PIN). The analysis suggests that the addition of PIN verification in the United States may not be consistent with network value optimization.

IV. Network Cost Minimization and EMV Adoption

In October 2015, Visa and Mastercard imposed a liability shift with respect to payment card use and acceptance. This regime was adopted by private action, not by government intervention. This liability shift had several notable characteristics. First, it provided that if merchants installed payment card processing terminals that were EMV capable, they would be protected from liability for fraud from counterfeit cards. Second, the networks provided that the required means of authentication would be a signature, not a PIN. Third, the networks provided that for many smaller transactions, usually under $50 depending on the merchant and type of transaction, no additional authentication would be required.

Public reception to the adoption of EMV technology has been mixed. Many small merchants have criticized the new standards, objecting that they will force the merchants to upgrade to new payment terminals or face the potential for increased liability. Many of the merchants have balked at these costs, complaining especially that what they will spend in precautions far exceeds the benefits from reduced fraud for their small shops.[35] As noted earlier,

[35] It has been reported that one small retailer in Michigan assesses a 3.75 percent surcharge on debit and credit cards, which the owner contends is to defray the cost of adopting EMV machines. The owner commented, “To convert my old system into a chip reader, you’re talking thousands in software.” *See* Anthony Sabella, *Chip Card Access Puts*

these costs can be especially trying for merchants who have just recently begun accepting payment cards by using payment dongles such as Square. In addition, the higher cost, larger size, and higher complexity of new chip-enabled terminals will be a barrier to many small businesses that currently do not accept payment cards but that otherwise might do so. According to a survey by Wells Fargo's small business group in July 2015, only 29 percent of merchants had planned to upgrade to EMV-enabled card processors.[36] Twenty-one percent stated that they never intended to adopt EMV-compatible terminals, and another 16 percent did not know whether they would do so. Of those who stated that they did not intend to upgrade before October 2015, 25 percent stated that they never planned to upgrade and would simply stop accepting payment cards at the POS.[37] Forty-six percent stated that they did not want to pay for the EMV terminal, and 41 percent stated that they were not concerned about the liability shift. Very small retailers (such as a corner sandwich shop) may be especially unlikely to upgrade because they are unlikely to be the target of counterfeit or lost/stolen fraud (because of the small gains available to criminals from such activites).

In this section, we use the model presented in part III to help explain the timing and method of adoption of the EMV liability shift standard in the United States, as well as to analyze the debate over the use of PINs to authenticate transactions. First, we examine the historical role that telecommunications costs have played in determining different mixes of POS and network security measures. Next, we examine the underlying forces that have led the United States to follow the European Union in adopting the EMV standard, which puts a greater reliance on POS

Service Fees on Credit Card Use at Local Businesses, ABC12.COM (Sep. 7, 2016), http://www.abc12.com/home/headlines/Chip-card-access-puts-service-fees-on-credit-card-use-at-local-businesses-392677441.html.

[36] Wells Fargo, *Q3 2015 Small Business Index Survey Results*, WELLSFARGOWORKS.COM (Aug. 6, 2015), question 17, https://newsroom.wf.com/press-release/community-banking-and-small-business/wells-fargo-survey-many-small-businesses-not.

[37] *Id.* at question 17_1.

security. We also use the joint-precautions model to suggest an explanation for the fact that the networks did not mandate EMV but rather have created incentives for merchants and issuers to adopt EMV technology by shifting the liability for fraudulent transactions. Finally, we examine the case for requiring PINs as an additional method of POS authentication, and we find reasons to suggest that this requirement may not hold up to a benefit-cost analysis.

A. The Role of Telecommunications Costs

In recent years, the payment security debate has focused in large part on the extent of the security devices built into cards (chips) and the verification method required by consumers and merchants (PINs, signatures, some other form of verification, or none at all). Ironically, however, the friction and cost of these forms of POS security have not been the determining factor as to whether they are required. Instead, the degree of security and verification required by consumers and merchants has been an indirect manifestation of a more fundamental factor—the speed and cost of a country's telecommunications technology.

Authentication of payment card transactions can take place in two distinct frameworks: online and offline. In an online system, payment card authorization takes place online and in real time—essentially, once a consumer swipes or dips a card, the information on the stripe or chip is transmitted from the payment card terminal to the issuing bank.[38] The issuer applies a set of highly complex computer algorithms and accesses information about the consumer's unique account—for instance, by validating the cryptogram (in the case of a chip card) and by determining (a) whether the transaction would exceed the consumer's authorized credit limit, (b) whether the card has been reported lost or stolen, or (c) whether the transaction appears odd in relation to normal consumer habits—to either authorize or reject the payment. Over time, of

[38] The transmission goes through several stages, such as the acquirer and the card network, to reach the issuer.

course, telecommunications have become speedier, more reliable, and less expensive, enabling authorization to be made even faster. The transaction is approved or rejected within seconds.

In an offline system, by contrast, final authorization from the issuer does not take place in real time. Instead, the transaction is made and held by the merchant—perhaps for days—and is later "batched" and sent for approval. In this sense, an offline system resembles the credit card imprinters of earlier eras, when the merchant made an imprint of the consumer's credit card and then submitted it to the financial institution for clearing.

Traditionally, the determining factor for whether a country's consumer payment card system standardized on online or offline authorization was the cost and reliability of the country's telecommunications system.[39] In particular, countries where telecommunications technology has been fast, reliable, and inexpensive have been late adopters of higher-cost cards and increased POS verification methods by consumers and merchants. In countries where telecommunications technology has been slow, unreliable, and expensive, consumers and merchants traditionally have had greater responsibility and greater cost for preventing fraud. In other words, countries where low-cost telecommunications has enabled card issuers and networks to prevent fraud at a comparatively lower cost have been able to avoid the higher costs imposed on merchants and consumers of requiring cards with more secure technologies built in (such as chips) and the increased payment friction that accompanies such methods.

This technologically motivated decision explains the variation among countries in their migration toward new authentication systems, particularly EMV systems. In particular, the absence of real-time online verification in the United Kingdom and Europe made it essential to

[39] *See* Julie Conroy, *EMV: Lessons Learned and the U.S. Outlook*, AITE GROUP, LLC (June 2014). Ronald Mann has also noted that countries with relatively more expensive telecommunications costs should be predicted to have higher fraud rates *ceteris paribus*, although he does not discuss the joint-care model we discuss here or the use of alternative authorization technologies. *See* Ronald J. Mann, *Credit Cards and Debit Cards in the United States and Japan*, 55 VANDERBILT L. REV. 1055 (2002).

strengthen verification procedures at the time of purchase. Because the payment might not be authorized or rejected for hours or even days, merchants and financial institutions needed some alternative system to reduce fraud at the time of purchase. As a result, they developed the concept of Chip and PIN as a substitute for real-time authorization. Note that at that time, online authorization (with magnetic-stripe cards) was the preferred authorization method because of the low cost and high convenience of real-time authorization. In offline-authorization countries, however, eventually it was thought that although Chip and PIN was more expensive and less convenient, the additional expense was justified in light of the difficulty of preventing fraud in other ways.

Consistent with the joint-precautions model's predictions, the burden on consumers and merchants for security at the POS, therefore, historically has been a negative function of the degree to which the networks and issuers themselves can engage in timely and accurate verification of payments. As will be discussed, the recent adoption of chip technology in the United States in its particular form (i.e., without required PIN) reflects the economic tradeoffs embedded in this underlying economic model.

