Risk and Privacy Implications of Consumer Payment Innovation in the Connected Age

Ross Anderson

I. Introduction—Sofort or Surcharge?

One might think that innovation in consumer payment systems is hard because payment networks tend to be slow-moving cartels with high barriers to entry, thanks to two-sided market effects and other externalities. And if innovation is hard, then surely new security and privacy risks should be moderate?

Then consider the case of Sofortüberweisung, a controversial entrant to the payment market in Germany. Its name means "instant payment," and its service has taken off rapidly in the past 2-3 years. Branded as Sofort (Instant), this service provides merchants with a low-cost payment service for online shopping. It is promoted by some large sites (such as airlines) by exempting users from the surcharges normally made for credit card payments. So far so good. What might be of interest to regulators is how Sofort managed to break the payment-card cartel. When a German bank customer clicks to pay at a website, Sofort asks for her bank account number, then goes to her bank's website and impersonates her. The bank asks for a PIN and a TAN (a one-time code, typically mailed to the customer); Sofort in turn questions the customer. If her responses lead to a successful logon, Sofort checks her available funds and uses her funds transfer facility to pay for the purchase directly from her account. In effect, Sofort is doing a middleman attack on her bank account in order to deprive the bank of card transaction fees. The merchant typically pays 75 basis points plus 10 cents per transaction rather than 250 or more for online credit card payment. Analysts estimate that Sofort had 1.2 billion euros of the 20-billion euro market for online payments in 2009.

One might think that Germany's 300 banks would object to this, and indeed they did. There was a technical arms race; the banks tried one security measure after another, from CAPTCHAs to IP address blocking. Sofort generally won that race. The banks' payment cooperative sued Sofort for unfair competition and for inciting customers to breach bank terms of service by entering their credentials at

Sofort's website. The case was suspended after the intervention of the German Federal Antitrust Office, which argued that the banks' harmonized terms of service hindered competition and were designed to exclude new business models like Sofort's.

The banks do make clear via their public relations machinery that any customer who gives their PIN and TAN to any third party breaches their terms and conditions and is on their own. Yet while geeks denounce Sofort in blog posts, the consumer-protection issue is far from salient to Sofort's many happy users. The company's own information on system security is reassuring: "Shopping-glück oder Geld Zurück" (Happy shopping or your money back); your banking is protected by your PIN and TAN (not pointing out that it's the PIN and TAN issued by your bank and used against its wishes); and their data protection is approved by the local standards body (whatever that means).

What lessons might be drawn from this? First, while a geek would consider it imprudent to enter bank credentials into the website of a low-cost airline, banks worldwide have trained their users to do just this through the Verified by Visa/MasterCard SecureCode (VbV/MSC) program. In (Anderson, Murdoch 2010) we discussed how VbV/MSC has become perhaps the most successful authentication protocol ever despite poor technical design, because of strong adoption incentives on merchants (who get cardholder-present fees and liability rules). In practice this means that in many countries, transaction disputes are being charged to the cardholder rather than to the merchant. The explanation: "Your password was used so you must have been negligent." So banks trained their cardholders to enter bank credentials into merchant sites, and trained merchants to adopt insecure systems in return for low fees. They sowed the wind with VbV/MSC, and reaped the whirlwind with Sofort.

Second, German banks had already introduced a Giropay system, which they had planned to extend to SEPA e-mandates (Anderson, Murdoch 2010). Such payments have much the same look and feel as Sofort: a customer making a payment at a website is redirected to their bank's logon page to authenticate it. By sending the customer to the bank directly, this mechanism does not have the same potential single point of failure provided by an active middleman such as Sofort, but is still vulnerable to many of the problems with VbV/MSC such as phishing.

Third, the payment-system innovation provided by Sofort may facilitate innovation elsewhere in the economy. The main alternative in Germany, which historically has had low credit-card usage, is direct debit. Tech-savvy Germans may have direct debits set up with large online businesses such as Amazon, but may be reluctant to trust small startups, who as a result might have to operate through Amazon or other portals that charge much higher fees than the card payment system.

Fourth, it is quite normal for firms competing in two-sided markets to offer insecure products in the race for market share and then lock things down later (Anderson 2002). This pattern has been seen in operating systems, mobile phones and social networking systems; there is no reason for payment systems to be any different.

Fifth, if Sofort becomes the dominant player in its market then there will be systemic consequences. It will be a natural destination of an investigator with a warrant; some will consider this a privacy risk (but then so is Visa). Others may see it as a control point where governments could interfere with trade (as Visa blocked WikiLeaks). A compromise of its systems could be expensive, leading to large-scale credential reissue (but the same can be said of Visa, and of firms like Cyota that provide VbV service).

II. Might Mobile Compete on Cost?

The Sofort model is spreading. Not only is Sofort Bank expanding its operations to Austria, Switzerland and Belgium. We now see the beginnings of payment service competition along similar lines in the U.K. This time, it comes from an insider. Barclays Bank has recently piloted a service called Pingit for small payments on mobile phones. In the initial phase, the bank's own customers can make payments up to £300 via a mobile phone app to other individuals and to businesses; the innovation is that the mobile phone numbers of the payer and the payee act as names in the system, as more familiar proxies for bank code and account number. Now that the usability issues have been debugged, the second phase will enable anyone with a U.K. bank account to make payments. The payer will make a single authorization for direct debits to be made to her account; thereafter whenever she presses the "pay" button, Barclays will direct-debit her and send money to the payee directly. This service has the potential, like Sofort, of breaking the payment card cartel (of which Barclays is a prominent member). In the short term, consumers and merchants will win as costs fall. There are already calls for regulation: industry people complain that Pingit will break money-laundering traceability (which is nonsense; if we end up with one interbank payment service provider the police can just subpoena them for everything). But in the medium term, consumer advocates may worry that pressure on margins may erode fraud protection still further.