B. Explaining Timing and Method of EMV Adoption in the United States

The model predicts that fraud losses will rise as the cost of security rises. Accordingly, we should expect jurisdictions with high telecommunications costs to have higher fraud rates than those in the United States and other jurisdictions with low telecommunications costs. The data tend to support this prediction.

For example, at the time that EMV was adopted in the United Kingdom, the fraud rate in that country was 14 basis points—almost three times higher than the fraud rate in the United

States at the time (just 5 basis points). As noted earlier, telecommunications costs in the United Kingdom were very high at that time; therefore, the country used offline authorization. In the short run, the UK adoption of EMV had the desired effect of reducing POS fraud. For example, losses from counterfeit fraud dropped from £129.7 million in 2004 (immediately before the country's liability shift) to £43.4 million in 2013. Losses from lost/stolen fraud fell from £114.4 million in 2004 to £58.9 million in 2013.

On the other hand, fraud rates in the United States have been relatively low because of the sophistication of data analysis by processing networks and the availability of real-time online transaction authentication. Between 2011 and 2013, however, US credit card fraud losses increased by 31 percent. This increase primarily was driven by counterfeit fraud, which increased from $1.652 billion to $2.41 billion in 2013. Ironically, another factor in the increase in US fraud was the introduction of EMV verification in Europe and other parts of the world, which pushed criminal activity involving counterfeit cards to the United States.[40]

The increase in fraud was not lost on consumers, who have expressed concern regarding security in the wake the high-profile data breaches.[41] According to one consumer survey, 90 percent of consumers were aware of the data breaches at major retailers and 93 percent were concerned about the security of their credit card information.[42] Another survey found that 77 percent of consumers were anxious about their financial information and social security numbers

[40] Conroy, *supra* note 39, at 28, Fig. 21. Lost/stolen fraud, by contrast, was less than half the size of counterfeit fraud in 2011 ($811 million), had increased to only $825 million in 2013, and was projected to rise to only $850 million in 2015. Unlike counterfeit fraud, lost/stolen fraud is not easily scalable by criminals.

[41] *See* Claire Greene & Joanna Stavins, *Did the Target Breach Change Consumer Assessments of Payment Card Security?* (Fed. Res. Bank of Boston Research Data Reports No. 16-1, Aug. 2016) (finding that consumers expressed less confidence in their data security after the Target breach).

[42] *See* David Braue, *Consumers More Concerned About Credit-Card Security Than Their Health*, CSO.COM, http://www.cso.com.au/article/558332/consumers-more-concerned-about-credit-card-security-than-their-health/ (last visited Aug. 20, 2016).

being stolen or compromised.[43] Industry surveys of consumers also found some significant support for the adoption of EMV cards.[44] The primary reason consumers stated for wanting EMV cards was the increased security that those cards provide.

This rapid increase in fraud, with its attendant consumer reaction, was a primary impetus for the US adoption of EMV. Again, this pattern is consistent with the model predicting that exogenous shocks to expected harm—such as increases in fraudsters' technological capabilities—would lead to improvements to security. Moreover, to the extent that the marginal product of POS precautions is likely to be larger than that for network precautions—for example, in preventing the interception of credit card data at the POS or preventing the ability to use counterfeit cards—the increase in precautions primarily will be along the POS dimension.[45]

An interesting facet of the US movement to EMV was that it was accomplished not only without any government involvement, but also without being privately mandated. Before EMV, the status quo provided that as long as merchants abided by contractually obligated security measures, issuers would be liable for counterfeit and lost/stolen fraud.[46] This rule was akin to a strict liability rule on issuers. In a bilateral care context, strict liability is known to create moral hazard on the part of the nonliable party. However, if POS measures were unlikely to contribute much to security or were too expensive to be cost justified, a strict liability rule would be

[43] *See* Press Release, Mastercard, MasterCard Survey Reveals Americans Anxious About Personal Security but Optimistic About New Ways to Pay (July 9, 2015), http://newsroom.mastercard.com/press-releases/mastercard-survey-reveals-americans-anxious-about-personal-security-but-optimistic-about-new-ways-to-pay/. According to the survey, 55 percent of respondents "would rather have naked pictures of themselves leaked online than have their financial information stolen."

[44] *See* Mastercard, *Consumer Enthusiasm and Desire for Chip Cards Growing* (May 2015), http://docplayer.net/12793239-Consumer-enthusiasm-and-desire-for-chip-cards-growing.html.

[45] This will occur as long as L_P is sufficiently larger than L_N, which would be the case if a system were using a large level of network security in relation to POS security.

[46] *In re* Sony Gaming Networks and Customer Data Sec. Breach Litig., 903 F. Supp. 2d 942, 951–52 (S.D. Cal. 2012). For example, merchants were required to adhere to the Payment Card Industry Security Standards to guard against data breaches. Merchants experiencing a breach because of suboptimal security would potentially be liable to the issuer for fraudulent charges. Indeed, this is the subject of lawsuits arising from the Target breach and from the massive Wyndham breach.

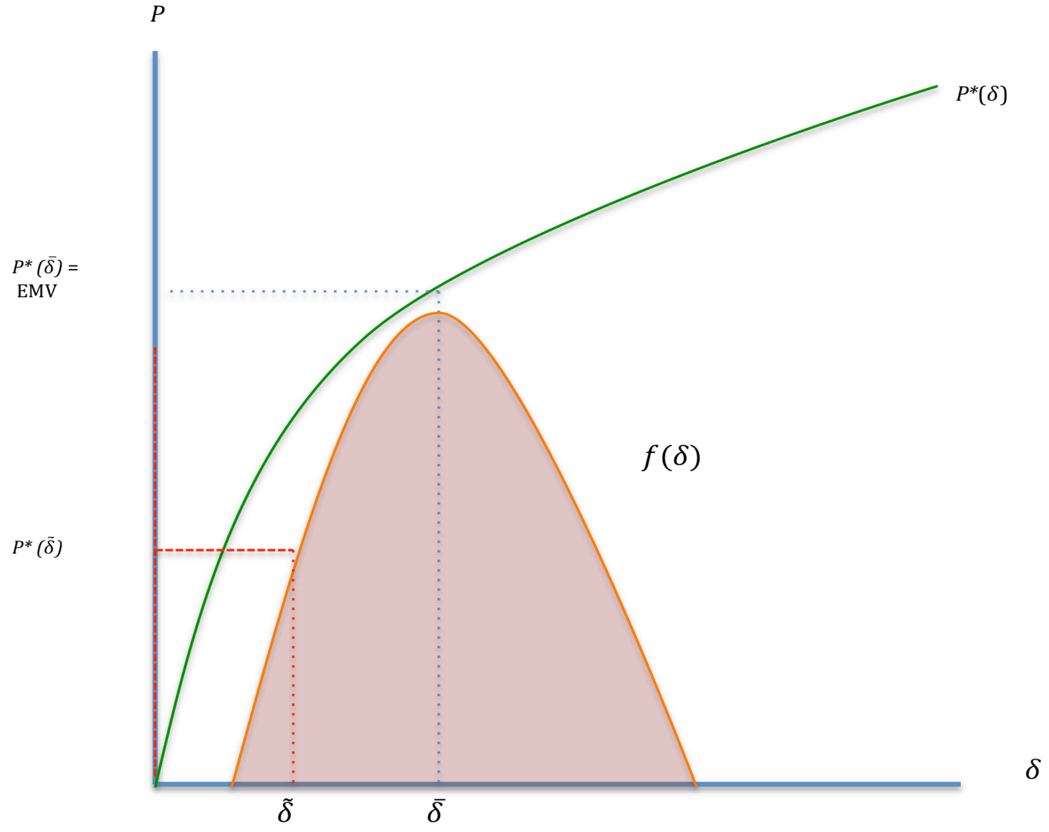
superior to others because it would economize on administration costs. Further, contractual obligations for network memberships could be used to mitigate moral hazard through direct regulation of behavior.

Rather than requiring merchants to adopt the EMV standard as a condition for network membership, the major networks moved to what can best be described as a rule of strict liability with a defense of contributory negligence: the issuers remain strictly liable for counterfeit fraud *unless* the merchant has failed to adopt the EMV standard.[47] In the standard joint-precaution tort model, it is well known that this rule will lead each party to adopt optimal care.[48] The victim will choose to take due care, because otherwise he or she will be stuck with the full costs of accidents, which are less than the cost of optimal care. Knowing that the victim will never be contributorily negligent, the injurer internalizes all the accident and avoidance costs and therefore has an incentive to take optimal care.

[47] For example, in the Visa network, merchants not adopting EMV are liable for (a) losses due to data stolen from chip cards at non-EMV compliant terminals, and (b) losses resulting from use of a counterfeit card at a non-EMV-compliant terminal. The Mastercard and American Express networks similarly place liability on the party who took the least precaution. For example, if a merchant is EMV compliant but a customer's bank has yet to issue EMV-compliant cards, losses from data stolen from this card would be the issuer's responsibility. *Visa Core Rules and Visa Product and Service Rules*, April 15, 2015, PSR-325.

[48] LOUIS KAPLOW & STEVEN SHAVELL, *Economic Analysis of Law*, *in* 3 HANDBOOK OF PUBLIC ECONOMICS 1666 (A.J Auerbach & M. Feldstein, eds., 2002); Andrew Daughety & Jennifer Reinganum, *Markets, Torts, and Social Inefficiency*, 37 RAND J. ECON. 300, 300–23 (2006).

Figure 4. Optimal POS Care as Function of Expected Harm (δ)



The liability shift solution to the EMV migration highlights the decentralized nature of the optimal care problem. Although our model couches this as a joint-care problem with a uniform optimal POS solution, the reality is that it may not be optimal for every merchant to adopt EMV when expected damages are heterogeneous. Suppose that the amount of fraud damages that a merchant faces is represented by a parameter δ, which is distributed $f(\delta)$. The optimal level of POS care for each merchant, $P_i^*(\delta)$, is shown in figure 4 to be a positive function of expected damages.[49] Even though a uniform EMV mandate (P*($\bar{\delta}$)) may be optimal if one rule has to be applied to the entire population,[50] those suffering harm away from the average ($\bar{\delta}$) are forced to take too much or too little care. Consider the small merchant who is unlikely to be a victim of counterfeit fraud located at $\tilde{\delta}$. This merchant will be better off if he or she is able to opt out of EMV as long as expected liability from fraud damages *without* EMV is less than the increase in precaution costs associated with adopting EMV, which is more likely to hold at lower levels of expected harm: $L\left(P^*(\tilde{\delta})\right) < \phi[P^*\left(\bar{\delta}\right) - P^*\left(\tilde{\delta}\right)]$.

In this manner, a uniform EMV mandate would force small merchants facing minimal risks from counterfeit fraud to engage in care that would not be cost justified. The liability shift rule, then, could be thought of as an efficient way to use the decentralized nature of the POS care decisions to harness private information about damages. Small merchants who view their risk of being targeted by fraudsters for data theft as small and who also view the potential losses from customers using counterfeit cards as small rationally may decide to forgo EMV adoption because the marginal benefits are less than the marginal costs of precaution. Importantly, this reticence to adopt EMV is optimal from a network point of view as well. If those merchants were forced to adopt EMV, the higher costs would be passed along to consumers without sufficient offsetting benefits in terms of reduced risk of payment card fraud.

By allowing self-selection, this approach has an added dynamic benefit. Today, the largest underpenetrated market in the United States for acceptance of payment cards is these very small businesses. It is estimated that some 20 million small businesses today that do not accept

[49] This can be seen from the individual merchant's loss maximization problem, in which the merchant selects P taking N as a given: max $q_i[u_i - L(P_i; N) - \phi P_i]$. Assuming that δ is a scaling factor for L, differentiating the first-order conditions with respect to δ yields: $\frac{\partial P_i^*}{\partial \delta} = \frac{-L_p}{L_{PP}} > 0$. Comparing this result with expression A9 in the appendix, which does not hold N constant and depends on the substitutability of network for POS care in light of changed damages, highlights the divergent incentives of network managers and individual merchants.

[50] KAPLOW & SHAVELL, *supra* note 49, at 1666–1784 (2006).

payment cards could convert a mobile phone or tablet into a card reader or cloud-based payment device using a payment dongle such as Square. Not only does the inability to accept payment cards increase payment friction for both consumers and these businesses, but it is also a primary source of tax evasion because cash transactions are largely untraceable. Thus, to the extent that certain elements of payment security increase the cost to particular merchants (such as small merchants) of accepting cards, that expense can deter the general spread of electronic payments in the economy. As analysts at J.P. Morgan observed, "In other words, mobile phone and tablet card readers could do to the physical world what PayPal did to the online space over 15 years ago, by [providing] casual merchants that previously couldn't afford to maintain a merchant account with a cost effective means of taking credit or debit cards."[51]

A potential concern about employing EMV through a liability shift could be moral hazard on the part of issuers: if issuers perceive that merchants are unlikely to adopt EMV, then issuers will no longer be liable for losses and hence will have suboptimal incentives to take precautions (e.g., invest in fewer network-based tools or solutions). There are at least two reasons, however, to believe that moral hazard will be muted. First, cards are issued to customers to be used at myriad merchants. As long as merchants who view EMV adoption as an uneconomical proposition represent a relatively small proportion of charges made by issuers' customers (which is likely to be the case), issuer incentives will remain essentially unchanged. Second, because the level of network care influences POS care, sufficiently low levels of issuer care may increase risks to a point that causes merchants to adopt EMV, which would shift liability back to issuers.

[51] J.P. MORGAN NORTH AMERICA EQUITY RESEARCH, *Payment Processing: Payments Market Share Handbook* 17 (6th ed. May 15, 2015).

V. Customer Verification and EVM: PIN versus Signature

The introduction of EMV into the United States has been contentious, in large part because of the costs of implementation (which, as discussed earlier, are substantial).[52] Another major debate accompanying the EMV transaction is the appropriate CVM. The current EMV standard in the United States calls for signature CVM, with no CVM for transactions below certain thresholds or where the risk of fraud is likely to be low. Some, however, have argued for the adoption of a PIN as the required CVM. For example, federal lawmakers have held hearings on the matter, and bills about requiring PINs have been introduced at the state level.[53] Furthermore, although many small businesses have complained about the cost of EMV, many big-box retailers and other special interests have argued that PINs should be adopted. In an intensive lobbying and litigation effort, Walmart, Kroger, and many other large retailers have joined forces to seek legislative, regulatory, or judicial mandates that *require* payment card networks to use PIN verification rather than signature verification.[54]

In this section, we examine the net worth of PIN verification from the point of view of maximizing network value, and we also explore the political economy of the PIN debate. Applying the foregoing model of the evolution of payment card security to the specifics of the Chip and PIN debate suggests that there is no evidence that the decision of card networks and issuers to provide a liability shift with respect to EMV adoption—but *not* to require PIN verification—reflects a market failure. Instead, as the foregoing analysis has suggested, the decision to incentivize EMV adoption but not PIN verification appears to be consistent with a desire to maximize the overall value of the system to all parties, taking into account the costs and benefits of greater security as well as the costs of alternative security precautions. Moreover, the

[52] See Kitten, *supra* note 5, and *supra* note 5**.**

[53] A4422 New York State Senate (2017–2018 Legis. Sess.).