So can mobile and online payments challenge the existing payment-card cartel? This is a fascinating question. Handling cash costs merchants 2.5 percent to 3 percent of turnover, and credit-card merchant discounts are set to be just competitive with this at 2.5 percent. The case for using cards rather than cash rests on factors such as convenience, credit and marketing rather than cost (Garcia-Schwartz and others 2006). In their history of the credit card industry, Evans and Schmalensee describe the vigorous competition between both issuers and acquirers within the framework set by Visa/MasterCard, which they describe as "co-opetition"; they recount how it drove merchant discounts down from the higher levels in the days of go-it-alone operators such as Diners and American Express. But the industry has largely resisted attempts to make electronic payments substantially cheaper than cash. PIN debit does cost about 1.5 percent but U.S. banks have been resisting attempts by retailers to move their customers to this—for example, MasterCard prevents the U.S. version of EMV (and mobile-wallet versions of PayPass) from supporting PIN debit.

Could a U.S. bank or an "outsider" like Barclays, break the U.S. payment-card cartel by offering a mobile payment service such as Pingit? An instant peer-to-peer payment service, delivered over a mobile channel, could be transformational. If consumers could pay for purchases not just online but also in-store, and merchants benefit from a discount of under 1 percent, then it could give the payment card cartel a real challenge. Merchants might offer triple air miles to entice customers, and even install femtocells at checkouts so that mobile phones would work there. Alternatively, a scheme operator could offer a contactless Pingit card for use in wireless dead zones. Competition of this kind could be economically significant; an efficiency gain of about 1 percent of retail sales would bring real benefits. And the incentives are certainly right for retailers: Wal-Mart processes $200 billion in credit-card transactions in the United States alone.

What might be the lessons for U.S. regulators? If the payment-card cartel is to be seriously challenged then a mobile system backed by ACH might be the way to do it. At present ACH-based consumer payment services are mostly niche players, with the largest being probably PayPal (we suspect most people top up their PayPal account from their credit card rather than using the ACH option, though we're not aware of any data). Mobile platforms might just possibly provide the opportunity to shake up the industry.

Three words of warning though. First, many people have predicted a mobile payment revolution; since about 2002 we've repeatedly been told that within five years m-payment will be big time with a billion users and a trillion a year in turnover, yet it hasn't happened. It is instructive to read and compare the Innopay market analyses for 2010 and 2012 to see how expectations are subsiding (Innopay 2010, 2012). Mobile has taken off in less developed countries that have no alternatives, rather than in developed ones with mature payment ecosystems: they account for 3.3 percent of GDP in Kenya but only 0.05 percent in Japan, the developed country with the highest uptake (IFC 2011). The U.S. market has multiple mobile offerings, some well-established (Obopay was founded in 2005) and some backed by large players (Obopay by Nokia, PayPal X by PayPal, Google Wallet by Google). Yet these remain niche players. The Innopay view is that to prevail they will have to offer speed and security of functionality. To these we might add cost; if mobile payments become cheaper than debit cards, we might see real change.

Second, there will be continuing pressures to reduce, undermine or circumvent the relatively strong consumer protection that U.S. account holders enjoy, and this will be especially the case if mobile succeeds as a low-cost payment channel. We will return to this later. Meantime, it makes sense to regulate Sofort or Barclays in the same way as Visa or MasterCard. In fact, Sofort now has a company in its group with a full banking license, so if the German government had acted against it on security grounds, rather than backing it on antitrust grounds, that would probably have led to a suboptimal outcome. There are outstanding issues around liability, dispute resolution and truth in advertising, but the same can be said for the banking industry as a whole.

Third, a large-scale move to mobile payment platforms will introduce new privacy and security tussles. Customer tracking via cookies is well-established online but has still led to an EU Directive whose implementation is controversial with both businesses and privacy advocates. The tracking of mobile platforms is even more likely to lead to conflict. A consumer's cell site location history is sensitive data, as is her address book; both are collected surreptitiously by mobile companies, which has led to a class action against path.com and congressional investigations into the privacy policies of Apple and of mobile apps generally. Even the late Steve Jobs publicly criticized mobile analytics in 2010 after he found that flurry.com's apps were monitoring devices on the Apple campus (Tofel 2010). There are also issues of security as malware writers turn their attention from the desktop to the handset, now that there's money to be stolen. (I'll discuss malware in more detail below.)

III. Regulation and Risk—130 Years on the Treadmill

The social objectives of payment system regulation may be some combination of efficiency, access (the absence of unlawful discrimination), consumer protection against fraud, rip-offs and liability dumping), privacy protection, and finally the avoidance or management of systemic risk. It is natural for supervisors to pay most attention to whatever aspects currently generate the most controversy, such as the interchange fee issue in recent years (Rochet, Tirole 2006; Chakravorti 2010). But neglected issues can move rapidly up the agenda—so we might perhaps pay more attention to operational risks, consumer protection, privacy and systemic risk.

There is a long history of payment system supervisors acting to protect consumers, only to find that the protection was only partial, and that eventually technological changes allow service providers to wriggle out. An early consumer-protection measure was the Bills of Exchange Act 1882. This responded to fraud as checks became widely used by ordinary citizens as well as by sophisticated merchants. The Act made a forged signature "wholly inoperative," so that a bank in the British Empire could not make its customers liable for a forged check by means of its terms and conditions (unlike in Switzerland where banks did just that). The responsibility for signature verification now fell on the relying party, as it should. But nothing was done about stolen checks. If a thief could open a bank account in the payee's name and cash the check, the drawer had no recourse. This shifted the tussle to the conditions under which a check could be negotiated by endorsement. When the thief of a check payable to "J. Bloggs" found it hard to open an account in that name, he could try to negotiate it by endorsing it with a forged signature and passing it through an account in a different name. Banks responded by overprinting check stock "not negotiable," and the arms race continued when courts in some countries found circumstances in which checks crossed in this way were in fact negotiable after all, leading to more fussy local detail about prudent check crossings.

The pace picked up in the 20th century. The introduction of payment cards into elite markets in the 1960s, followed by their spread into mass markets in the 1980s, made available a new payment instrument in the form of the credit card, with generally good consumer protection worldwide. From the late 1960s banks also started to deploy ATMs, leading to debit cards, which have had a more mixed history and were driven initially by a desire to save staff costs rather than to provide elite service (Batiz-Lazo 2010).