[54] Wal-Mart Stores, Inc. v. Visa U.S.A. Inc. (N.Y. County 2016) (No. 652530/2016).

dynamic nature of evolving payment security protocols with respect to consumer payments suggests that government should take great caution before second-guessing these decisions.

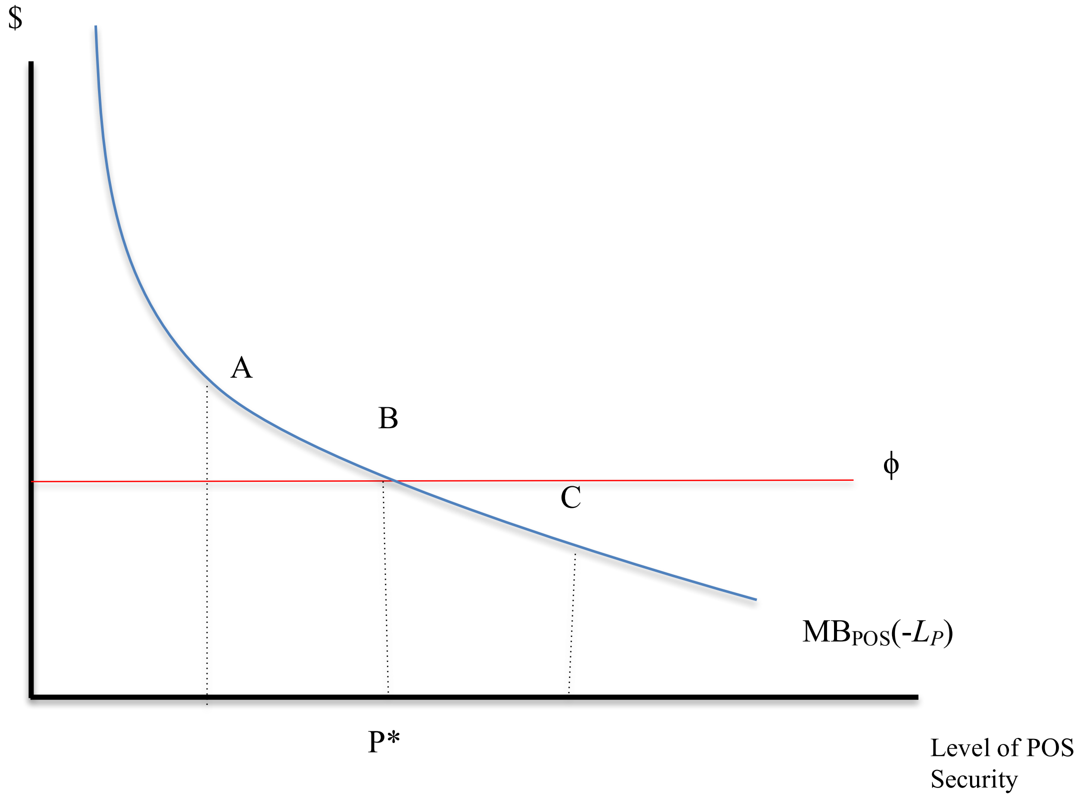
A. Does PIN Increase the Value of the Network?

As shown in part III, even if consumers are not financially responsible for losses from fraud, they end up paying for fraud and prevention costs indirectly through fees and transaction costs. Accordingly, consumers have an interest in the adoption of only those additional security measures that have a marginal value beyond their marginal cost.

Figure 5 puts the question of PIN adoption in the framework of the joint-care model. The point at which the marginal cost of POS care (ϕ) and the marginal benefit from increasing POS security measures ($-L_P$) meet defines the optimal level of POS precautions ($P*$). The question is whether the adoption of EMV alone is more like point A (where the marginal cost of care is less than the marginal benefit of additional precaution) or point B (the level of optimal care). If EMV alone gets us to point A, then adoption of PIN verification may move us closer to the optimal level because the additional friction introduced by PIN verification is less than the marginal benefit in terms of reduced expected fraud losses.

On the other hand, if EMV alone is closer to point B, the optimal level of care, PIN adoption will result in too much care. Although using PINs will provide additional protection against fraud, this marginal benefit will be too small in relation to its marginal cost to be beneficial to society, moving us toward point C. We next explore the available evidence, which in our view suggests that although PIN may provide some temporary relief from lost/stolen fraud, this marginal benefit is likely to be meager in relation to its substantial marginal costs.

Figure 5. Marginal Costs and Benefits from PIN



1. Marginal benefits. As previously discussed, the push to implement EMV in the United States was animated by the rising rate of fraud, particularly counterfeit fraud.[55] That rapid increase in counterfeit fraud explains the move to adopt EMV in the United States, notwithstanding its additional cost and payment friction. The introduction of EMV in the United Kingdom, for example, cut counterfeit fraud losses in that country to less than one-third of their prior rate—from £170 million in 2008 to £45 million in 2015.[56] It is expected that once EMV is implemented in the United States, counterfeit fraud should drop dramatically in that country as well.[57] Thus, implementing EMV alone addresses the largest source of preventable fraud. Moreover, adopting

[55] *See* discussion *supra* notes 41–43 and accompanying text.
[56] PETERSON & CONROY, *supra* note 14 at 7.
[57] Conroy, *supra* note 39, at 28.

EMV appears to be having the intended effect already—according to Mastercard, merchants who have adopted EMV technology have seen a 54 percent reduction in fraud.[58]

EMV alone, however, does little to address lost/stolen fraud. A valid (not counterfeit) card in the hands of an unauthorized user will work until it is reported lost or stolen, or until purchasing patterns result in the card being flagged as such. Although EMV as implemented in the United States allows signature verification as the preferred method, a signature can be easily faked and is rarely checked. Further, many transactions do not require a signature or other CVM. The addition of a PIN works primarily on this margin. PIN verification adds a layer of security against lost/stolen fraud because a lost or stolen card is worthless without the PIN. Indeed, the experience in the United Kingdom illustrates this observation: after the introduction of Chip and PIN, lost/stolen losses fell from £68.5 million in 2006 to £44.4 million by 2010. Although some of this decline may have been associated with the overall reduction in economic activity during the financial crisis, the use of PINs appears to have had an effect.

Despite its potential to ameliorate some fraud, the overall impact from the addition of PIN to EMV cards is likely to be small; Aite Group estimates that only about 2 percent to 2.5 percent of fraud would be prevented by adding the PIN verification method to EMV.[59] There are at least three factors behind this small marginal benefit.

First, criminals adapt. For example, the initial decrease in lost/stolen fraud after the introduction of PIN security in the UK was short-lived. Lost/stolen fraud began a dramatic reversal, reaching £74.1 million by 2015, higher than before Chip and PIN was introduced.[60] This reversal in lost/stolen fraud suggests that criminals sought new tactics as counterfeiting

[58] *See* Kim S. Nash, *MasterCard Seeks to Stop Online Fraud with Selfies, Fingerprints*, WALL ST. J.: CIO J. BLOG (Sept. 14, 2016), http://blogs.wsj.com/cio/2016/09/14/mastercard-seeks-to-stop-online-fraud-with-selfies-fingerprints/.