The treatment of specific payment instruments can vary across jurisdictions. In the United States, the signal ATM case was Judd vs Citibank. Dorothy Judd claimed $800 from Citi in disputed ATM transactions; Citi said that as its systems were secure, she must be responsible. The judge ruled that he was "not prepared to go so far as to rule that where a credible witness is faced with the adverse 'testimony' of a machine, he is as a matter of law faced also with an unmeetable burden of proof" and found in her favor (Judd 1980). Regs E and Z now entrench that view in the U.S. regulatory system. In the U.K., the first serious case was McConville and others v Barclays and others, where 2,000 plaintiffs sued 13 financial institutions for £2 million in disputed transactions. The banks' lawyers persuaded the court to split it up into separate small-claims cases, arguing that they would all be too different for a class action to make sense. Two years later, it turned out that the judge had got it wrong: Andrew Stone was sent to prison for 6.5 years for leading this crime wave. (The McConvilles, however, never got their money back.)

The banks introduced a Banking Code under which customers are supposedly only blamed for fraud if they were grossly negligent; but once the media fuss had died down, banks started claiming that cardholders whose card details and PINs were used in fraud were grossly negligent. Online banking was the scene of the next tussle as the dotcom boom in the late 1990s saw banks rush to offer services via the Web. The effects were documented by Bohm, Brown and Gladman: after some vacillation, banks harmonized their terms and conditions to the effect that a customer who accepted a password for Internet banking would be held liable for any transaction that the bank claimed had been authorized using it (Bohm, Brown, Gladman 2000). So as passwords replaced signatures, the protection introduced by Gladstone was quietly sidelined. People who complain of fraud are routinely told, "Your password was used, so you're liable."

The *danse macabre* of banks and regulators in the U.K. continued with the Financial Services Act 2000, which established the Financial Ombudsman Service, an arbitration system for dispute resolution between banks and customers, but which appears to have been largely captured by the banks (Anderson, Bohm 2008). The European Union's Payment Services Directive of 2007 brought in various provisions for consumer protection. This was advertised as stopping banks dumping fraud liability on customers, yet seems to have had little effect on national practices.

The situation across Europe is variable, but generally better than in Britain. The 2010 Eurostat crime survey ranks all 27 EU countries by online users' concerns and finds that the U.K. is second worst after Latvia for fear of online payment card fraud, fear of phishing attacks on online bank accounts, and fear of privacy violations; it's also fourth for spam and sixth for virus infections (Eurostat 2012). In a report to ENISA in 2008 we recommended that comparable bank fraud statistics be recorded for all EU member states (Anderson and others 2008); such figures will be collected from 2012 for all seven eurozone countries. There will also be a further Eurostat survey of citizens' experiences of cybercrime in 2014. We will be interested to see whether fraud is higher in countries with good consumer protection, such as Finland and the Netherlands, or in countries with weak protection such as Britain, Latvia and Spain. It is noteworthy that the United States does not have central fraud reporting, a topic we'll revisit later.

Another variable that may bear watching is finality of settlement. In a previous study, we observed that fraudsters preferred to attack payment mechanisms with rapid final settlement, and to avoid those that permitted stolen funds to be clawed back for an extended time period (Anderson 2007). The Payment Services Directive imposed a uniform 48-hour settlement deadline for electronic transactions in the Single European Payment Area. Yet there are still variations. The U.K. government, for example, prodded banks to introduce a Faster Payments Service, which reduces the delay in electronic payments from one customer account to another from three days (under the old BACS system) to near real time. It will be interesting to see what this does for fraud; anecdotally, industry insiders suggest losses are on the uptick. We're not aware of any published data, but Faster Payments limits vary so widely from one bank to another (from £5,000 to £100,000) that we expect some interesting data in due course.

IV. Cybercrime Patterns

In order to put the likely risk evolution in context, it may be useful to consider the overall cybercrime picture. A recent study for the U.K. Ministry of Defense (Anderson and others 2012) classifies cybercrime into four categories:

1. Traditional offenses such as tax fraud and welfare fraud that are now classed as "cyber" by virtue of the fact that tax returns and welfare claims are filed online, but where the substance is much the same as a generation ago (in the case of tax and welfare fraud, misrepresentation of income/capital/relationships);

2. Offenses such as card fraud that have been around for a generation, but where both the modus operandi and the main countermeasures are changing rapidly with technology. The report calls these "transitional" offenses;

3. "Pure" cybercrimes against individual victims of a kind that did not exist offline, such as extortion using fake antivirus software;

4. "Platform" cybercrimes that provide illegal services to criminals committing offenses of types 2 and 3, such as the provision of botnets and cashout services.

The big picture is that in traditional frauds, the direct losses are much greater than either the costs in anticipation (such as security measures) and the costs in consequence (such as law enforcement); in pure cybercrimes, the reverse holds, with cybercriminals imposing billions of dollars of costs on the world economy while managing to steal only a few hundred million. Payment systems are a microcosm: the direct costs of card fraud ($9.2 billion) exceed the indirect ones ($2.4 billion) while for online bank fraud, the indirect costs are greater ($1 billion versus $690 million). In short, the more "modern" or "cyber" a payment system is, the harder it seems to be to defend it efficiently. This may be partly a learning effect, but externalities surely play a role, too.[1]

There is a further rider: if we include in the indirect costs an estimate of the opportunity costs—the value of business foregone, by both customers and merchants, because of the fear of fraud—then these numbers may be several times higher. The actual amounts are uncertain, but we can perhaps get defensible order-of-magnitude estimates from survey data. One Visa merchant survey, for example, suggested that merchants turn away $4 in business for every $1 they suffer in fraud (Khan, Hunt 2012). Yet it is not clear that all these $4 were lost to the economy; people who fail to shop at one website may shop at another or at a physical store. As a reasonable guess, we might end up with global indirect costs on the order of $10 billion for users and $20 billion for firms. (For a more detailed discussion, see Anderson et al. 2012.)

The takeaway message is that payment fraud is a large business. It's worth on the order of $10 billion a year to the bad guys—bigger than Facebook's turnover, but not as big as Google's. Specific defenses against fraud, and generic defenses against cybercrime, are worth maybe $3 billion each, while the indirect costs of cleanup and of lost business and confidence might be in the low tens of billions each. So if we include the indirect costs too, payment fraud might lie somewhere between Google and Microsoft in turnover. As for the growth prospects, fraud accounts for about 5 basis points of cardholder-present transactions but 30 basis points for cardholder-not-present. So if a further 10 percent of world GDP moves online over the next 10 years, we might see fraud increase by 0.025 percent of world GDP, which is $15.7 billion (though we'd hope we'd get better at fraud prevention and perhaps limit the rise to half that). It's important to realize that the move online is associated with real improvements in social welfare because of efficiency gains, and the same will almost certainly be true of mobile. Becker pointed out in the 1960s that the socially-optimal level of crime is not zero (Becker 1968), and that certainly holds for payments.

What's more, this isn't just a macro effect, of decreases in transaction costs improving welfare despite higher fraud; there are micro effects, too. The United States, for example, accounts for 47 percent of all card fraud despite generating only 27 percent of the transaction volume. This is partly because of much greater competition between issuers; they are reluctant to decline transactions as customers will just start using a different card (Business Wire2011). Yet no sane lawgiver would want the United States issuing market to be as concentrated as the typical European one is. And if reasonably open mobile wallets take off, then there should be the same issuer competition as with cards; combined with the technological novelty and the strong externalities, this should lead us to expect a significant increase in fraud.[2]

V. Trends in Mobile Payment Systems

Mobile payment systems have been around for about a decade and are now widely used in less developed countries. A typical system, such as Kenya's M-PESA, lets a user access a bank account from a mobile phone, authenticating herself using a PIN that is encrypted in the SIM card and verified using standard banking technology. Payments can be made from one account to another by encrypted SMS messages. Such phone payment systems are expanding from phone-to-phone to phone-to-agent and even agent-to-agent; M-PESA does this, and Easypaisa is doing it in Pakistan. A phone payment system can thus grow into a physical network that looks somewhat like a bank branch system or a network of Western Union franchisees. The establishment of such systems in countries with poor banking systems leads to significant social gains; philanthropists such as the Gates Foundation have invested in supporting them (The Economist 2011).

A different technology, near-field communication (NFC) payment, was pioneered in Japan and introduced to the U.S. market in 2011. NFC is a radio communications standard designed to communicate with RFID (radio frequency identifier) tags, contactless smart cards and similar low-cost devices over a range of an inch or so. Contactless cards are already used in ticketing applications such as London's "Oyster" card for public transport. NFC technology allows a suitably equipped mobile phone or tablet to act as either the payment card, or the terminal, or both. Contactless payment used to involve dedicated tickets or cards talking to dedicated terminals; now it can become a software platform at one end or both, and this can support innovation in all sorts of new ways not just for payment but for apps such as transport and event ticketing, marketing coupons and loyalty programs.

An interesting general example is the Google Wallet.[3] This is a software app for the new NFC Android phones that supports NFC payments and enables other phone apps to interface to the payment system. Such phones contain a Secure Element (SE), a smart card chip mounted in a tamper-resistant package with an NFC chip and antenna. A bank can load a payment card into the SE chip in the form of a signed Java card applet; the user can then select it using the phone's screen and use it to pay, whether by tapping it against a payment terminal in a physical store, or by an online transaction. The wallet and its associated infrastructure deal with the tedious problems such as provisioning the phone with the right cards, revoking them should the phone be lost or stolen, and logging transactions to resolve disputes. (This is a simplified description; see Anderson 2011 for more detail.)

Mobile wallets will in future mediate access to the payment mechanism by other apps, which are assumed to be untrusted. Without this, an evil app could phish the user by saying "please enter your PIN to pay $2.50 to play this online game" while actually kicking off a large transaction elsewhere. By providing a trustworthy user interface and logging, the wallet can create a payment platform that supports innovation by other businesses. As Google is an advertising firm, their wallet is designed to support coupons and offers; platforms offered by other firms might have a different flavor. For example, Isis is a venture backed by Verizon, AT&T and others, working on standards for phone banking, prepaid cards and charge cards.[4] This will no doubt reflect the mobile operators' view of the world, as tends to be the case with the SIM-based payment platforms offered by operators in many less developed countries. And then there are the disruptive small entrants, such as Square, a company started by the founder of Twitter; its product line is aimed at challenging not just Google on wallets but VeriFone on terminals.

Darin Contini and others report a 2010 Federal Reserve meeting whose participants advocated an open platform for NFC payments, envisaging collaboration between financial regulators, the FCC, the FTC and bodies such as NACHA (Contini et al. 2011). They envisaged a single platform supporting multiple payment channels, from ACH to carrier billing, and common technical standards including dynamic data authentication (DDA) and for certification. They held out the hope that with the mobile phone used as a security tool for authentication at the point of sale and over the Internet, as well as in new NFC and peer-to-peer payment channels, there is a prospect of significant fraud reduction. Furthermore, eliminating physical cards would cut issuer costs, while removing magstripe data from merchant systems would cut the cost of PCI compliance. This vision helped guide industry players in the development of mobile wallets.

There are certainly cost savings to be aimed at, and the early experience of Google, Isis and others should help quantify them. But DDA is no panacea, and certification is hard, too. Europe rolled out EMV first, and has had many failures of hardware, software, protocol design and certification. Once the PIN entry devices (PEDs) used in EMV (chip and PIN) transactions were fielded at scale, terminal-tampering attacks turned out to be trivial, despite a much-trumpeted evaluation scheme (Drimer, Murdoch, Anderson 2008). We then discovered that a thief can use a stolen card (for which he does not know the PIN) by using an electronic device to manipulate communications between the card and the PED. The card believes it's doing a signature transaction while the PED believes that the card accepted an entered PIN; and this works regardless of whether DDA is used (Murdoch et al. 2010). The flaws in the DDA payment protocol design are simple enough but fixing them appears to be intractable because of the incentives facing different actors. Governance is hard in a payment system involving hundreds of vendors, tens of thousands of banks and millions of merchants. Everyone wants to cut costs and customize systems, both of which undermine security; and when a systemic vulnerability emerges, no one will step up to the plate. More complex value chains involving more diverse stakeholders will make governance even harder.