[59] PETERSON & CONROY, *supra* note 14, at 30.

[60] PETERSON & CONROY, *supra* note 14, at 14, Fig. 3.

became more difficult: they focused on new ways of capturing both the card and the consumer's PIN.[61] For example, thieves use such methods as false keypads that overlay the POS checkout and capture consumer PINs, installation of small cameras focused on a store's keypad, and even old-fashioned techniques such as looking over a consumer's shoulder as he or she enters a PIN. Similar techniques have been used to capture consumer PINs from ATM transactions.[62] Phishing scams also become more profitable if consumers can be tricked into providing their PINs. According to Financial Fraud Action UK, ATM attacks in the United Kingdom increased from 2,553 in the first four months of 2012 to 7,525 during a similar period in 2013.[63]

Second, consumers who have their PINs captured in addition to their card numbers can suffer much greater loss than those who merely have their magnetic stripe compromised. In particular, not only can a criminal who captures a consumer's PIN engage in fraudulent transactions, he or she can also go to an ATM and empty a consumer's bank account. According to data collected by the Federal Reserve, the average loss per fraudulent transaction is approximately twice as large for PIN debit fraud as for signature debit.[64] Moreover, many consumers reuse their PINs for multiple purposes to reduce the risk of forgetting them; thus, a consumer whose PIN is breached for one card may suffer other losses. So even though a PIN might provide a consumer with increased marginal protection from fraud—in this case only lost/stolen fraud because it is the EMV chip that prevents counterfeit fraud—this additional reduction in risk must be tempered by the cost of risking higher loss in the event of a breach or skimming of the consumer's PIN.

[61] *See* discussion *supra* notes 59–61 and accompanying text.

[62] Conroy, *supra* note 39, at 9–10; *see also* Brian Krebs, *Secret Service Warns of "Periscope" Skimmers*, KREBS ON SECURITY BLOG (Sept. 13, 2016), https://krebsonsecurity.com/2016/09/secret-service-warns-of-periscope-skimmers/.

[63] Jessica Winch et al., *Warning to Cash Machine Users*, THE TELEGRAPH (Jun. 2013), http://www.telegraph.co.uk/finance/personalfinance/bank-accounts/10103786/Warning-to-cash-machine-users.html.

[64] *See* FEDERAL RESERVE, *supra* note 15, at 21, Exhibit 16.

Third, in addition to any change in criminal behavior, the fact that lost/stolen fraud remains the smallest portion of payment card fraud—about 9 percent—further mitigates benefits from adding PIN verification to an EMV card. As noted, the largest component of fraud has been counterfeit cards, which EMV addresses. And the most rapidly growing component of fraud is CNP fraud, which does not require a PIN.[65] In the United Kingdom, between 2004 and 2008 CNP fraud increased from £151 million annually to £328 million annually.[66] British issuers and merchants responded to this skyrocketing fraud by increasing protections for CNP; that step led to a decline in CNP fraud to £221 million annually in 2011. By 2013, however, this trend had reversed itself, and CNP fraud had increased to £301 million. Another avenue of fraud in response to greater POS security is fraudulent application fraud, in which a criminal submits an application for a new card in the victim's name. In Australia, for example, from 2011 to 2012 fraudulent application fraud rose threefold as the adoption of EMV technology accelerated.[67]

2. Marginal costs. Although layering PIN verification onto an EMV card is likely to provide some additional protection against lost/stolen fraud, it also will add substantial new friction to the consumer payment system. First, and perhaps foremost, adding new terminals will increase implementation and certification time. Transaction time will increase, too, after those terminals are installed. For example, if a merchant sought not only to process EMV transactions but also to require a PIN, that would require still further security and other costs. For many small merchants, it is not uncommon to keep a small payment-processing terminal behind the sales counter and to physically swipe or dip the card for the consumer. If merchants were forced to accept chip and PIN

[65] As a report from the Federal Reserve notes, the low level of fraud for PIN debit is driven in part "by the fact that single-message transactions rarely take place online, where most card-not-present fraud originates." FEDERAL RESERVE, *supra* note 15, at 20.
[66] CONROY, *supra* note 39, at 10.
[67] CONROY, *supra* note 39, at 14.

verification, by contrast, the merchant would be required to (a) place the card terminal in a location that is easily accessible to consumers, and (b) take proper precautions so that consumers can shield the keypad when entering their PIN to prevent unauthorized surveillance of that process. Sit-down restaurants would need one or more portable payment terminals for consumers to use, which in turn would raise new security issues as well as the potential for damage to payment terminals. According to one survey, in June 2016 only 39 percent of "eating and drinking" establishments had PIN capability.[68] Furthermore, certain types of CVMs may be excessively inconvenient, cumbersome, or even infeasible in many transaction contexts, such as trying to enter a PIN when paying at a fast-food drive-through window or paying a toll on the highway. Along with adding time and inconvenience to the transaction, PIN verification also increases the likelihood of a "false rejection" that occurs when a legitimate user forgets his or her PIN.

The ambivalence about PIN verification is reflected in consumer surveys. For example, in a survey of debit customers conducted in May 2016 by Visa, about half (47 percent) of Visa debit cardholders expressed concern about using their PIN to make debit card purchases, with 24 percent of respondents saying that they "don't think it is safe to use [their] PIN," 9 percent saying "it takes longer," and 8 percent saying that they "don't always remember their PIN."[69] Indeed, consumer experience with the choice between using signature debit or PIN debit shows a revealed aversion to PIN. [70] In the United States, consumers have traditionally preferred signature debit over PIN debit. For example, in 2014, 65 percent of debit transactions were made

[68] PETERSON & CONROY, *supra* note 14, at 17, Fig. 5.

[69] Visa Debit Cardholder Research Results (May 2016) (on file with authors).

[70] After a major data breach, many consumers also state that they will no longer shop at the store because of security concerns, but many of them do not follow through on their claim. *Id.*

with signature debit, compared with only 35 percent for PIN debit.[71] Many consumers who could use PIN debit obviously prefer to use signature debit.

Further, according to a report by the Aite Group, one unnamed "sizable" US card issuer initiated its migration with chip and PIN as its preferred CVM for credit cards.[72] The results of the experiment were revealing about consumer willingness to incur higher costs and friction in exchange for PIN verification: "[The issuer] only deployed EMV credit cards to a sample population to test the impact of the more cumbersome CVM. This issuer experienced an 8% drop in transaction volume among the pilot portfolio and is now working on a plan to transition to chip and signature for its credit card CVM." In other words, if consumers valued the added security of PIN verification, they should have used the issuer's card more. Instead, the increased cost of using a PIN card caused consumers to push the PIN-based card to the back of their wallets in favor of other cards that lacked PIN functionality but that consumers evidently found easier to use, regardless of what they said they preferred.

In addition to the per-transaction marginal costs that a PIN regime would introduce, transitioning to PINs would cause large fixed expenditures that likely would be passed on to consumers. According to Aite's estimates, it would cost approximately $3.1 billion to enable all non-PIN-accepting merchants (such as small merchants who lack PIN-capable devices) to accept PIN verification. This figure excludes the cost for sit-down restaurants to purchase pay-at-table terminals (which cost about $500 each, amounting to about $665 million in aggregate). Merchants who currently accept PIN verification (for PIN debit cards) would spend approximately $380 million to upgrade. Finally, staff training time would likely cost about $389 million. Overall, Aite estimated that it would cost merchants $4.534 billion to transition to Chip and PIN.