The killer is Wilkes' law. Imagine there's a sudden problem with relay attacks. At present, it's possible to connect a false EMV terminal remotely to a false card, so that when the victim buys coffee from a vending machine on which the false terminal has been fixed, a crook can take money from an ATM hundreds of miles away using the false card. With conventional EMV this requires specialist equipment, so it's not been industrialized at any scale (suspected losses are only in the hundreds of thousands). But once mobile phones do NFC, a crook can program one phone to act as a false terminal, and another to act as a false card. An attack that used to require serious engineering is now just a software app. This is Wilkes' law: "everything becomes software in the end." It applies to crime, too; while pick pocketing used to take long and arduous training, a pervasive mobile platform can reduce it to a piece of software that might take real skill to write, but can then be copied infinitely. Crimes can be pirated just as easily as music. Once a card cloning scam gets into widespread use, who's going to stop it, and how?

There are problems with carrier billing, whose viability is threatened by fraud according to some industry sources. First, there's a problem with malicious smartphone apps: most bad apps being removed from the Android app store in 2011 were dialers that called premium-rate numbers. Second, there's sharp growth in PBX fraud, where bad guys acquire accounts on corporate switchboards (often by exploiting default passwords) and use them to call premium-rate numbers. Third, enforcement against premium-rate fraud is poor; while victims are too dispersed to shout loudly, the telcos share the proceeds and so have no real incentive to crack down. Finally, no one really knows how much is being stolen, with estimates ranging from the low billions per annum globally right up into the tens of billions. If payments migrate to carrier billing on a large scale, this might become a big deal for financial regulators. But the fees for carrier billing are so high (typically 30 percent) that this channel competes mostly for virtual goods that sustain large markups, for poor customers and for tied services. And with chargebacks in some countries now over 20 percent, even these markets may become unviable. As phone malware spreads from China to the United States, we may see some interesting times.

The payment services associated with cybercrime also bear watching. At present the payment system of choice for scamsters is Western Union, as it enables scam victims to make irrevocable payments that can be collected immediately overseas in cash. Other payment systems are favored for internal use by the online criminal underworld—the people who herd botnets, operate pay-per-install services and trade financial credentials. For them, both irrevocability and untraceability are at a premium (Anderson 2007). A popular service was eGold, but after it was raided by the FBI the action appears to have moved to services based in Russia such as WebMoney. Other payment systems feed "High Yield Investment Programs," also known as postmodern Ponzi schemes. There's an ecosystem of such schemes which pay very high yields to early investors and then stop paying, supported by ratings agencies which track on a daily basis which schemes are paying and which aren't. Many "investors" seem aware they're Ponzi schemes, and hope to get in and out of a scheme quickly before it stops paying (Moore et al. 2012). We know little about this ecosystem—we don't even know how many real principals lie behind it, let alone who they are. Perhaps the combination of phone payment networks with new international remittance services will open up new channels for laundering the proceeds of crime. The cautious regulator may prefer to tread carefully because of the net social gains from a more competitive remittance system; but those payment systems which serve only Ponzi schemes appear to break laws and merit investigation.

Pornography is big business online too, but rating firms such as FICO and Google are reluctant to try to tell the good from the bad and the ugly. Google, for example, will serve porn to those who ask for it, but won't optimize its search services for porn as it does in other sectors. There have been firms offering payment gateway services for pay sites but, as anyone familiar with the literature on adverse selection and moral hazard might expect, they have a bad history (Campbell 2005). The alternative to paid-for porn is free porn, but most pay-per-install services—villains who will install your choice of malware on thousands of machines in return for a modest payment—are linked to porn sites. The cost of free porn is often getting your machine infected (Wondracek et al. 2010). These problems will no doubt migrate to mobile platforms too as they become more pervasive.

The strategic risk with mobile payments generally is of an attack that makes fraud so easy that a platform or channel becomes unviable. The nightmare scenario of the wallet engineer is that malware on the mobile phone might take it over so comprehensively that a remote software attack becomes possible. If I can infect your phone, go into a shop, buy diamonds and bill the transaction to your phone while it sits quietly in your pocket, then its viability as a platform is at stake. Hardware security devices such as the Secure Element are designed to reduce such risks, but it's always possible that design error or governance failure could lead to catastrophe.

An optimist will take the view that disasters have been localized in the past. It's always been easy for a smart crook to loot a few accounts with a few million in them, but that doesn't threaten the system; and if someone invents a mass-pillage attack that can book a large volume of low-value debits, the problem is finding somewhere to send them without being caught. So far no one's managed to do that. Even the no-PIN attack has not been industrialized at any scale, and if the carrier billing mechanism breaks down because of fraud from mobile malware, it won't be the end of the world.

A pessimist will take the view that once all the authentication tokens we use in our lives—our credit cards, passports and car keys—become NFC apps on our mobile phones, we are creating a huge target and at the same time a serious governance problem. He will also argue that a constant low level of fraud can undermine confidence, dumping large opportunity costs elsewhere. (But then, that's already happened in countries with poor consumer protection like the U.K. and Latvia, and the world continues to turn; and phone vendors may be more motivated to fight malware than Microsoft used to be.)

VI. Where Else Might Competition Come From?

A large niche that may drive payment innovation is retail marketing. In the past, store loyalty cards have mutated into credit cards; the U.K. retailer Tesco launched a bank as a branding operation for the Royal Bank of Scotland (RBS) which handled its card issuance and ATM operations, then bought RBS out and set up a proper bank when RBS ran out of money in 2008. We already mentioned Facebook Payments; there is currently an explosion of interest in social marketing, with Groupon creating some excitement in the run-up to its IPO. There have also been rumblings from large retailers in some countries about setting up their own captive acquirers in order to cut card-processing fees. There are enough incentives here; the question is whether anyone capable will make a go of it.

Another possible source of new competition is managing merchants' risk. At present the heavyweight fraud-risk management is done by card issuers, as acquirers tend to be concentrated. Yet as more and more business goes online, merchants face an increasing share of the risk. The leading U.S. acquirer, First Data, is starting to offer risk management, but the industry perception is that the acquirer-side services are not yet as competitive as the issuer side.

Peer-to-peer payments are another source of competition. Some countries, like Germany, have almost abolished checks. U.K. banks said they would like to, but were stopped by the government, which worried about what might happen if the 9 million adults who do not currently bank online were suddenly forced to. But if I can no longer send my mum a check for the wool when she knits me a jersey, what am I to do? A number of startups have begun offering peer-to-peer payments, such as ZashPay and Popmoney. So far, they have tended to be bought by established players; these two firms were bought by Fiserv, whose model appears to be to buy payment service providers in many different niches, then industrialize them by improving the fraud detection and marketing.

Another class of financial-industry mold breaker is the low-cost remittance service. An example is oanda.com, a Canadian company that competes with high-street banks, Western Union and Hawala operators to help send money internationally at low cost. Oanda is a member of SWIFT; unlike traditional operators whose Forex rates include a bid-offer spread of 3 perecent to 10 percent, they offer interbank rates and a fixed fee of $25. According to Western Union's 2010 financial report, the main competitive factors in consumer remittances are brand, trust and distribution; building a direct competitor to their many thousands of franchisees in shops worldwide would be expensive. But with phone payment operators emerging in most LDC markets, a modern global payments business only has to link up to local or regional networks. The main problem now facing new payment market entrants, according to an executive of one of them, is the overenthusiastic interpretation of anti-money-laundering regulations, especially in the United States, which can lead to payments being blocked for days with no explanation or recourse.

A novel and controversial payment service is Bitcoin. This is a currency invented by "Satoshi Nakamoto," the pseudonym of an unknown cryptographer. People mine bitcoins by solving cryptographic puzzles and can then trade them; they are converted to and from U.S. dollars on a market run by several small firms. Bitcoins, being digital, have a number of features attractive to techies; there is a scripting language that enables you to make payments subject to time locks or other computational conditions. But their price depends entirely on demand in a small and not very efficient market; it peaked in June 2011 at almost $30, fell to under $3 by October 2011, and currently trades just over $5. It might be more accurate to think of them as bearer securities rather than currency: they are a store of value (of sorts) but not a medium of exchange except in that they can be tracelessly transferred from one holder to another. There is a concern that criminals with large botnets have been using their computational resources to mint bitcoin, and that they are used in Silk Road, an anonymous black market. This has led to U.S. senators calling for Bitcoin to be investigated by the U.S. Attorney General, and to bitcoin exchanges calling for the currency to be regulated (Bitcoin 2012).

The world of credit can also give us some pointers to possible future innovations in payments. Social credit has been established for some years, with the Grameen Bank earning its founder a Nobel Prize; there are now numerous online social lending systems such as zopa.com, prosper.com, lendingclub.com and smaba.de. These have a number of operational models; the "social" aspect can involve using social pressure to ensure payment or having individual lenders decide whether to offer loans. There may be privacy issues here as credit data can be disclosed to many potential lenders, and poorer borrowers are pushed to expose the private data of relatives (Böhme, Pötzsch 2010).

A recent development, from firms like Telrock, is to use a consumer's transaction stream for credit risk management. Cardholders who miss payments are encouraged to opt in to surveillance in order to escape aggressive calls, but get constant nagging and nudging instead: "How come you just spent $372 at Macy's when you need to make a card payment of $590?" This might conceivably be welfare-enhancing for people with poor self-control but also raises the question whether more "efficient" debt-collection mechanisms will be used to help the poor manage their finances better, or to get them deeper into debt, keep them there for longer and charge them even more. There is growing controversy in both the United States and the U.K. about payday lenders, with a new generation of online firms like wonga.com grabbing market share from old-fashioned pawnbrokers and check cashers despite interest and fees which can amount to thousands of percent per annum. Without regulation, we may see the emergence of a new underclass of digital sharecroppers, held in debt bondage by ever more sophisticated online and social tools. (In the United States, the concerns raised here may be more within the remit of the CFPB than the Fed but should still not be ignored.)

So far, we have not seen social mechanisms extending much into payment products. There are payments in social networks such as Facebook Credits, but

Facebook Credits is a centralized system used to levy a tax on user payments to game operators and other merchants operating within the Facebook ecosystem. (As Facebook takes 30 percent of all money spent via Facebook Credits, it's unlikely their system will spread beyond their tied services, digital goods and other niches unless the business model changes.)

We do know that more information sharing between banks helps cut risk of defaults (Jappelli, Pagano 2002) and could cut exposure to cybercrime (Moore, Clayton 2008). The FS-ISAC has existed for over a decade, and some banks are starting to get keen (Kapner 2012). But the most likely near-term future large-scale use of social data is by fraud analytics firms such as FICO that use dynamic profiles of cardholders to screen transactions on behalf of issuers; such firms do indeed see this as a hot opportunity (Zoldi 2012). Their systems cut fraud in cardholder-present transactions from 18 basis points in 1992 to 5 basis points now; if social data can be used to cut cardholder-not-present fraud from its current level of about 30 basis points, this could be a real benefit. Mobile data might also help: transaction location is already an input to some fraud engines. But the use of social and mobile data in fraud profiling might bring real problems of privacy and access.

VII. What Way Forward for Regulators?

The modern world demands ever more (and more complex) public goods—from a clean environment, through dependable critical infrastructure, to financial sustainability. Humanity's struggles to meet this challenge might be the defining story of the 21st century (Wolf 2012). The costs raise questions about the sustainable borders of the state, especially in post-industrial and post-credit-boom states with falling populations (Helm 2012). The upshot is that policymakers have to prioritize. But prioritize what?

Culture matters. In a recent review of the nuclear industry, *The Economist* wrote, "safety requires more than good engineering. It takes independent regulation and a meticulous, self-critical safety culture that endlessly searches for risks it might have missed" (*The Economist* 2012). Regulators can help shape culture over time. But which organizational cultures should be targeted, and with what interventions?

In the absence of a clear and present danger, the strategic priority of a smart regulator should be better information, so that when events suddenly demand action it has some hope of being effective. So let's summarize what we know about payment systems innovation. First, as the world moves online, fraud is likely to increase, as online card fraud is typically six times the level seen in face-to-face transactions. The net social welfare gains could still be considerable though. The same is happening with mobile payments, which are bringing huge social gains to countries like Kenya, Pakistan and South Africa, and will benefit the developed world too (though the revolution promised 10 years ago hasn't materialized yet).