[71] J.P. MORGAN, *Payments Handbook*, *supra* note 52, at 32, Fig. 28.
[72] PETERSON & CONROY, *supra* note 14, at 12.

Mandating Chip and PIN technology also can be cumbersome for very small merchants who use small, convenient, and simple portable card-processing devices.[73] Payment dongles such as Square allow very small merchants—such as a landscaper, handyman, or farmer's market vendor—to accept card payments by affixing a small payment device to their smartphone or tablet. Those small devices enable merchants to quickly accept payment cards without the cost and inconvenience of a large, PIN-enabled payment machine. Newer dongle models that can accept EMV cards are larger and more expensive than traditional magnetic-stripe receivers, and adding a secure PIN pad would dramatically increase their cost still more and reduce their convenience. In particular, not only must equipment have a PIN pad available, but it must also contain the software to encrypt or tokenize the consumer's PIN.[74]

What's more, Aite Group estimates that it would cost card issuers more than $2.6 billion to transition to universal PIN use. That figure includes the various costs and difficulties related to providing consumers with an initial, temporary PIN that consumers would then be able to reset. Overall, the Aite Group estimates that the total direct cost to issuers and merchants of adopting Chip and PIN would exceed $7 billion. Moreover, that figure excludes any costs from lost sales from payments failures. It also excludes the opportunity cost of slowing many small merchants from adopting technologies (such as Square) that would permit them to accept payment cards (because of the higher cost and size of PIN-enabled devices).

There are also potential dynamic and second-order costs associated with a PIN mandate. Issuers and networks are rapidly developing more secure and less expensive CVM methods that

[73] This cost to small businesses has produced something of a rift in the merchant community between smaller merchants who opposed the EMV migration and larger merchants who supported it, likely for reasons unrelated to consumer security or risk of loss. As Julie Conroy of Aite Group told Bloomberg, "Merchants aren't crazy about this migration to EMV. Many of them are fighting it tooth and nail." Quoted in Kharif & Toness, *supra* note 1.

[74] PCI security protocols prevent entering PINs directly into a business's tablet, and moreover, consumers are likely to be uncomfortable doing so. EMERGING TECHNOLOGIES AND PCI SECURITY STANDARDS COUNCIL, PCI MOBILE PAYMENT ACCEPTANCE SECURITY GUIDELINES VERSION 1.0 (2013), https://www.pcisecuritystandards.org /documents/Mobile_Payment_Security_Guidelines_Merchants_v1.pdf.

can improve security without the additional friction of PIN verification or other similarly high-friction technologies. For example, new methods of customer verification are being developed, including biometrics (fingerprint or retina scans), voice recognition, and device identification with smartphones (for example, verifying the presence of the consumer by geolocating his or her smartphone).[75]

A PIN is often referred to as a static anti-fraud technology because once consumers establish a PIN, they rarely change that PIN and they frequently reuse it for multiple cards and across multiple platforms. In this sense, one can draw an analogy between a static transaction-and-authentication technology such as Chip and PIN and the infamous Maginot Line that France built and relied on following World War I. The Maginot Line was built to repulse German hostility through what was thought to be the most likely direction of a German attack—head on. That direction was chosen because it was generally believed that the German army would not be able to penetrate the Ardennes forest. The French military's reliance on an expensive static defense technology turned out to be tragically shortsighted in the face of a dynamic threat. Instead of relying solely on chip and PIN (the card industry's figurative Maginot Line), card-processing networks are investing major resources in biometrics and other forms of authentication such as fingerprint, retina, and voice scanners. Visa has already introduced a technology that uses a consumer's cell phone to help authenticate a card transaction. In short, this service provides information about whether a cardholder's cell phone is located near the merchant. For example, a transaction in a foreign country—which might otherwise be flagged as potentially fraudulent—could be authenticated through cell phone geolocation.

Variety and experimentation in authentication measures provide for innovation—increased security at lower transactional friction—but constant experimentation also prevents the

[75] *Id.*

Maginot Line problem by reducing the ability for criminals to target one particular, static technology over time. In response to consumer frustration about the perceived slow nature and inconvenience of dipping an EMV card, financial institutions are already rolling out new cards that combine EMV technology with near-field communication (NFC). In January 2017, TCF Financial Corporation announced that it is adding NFC to all of its newly issued EMV cards to increase convenience and to speed checkout.[76] Citigroup is also equipping all of its new co-branded Costco Visa credit cards with NFC, and other issuers are following suit.[77] In many areas outside the United States, contactless EMV cards "are increasingly becoming the norm."[78] Use of such cards is expected to grow rapidly in the United States, further obviating the relevance of a traditional PIN authentication procedure.

Still more dramatic are payment technologies that do not require a physical card. Most notable, of course, is the booming popularity of near-field, contactless payment services such as Apple Pay. These services enable customers to make purchases with high security, without a physical card and with minimal friction. With respect to Apple Pay, the magnetic-stripe information never comes in contact with the merchant's terminal, and consumers need not run the risk associated with inputting one's PIN. Indeed, technologies are being developed today that would eliminate any physical card or device presence. (Instead, fingerprint scans or retina scans—or both—would be used to both make and authenticate purchases.)

At a still higher level, issuers and payment networks are creating ever-more-sophisticated and accurate authentication algorithms to verify transactions. In this sense, the traditional distinction between processing and authentication is increasingly being erased—in the world of big data, every transaction presented for processing also feeds new information into the database

[76] *See* Kate Fitzgerald, *TCF Finds Edge with Chip Card Haters*, AM. BANKER, Jan. 6, 2017.
[77] *Id.*
[78] *Id.*

that processors and issuers use to analyze transactions and develop better models. The major processing networks and issuers are always working to develop better models of fraud prevention and protections for consumers. Increasingly, processing *is* authentication.

In light of the preceding discussion, one could make a plausible case that the marginal benefit from the additional security of PIN authentication at the POS is too small to justify its marginal cost. All told, although Chip and PIN can be a valuable measure in fighting lost/stolen fraud, the marginal value overall is limited once other precautions (such as EMV) are adopted. As industry analysts Thad Peterson and Julie Conroy of Aite Group observe, "[S]ince implementation of EMV without any CVM [cardholder verification method] dramatically reduces the incidence of counterfeit card risk, and since lost/stolen card risk accounts for approximately 9% of fraud losses in payment cards, the relative negative impact of implementing EMV without PIN was low."[79] Therefore, although using PINs likely will reduce lost/stolen fraud, these small—and potentially transitory—gains are likely to be small in relation to the friction from longer checkout times, forgotten PINs, and reduced innovation around payment card security.

B. The Political Economy of the PIN Debate

The foregoing analysis suggests that adding PINs to EMV cards makes little economic sense from a benefit-cost perspective. PIN verification likely will be a passing technology with rapidly declining relevance to the world of electronic payments, and most consumers are ambivalent or hostile toward a PIN mandate. Yet merchants are highly divided on the issue, with many larger merchants in favor and many smaller ones opposed. For example, in a survey of merchants during June 2016, Aite Group found that 77 percent of very large merchants (with more than $50

[79] PETERSON & CONROY, *supra* note 14, at 10.

million in revenue) favored implementation of Chip and PIN, but only 50 percent of smaller merchants (with $500,000 to $2.4 million in revenue) did so. Why are some large merchants so adamant about their support and intensive lobbying efforts in favor of a PIN mandate, including launching several major class-action lawsuits? This intensive and expensive effort seems especially puzzling in light of the fact that merchants that install EMV devices bear no risk of loss from lost/stolen fraud, the only source of fraud that PIN verification addresses.