Second, innovation in developed markets is likely to be driven by the high costs of the existing core cartel. Competition can come from either insiders who break ranks, or external challengers—whether new platforms like mobile or social, niche services such as global remittances or consumer credit, or maybe even off-the-wall ideas like Bitcoin.

Third, cost pressures will push innovators to circumvent consumer protection if they can. This may cause governance failures and erode the incentives on industry players to fight fraud, leading not just to higher costs for consumers but overall. There may be real tensions between competition and security; monopolies may be better at managing the costs of crime in the short run but impose large social costs in the long run. Fourth, there is a small risk of a large-scale technical failure, whether a sudden catastrophic compromise, or a rolling governance failure of a payment ecology where no single player has the incentive to step into the breach.

Fifth, there is a risk of a confidence failure if ever more people experience fraud losses against which they could not have taken effective precautions. The uptake of e-commerce is already slower than it should be, and worse in countries with poor consumer protection (though opportunity costs are hard to measure with any precision).

Sixth, given that both technology and business models are changing rapidly, it makes little sense to regulate technical details such as whether consumer logons to electronic payment systems should use cryptographic challenge-response mechanisms rather than passwords. The important thing is to regulate desired outcomes, which boil down to an optimal combination of innovation, competition and traditional consumer protection (against fraud and privacy compromise). In fact one can see the regulator's job as the protection of consumers, defined slightly more broadly: it's about preventing not just the fraud and embarrassment of operational security failures, but also the high costs and lost innovation that follow failures of competition, and the asset losses that flow from institutional collapse.

Under the circumstances, the immediate priority for payment system regulators must be to get better information about what's happening. Some countries are taking steps towards this; Singapore tightened regulation post-Leeson, bringing technical experts into its discussions with bankers, while the Banque de France has set up an Observatory to measure fraud.[5] In work done for ENISA in 2008, we recommended that the EU collect comparable statistics on fraud across member states; from this year this will happen within the eurozone.

What I suggest for discussion is that the Federal Reserve set up a fraud analysis center, whose mission will be to collect fraud statistics not just for cards but for mobile and all other payment channels. There are several possible models to consider.

One option would be a pure public-sector body, centrally funded (as is the Banque de France's Observatory) and given the power by Congress to demand reports from all payment service providers. Another might use as a model the

National Cyber-Forensics & Training Alliance (NCFTA), the hub of America's cybercrime effort, which has a substantial public-sector input in the form of agents seconded from the FBI and the Secret Service, but which also works with the big service firms and with academics to turn data into both actionable intelligence and a strategic picture. A third model could be the private-sector firms that accumulate information for the benefit of subscribers; they include both for-profit firms like FICO and Nilson, and nonprofits such as the U.K. Card Association, which collects fraud statistics in Britain and shares them with member banks. It may be simplest to try voluntary pooling of information to begin with.

A good start might perhaps be made by collecting what's available publicly and asking both banks and other system operators politely for the data, giving overall estimates to the public and sharing better data with providers who cooperate and bona fide researchers. Links to academic researchers and to cybercrime bodies like NCFTA could add real value. Finally, no regulator should neglect payment system architecture, as this can define the platform for innovation and set the parameters within which consumer protection and competition are traded.

VIII. Conclusions

The world of payments is getting more complex, fast. Fraud is quite likely to rise as more and more transactions go online, and consumer protection is likely to be eroded as new payment systems fall outside the traditional frameworks. This could give rise to problems of access, consumer protection and privacy protection; if new monopolies emerge, or old governance structures fail, it might increase systemic risk. Regulators will face new challenges, and it's hard to predict what they will be.

Technical security is getting harder. Each new technology evolution starts up the arms race of attack and defense once more, and mobile is no exception. It also expands the circle of stakeholders in the payments system. The nonbank players used to be specialist service firms like First Data and FICO; now they include Microsoft, Google, Apple, hundreds of mobile network operators and thousands of app developers. The governance issues of dealing with compromises are going to be seriously difficult. (Privacy may be harder still, but is likely to be driven by European data protection law more than by U.S. regulatory action.)

Yet America needs better data on fraud, as do we all. Defensible statistics for card payments will not be enough. Analysts need to be able to watch what's happening with mobile, with other new competitors, with telcos, with Facebook and with niche channels too. Financial supervisors have a vital role here. Eventually the Fed may decide to ask Congress for the regulatory power to collect data from all payment service providers; meanwhile a start can be made by building links, sharing data on a voluntary basis and growing the capability organically. Others, such as NCFTA and NACHA, may look for actionable intelligence; someone should be analyzing data for the strategic picture, and that might well be a role for the Fed.

Finally, although sharing information helps, compelling sharing could be difficult. The stakeholders are many and diverse, and mobile payments touch the turf of many government agencies. An appeal to providers' enlightened self-interest may be quicker than legislation, and a multistakeholder approach may work better anyway.

Author's note: I'm grateful to Johann Bezuidenhoudt, Rainer Böhme, Steven Murdoch, and Scott Zoldi for discussions on these issues.

Endnotes

[1]These figures give no more than order-of-magnitude indications; Nilson puts global card fraud at $7.6 billion (Business Wire 2011). There is also an open question about the proportion of general "cyber" defense costs to apportion to the prevention of online payment fraud (these costs include $3.4 billion expenditure on antivirus software and similar measures, and a whopping $20 billion for the costs to users and firms of cleaning up infected machines).

[2]The mobile value chain is also more complex. The processor designer may invent a new access control mechanism, but has to sell it to the chip designer, get it supported by the operating system vendor and then promote it to wallet designers. An operating system upgrade is only rolled out if both the handset vendor and the mobile network operator agree. As a result, most smartphone handsets have exploitable vulnerabilities.

[3]Full disclosure: I worked on the design of the Google Wallet in January-February 2011 while on sabbatical as a visiting scientist at Google.

[4]See *http://www.paywithisis.com.*

[5]See *http://www.banque-france.fr/observatoire/home.htm.* The Observatory was set up by a specific law with representatives from issuers, merchants, consumers and experts.