One possible explanation for large merchants' support for PIN verification relates less to the risk of fraud or merchant fraud losses than to long-standing efforts by merchants to steer consumers toward increased use of PIN networks, which tend to charge lower interchange fees than signature networks.

As noted earlier, signature debit remains very popular with consumers in the United States. Several factors may explain this popularity. First, consumers are averse to the inconvenience of paying with debit (the time and friction of remembering and entering a PIN), and they are afraid that PIN-skimming could dramatically increase their losses if it ended up draining their bank account. Second, consumers are simply unable to use PIN debit for many transactions, such as online transactions and those in sit-down restaurants. Third, consumers may have a heightened sense of confidence in the Visa or Mastercard processing networks as compared with the myriad PIN-debit networks that many consumers do not recognize. Fourth, many merchants, especially smaller merchants, do not accept PIN cards. In many contexts, consumers have a choice: if they prefer the additional security of entering their PIN, they frequently have that option.[80]

[80] In fact, they universally have that option at the very large retailers that are lobbying and suing to require PIN verification.

Larger merchants that own card terminals that can accept both signature and PIN debit, however, generally prefer that consumers use PIN debit. The rationale for their preference has little to do with increased security. Instead, it has to do with the level of interchange fees on different types of payment cards, which are incorporated into the discount rate merchants pay when a consumer pays by card. In general, the interchange fee for card payments is substantially lower for PIN debit than for signature debit. Those savings are passed through to merchants in lower merchant discount rates.

Annual data collected by the Federal Reserve reveals this cost differential. Debit card interchange fees today are set on a two-tier system: (a) large banks (with more than $10 billion in assets) that are subject to the price controls imposed by the Durbin Amendment to the Dodd-Frank financial reform legislation, and (b) exempt banks (with less than $10 billion in assets) that are not subject to the Durbin Amendment's interchange price controls. According to the Federal Reserve, exempt banks provide about 38 percent of the total volume of signature debit card transactions in the United States annually and about 35 percent of the PIN debit transactions.[81] For transactions made by cards issued by Durbin-covered banks, the average interchange fees for signature and PIN debit transactions were virtually identical: $0.23 and $0.24, respectively. For exempt banks, however, the differences were dramatic: the average interchange fee for signature debit was approximately $0.51, compared with $0.26 for PIN debit. Thus, with respect to signature debit transactions made with cards from exempt banks (approximately 38 percent of all transactions), merchants could save substantial sums of money if consumers were compelled to use PIN debit instead. In addition, 95 percent of the transaction volume for prepaid cards (which

[81] *See* BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM, AVERAGE DEBIT CARD INTERCHANGE FEE BY PAYMENT CARD NETWORK (July 22, 2016), https://www.federalreserve.gov/paymentsystems/regii-average-interchange-fee.htm.

constitute a rapidly growing segment of the market) are exempt from the Durbin Amendment's interchange price controls.

Thus, it appears that merchants' primary desire is to increase the overall use of PIN networks, rather than to reduce fraud; indeed, it makes economic sense for merchants to push for a requirement that would allow them to route a larger proportion of their electronic payments through cheaper networks.

Although lower interchange fees for PIN debit versus signature debit explain why merchants as a whole would prefer the former, that does not explain the difference between large and small merchants' support for PIN as part of the EMV rollout. The reality of the way in which interchange fees are set means that larger merchants also benefit disproportionately when consumers use PIN versus signature verification for electronic transactions.[82] For larger merchants, discount rates are typically set by cost-plus pricing, composed of the relevant interchange fee with certain costs added on. Smaller merchants, by contrast, typically have bundled pricing models, in which they are quoted an overall cost for a package of services, including debit and credit card payments. As a result, interchange fees are marginal costs for large merchants and tend to be passed through much more rapidly and completely for large merchants than for smaller merchants.

This observation suggests that the puzzling investment by large merchants to require the adoption of chip and PIN may have little to do with reducing their losses or protecting consumers, but instead may reflect efforts by large merchants to reduce their costs of card acceptance for transactions, using consumer protection as a rationale.

[82] *See* Todd J. Zywicki, Geoffrey A. Manne & Julian Morris, *Price Controls on Payment Card Interchange Fees: The U.S. Experience* (George Mason University Law & Economics Research Paper 14-18, June 4, 2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2446080.

VI. Conclusion

The evolution of payment card security has been driven by an economic logic of maximizing the value of the network for consumers, which is accomplished when network participants choose the mixture of POS and network-level security that minimizes transaction costs of using payment cards, while also adapting to a rapidly changing technological and threat environment. Consistent with experience, the joint-care model developed in this article predicts that the European Union would prefer to rely primarily on POS security because of higher telecommunication costs, whereas the United States would prefer to rely primarily on network authentication. The model predicts that an exogenous increase in the level of counterfeit fraud—largely a result of technological advances by fraudsters—would increase the use of POS security measures, a condition which is consistent with the timing of the adoption of the EMV standard by the United States. Moreover, the use of a liability shift model as opposed to a mandate is likely to act as an efficient selection tool: smaller merchants that are unlikely to be targets of counterfeit fraud can opt out if the risk of fraud is less than the cost of adopting EMV.

The US adoption of EMV was not full throated, in that signature rather than PIN remains the means for consumer verification. This decision makes sense from the perspective of network value maximization. There are strong reasons to believe that the marginal benefit from PIN verification—almost solely a reduction in lost/stolen fraud—is too meager to justify its adoption. Nonetheless, some large merchants have pushed for a PIN mandate. These proposals for government intervention with a PIN mandate now would likely disrupt the dynamic and evolving ecosystem of the evolution of payment cards and payment card security, imposing costs on consumers and merchants with very few benefits. In fact, there is some reason to believe that the recent push for command-and-control mandates on payment card security—particularly lobbying and litigation efforts in the United States by special interests to require Chip and PIN technology—are driven by financial self-interest in lower interchange fees, not by consumer welfare.

Before regulators intervene in a market, they must first determine that (a) there is a market failure, (b) an effective solution to that market failure can be identified, and (c) the benefits of any proposed solution exceed the costs of the intervention, including the unintended consequences. To date, it is difficult to see that there is a market failure in the consumer payment system. Instead, it appears that the system has evolved somewhat spontaneously over time in light of available technology and efforts to reduce payment friction while also protecting consumer security. It seems to make little sense to mandate a particular technology that will soon become obsolete rather than to allow the payment system to continue to evolve.

Appendix: Joint-Care Model and Simulation Results

The payment care industry would like to avoid losses from fraudulent transactions, L, which can be reduced by action both at the point of sale, P, and through the network, N. These actions have marginal costs ϕ and θ, respectively. A consumer's marginal willingness to pay for a payment card transaction is u, and his or her net value from using the payment card network is

$$Max\ V_{P,N,q} = q(u - L(P,N) - \phi P - \theta N). \tag{A1}$$

Maximization implies that the following conditions will hold in equilibrium:

$$-L_P = \phi \tag{A2}$$

$$-L_N = \theta \tag{A3}$$

$$u = L(P,N) + \phi P + \theta N. \tag{A4}$$

These conditions simply indicate that each type of precaution will be used until its marginal benefit ($-L_P$ and $-L_N$, which are avoided fraud losses from additional care) equals its marginal cost (ϕ, θ). The third condition, equation A4, shows that network value is maximized when the marginal value to a consumer from a transaction, u, is equal to the marginal cost, which here is fraud and precaution costs. Clearly, by minimizing the right-hand side of this condition—the sum of fraud and fraud-avoidance costs—welfare is maximized. Because the optimal level of POS and network care is unrelated to output, we focus on the loss minimization problem.

Comparative Statics

How a change in POS usage affects optimal network usage and vice versa can be found by differentiating the first-order conditions with respect to N:

$$L_{PP}\frac{\partial P^*}{\partial N} + L_{PN}\frac{\partial N^*}{\partial N} = 0 \tag{A5}$$

$$L_{NP}\frac{\partial P^*}{\partial N}+L_{NN}\frac{\partial N^*}{\partial N}=0. \tag{A6}$$

Solving yields: $\frac{\partial P^*}{\partial N}=\frac{-L_{PN}}{L_{PP}}$. Because $L_{PP} > 0$, network and POS care are substitutes as long as $L_{PN} > 0$, which implies that the marginal product of POS rises with increase in network care.

The impact of an increase in price of network care on the use of both network and POS care can be found by differentiating the first-order conditions with respect to θ:

$$L_{PP}\frac{\partial P^*}{\partial \theta}+L_{PN}\frac{\partial N^*}{\partial \theta}=0$$

$$L_{NP}\frac{\partial P^*}{\partial \theta}+L_{NN}\frac{\partial N^*}{\partial \theta}+1=0.$$

Solving yields the following two expressions:

$$\frac{\partial N^*}{\partial \theta}=\frac{-L_{NN}}{SOC}<0 \tag{A7}$$

$$\frac{\partial P^*}{\partial \theta}=\frac{L_{PN}}{SOC}>0, \tag{A8}$$

where *SOC* is the determinant from the second-order condition matrix, assumed to be positive for minimum. Because of symmetry in the model, these results imply that $\frac{\partial P^*}{\partial \emptyset}<0$, and $\frac{\partial N^*}{\partial \emptyset}>0$.

Finally, we examine the impact of an exogenous increase in losses associated with any level of care. To formalize this, consider a parameter $\delta > 0$ that represents an exogenous shock to $L(P,N)$:

$$L(P,N)\delta+\phi P+\theta N. \tag{A9}$$

First, we can see from the envelope theorem that total costs increase with α:

$$\frac{\partial TC(P^*,N^*)}{\partial \delta}=L(P^*,N^*)>0. \tag{A10}$$

Differentiating the first-order conditions with respect to δ yields the following:

$$L_{PP}\frac{\partial P^*}{\partial \delta}+L_{PN}\frac{\partial N^*}{\partial \delta}+L_N=0$$

$$L_{NP}\frac{\partial P^*}{\partial \delta}+L_{NN}\frac{\partial N^*}{\partial \delta}+L_P=0.$$

Solving yields the following:

$$\frac{\partial N^*}{\partial \delta} = \frac{L_P L_{PN-} L_{PP} L_N}{SOC} \lesseqgtr 0 \tag{A11}$$

$$\frac{\partial P^*}{\partial \delta} = \frac{L_N L_{PN-} L_{NN} L_P}{SOC} \lesseqgtr 0. \tag{A12}$$

The signs of A11 and A12 are ambiguous because the change in optimal POS and network care in response to a change in potential damages will depend on their relative substitutability.

Simulation

The simulation was based on the following baseline model:

$$TC(P, N) = (PN)^{-0.5} + P + N. \tag{A13}$$

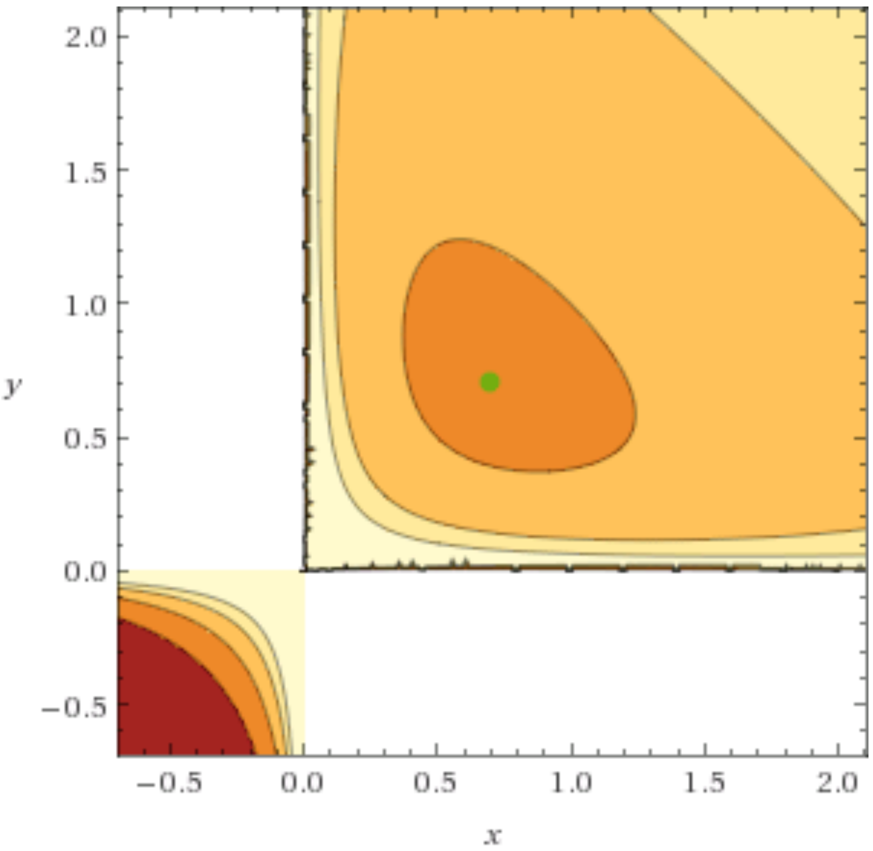
In A1, $L = (PN)^{-0.5}$, and $\emptyset = \theta = 1$.[83] The solution to minimizing (A13) with respect to P and N yields the following values:

- P^* = 0.71
- N^* = 0.71
- $TC(P^*,N^*)$ = 2.82
- $L(P^*,N^*)$ = 1.40

The solution is shown graphically in figure A1.

[83] It can easily be confirmed that $L_P < 0$ and $L_{PP} > 0$.

Figure A1. Graphical Solution to Joint-Care Simulation



Note: X = network; Y = POS care.

To generate the data underlying figure 3, θ was varied from 0.25 to 3.0, holding $\emptyset$ constant at 1.0. The results are listed in table A1.

Table A1. Simulation Results

Marginal cost of network (MC_P held constant at 1)	*P**	*N**	Total costs at *P** and *N**	Losses at *N** and *P**
0.25	0.50	2.00	2.00	1.00
0.50	0.60	1.20	2.38	1.18
1.00	0.71	0.71	2.82	1.40
1.50	0.78	0.52	3.13	1.57
2.00	0.83	0.42	3.30	1.63
3.00	0.93	0.31	3.72	1.86
4.00	1.00	0.25	4.00	2.00