References

Anderson, Ross. 2002. "Why Information Security is Hard—An Economic Perspective," in ACSAC 2001, pp. 358-36.

____________. 2007. "Closing the Phishing Hole—Fraud, Risk and Nonbanks," at *Nonbanks in the Payment System*, Santa Fe, N.M., May 2007.

____________. 2011. "Can We Fix the Security Economics of Federated Authentication?" at *Security Protocols Workshop 2011*, Springer LNCS 7111, pp. 25-48.

Anderson, R., and N. Bohm. 2008. "FIPR Submission to The Hunt Review of the Financial Ombudsman Service," January 2008, at *http://www.fipr.org.*

Anderson, R., R. Böhme, R. Clayton, and T. Moore. 2008. "Security Economics and the Internal Market," European Network and Information Security Agency (March 2008) at *http://www.enisa.europa.eu/pages/analys_barr_incent_for_nis_20080306.htm.*

Anderson, Ross, Steven Murdoch. 2010. "Verified by VISA and MasterCard SecureCode: or, How Not to Design Authentication," at *Financial Cryptography,* Springer LNCS 6052, pp. 336-342.

Anderson, R., C. Barton, R. Böhme, R. Clayton, M.J.G. van Eeten, M. Levi, T. Moore, and S. Savage. 2012. "Measuring the Cost of Cybercrime," in submission to WEIS 2012.

Batiz-Lazo, B. 2010 "Emergence and Evolution of ATM Networks in the UK, 1967-2000," SSRN working paper no. 1534713 (2010) 73.

Becker, Gary. 1968. "Crime and Punishment: An Economic Approach," *The Journal of Political Economy*, vol. 76, (1968) pp. 169-217.

Bohm, N., I. Brown, and B. Gladman. 2000. "Electronic Commerce: Who Carries the Risk of Fraud," JILT 2000 3, at *http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2000_3/bohm/.*

Böhme, R., and S. Pötzsch. 2010. "Privacy in Online Social Lending," AAAI Spinr Symposium on Intelligent Information Privacy Management.

Business Wire. 2011. "U.S. Leads the World in Credit Card Fraud, States the Nilson Report," Nov. 21, 2011, at *http://www.businesswire.com/news/home/20111121005121/en/U.S.-Leads-World-Credit-Card-Fraud-states.*

Campbell, D. 2005. "Operation Ore Exposed," PC Pro, July 1, 2005.

Chakravorti, S. 2010. "Externalities in Payment Card Networks: Theory and Evidence," *Review of Network Economics*, vol. 9, no. 2.

Contini, Darin, Marianne Crowe, Cynthia Merritt, Richard Oliver, and Steve Mott. 2011. "Mobile Payments in the United States—Mapping Out the Road Ahead," Federal Researve Bank of Atlanta, Federal Reserve Bank of Boston, 2011.

Drimer, S., S. Murdoch, and R. Anderson. 2008. "Thinking Inside the Box: System-Level Failures of Tamper Proofing," at 2008 IEEE Symposium on Security and Privacy, pp. 281-295.

Eurostat. 2012. Crime Statistics, Eurostat, at *http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Crime_statistics.*

Evans, David, and Richard Schmalensee. 2005. "Paying with Plastic," MIT Press.

Garcia-Schwartz, Daniel D., Robert W. Hahn, and Anne Layne-Farrar. 2006. "The Move toward a Cashless Society: A Closer Look at Payment Instrument Economics," *Review of Network Economics*.

Helm, D. 2012. "The Sustainable Borders of the State," *Oxford Review of Economic Policy*, vol. 27, issue 4.

IFC. 2011. "IFC Mobile Money Study 2011," Anato Onoguchi, Leila Search, Piya Baptista.

Innopay. 2010. "Mobile Payments 2010," Chiel Liezenberg, Douwe Lycklama.

________. 2012. "Mobile Payments 2012," Chiel Liezenberg, Douwe Lycklama.

Jappelli, R., and M. Pagano. 2002. "Information Sharing, Lending and Defaults: Cross-country Evidence," *Journal of Banking and Finance*, 26, pp. 2017- 2045.

Dorothy Judd v Citibank, 435 NYS, 2d series, pp. 210-2, 107 Misc.2d 526; at *http://ny.findacase.com/research/wfrmDocViewer.aspx/xq/fac.19801103_0045480.NY.htm/qx* (Nov. 3, 1980).

Kapner, Suzanne. 2012. "Banks Unite to Battle Online Theft," *The Wall Street Journal*, p. C1, Jan. 10, 2012.

Khan, A., and J. Hunt. 2012. "UK Online Fraud Report," Cybersource, 2012.

Moore, T., and R. Clayton. 2008. "The Consequence of Non-Cooperation in the Fight Against Against Phishing," Third APWG eCrime Researchers' Summit, Oct. 15-16, 2008.

Moore, Tyler, Jie Han, and Richard Clayton. 2012. "The Postmodern Ponzi Scheme: Empirical Analysis of High-Yield Investment Programs," *Financial Cryptography* 2012.

Murdoch, S., S. Drimer, M. Bond, and R. Anderson. 2010 "Chip and PIN is Broken," at IEEE Symposium on Security and Privacy (2010) pp. 433-444.

Rochet, J.C., and J. Tirole. 2006. "Externalities and Regulation in Card Payment Systems," *Review of Network Economics*, vol. 5, no. 1 (March).

The Economist. 2011. "Not Just Talk," Jan. 27, 2011.

____________. 2012. "The Dream that Failed," March 10, 2012.

Tofel, Kevin C. 2010. "Apple to Flurry: Kiss Our Data Good-bye," gigaom.com, June 2, 2010.

Wikepedia. 2012. "Bitcoin," March 11, 2012.

Wolf, M. 2012. "The World's Hunger for Public Goods," *Financial Times*, Jan. 24, 2012.

Wondracek, G., T. Holz, C. Platzer, E. Kirda, and C. Kruegel. 2010. "Is the Internet for Porn? An Insight into the Online Adult Industry," WEIS 2010.

Zoldi, S. 2012. "Analytic Techniques for Combating Financial Fraud," keynote at Financial Cryptography 2012, Kralendijk, Netherlands Antilles, Feb. 27, 2012.