STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's thesis in economics
Academic year 2016–2017

Blockchains and Coordination Failure in Cross-Border Payment Systems: Overcoming the Penguin Effect

James Finucane (40916)

Abstract: The blockchain has been described as a "foundational" technology that has the potential to fundamentally alter the banking system. Its core feature is the enablement of an electronic peer-to-peer money transfer system without the presence of intermediaries. It is potentially a lower-cost and more efficient alternative to the transaction infrastructure of current payment systems.

In this paper I analyze blockchain's potential impact on remittances, which are an important component of the GDP of many developing countries. Remittances have operated in a suboptimal technology equilibrium, partly as a result of coordination failure. This market failure results in high consumer welfare costs. I use a Farrell and Saloner game-theoretic model to test this claim and demonstrate that, due to lower fixed costs and a lower critical mass than centralized technologies, blockchain reduces the excess inertia that has resulted in the persistence of the network effects that sustain legacy payment systems. As a result, correspondent banks face lower downsides to experimenting with blockchain and greater risks to remaining alone with the legacy technology, including disintermediation.

Although it may have significant advantages over centralized technologies, the potential for a rapid and widespread adoption of a blockchain currency such as Bitcoin presents unknown risks for payment system resilience and financial stability. I present general policy suggestions describing how policymakers might overcome potential risks in the adoption of blockchain technology, in an attempt to help ensure that the system does not trade safety for greater efficiency.

Keywords:	Coordination failure, game theory, network effects, excess inertia
JEL:	D700, L150, O330

Supervisor:	Karl Wärneryd
Date submitted:	May 11, 2017
Date examined:	May 30, 2017
Discussant:	Elsa Horn Af Rantzien
Examiner:	Martina Björkman Nyqvist

Acknowledgments

I am grateful to my supervisors Karl Wärneryd at the Stockholm School of Economics and Kent Smetters at the Wharton School of the University of Pennsylvania for their valuable advice and guidance.

Thanks also to Erik Lenntorp, Björn Segendorf, Gabriel Söderberg, and Albina Soultanaeva at Sweden's Riksbank, who gave me useful ideas and policymaking insights; Jacob de Geer of iZettle, who provided insights into private sector incentives; and Jesper Lindé, who provided me with helpful guidance in thinking about the costs of suboptimal infrastructure.

Lastly, thanks to my family, for everything.

Contents

1 Introduction

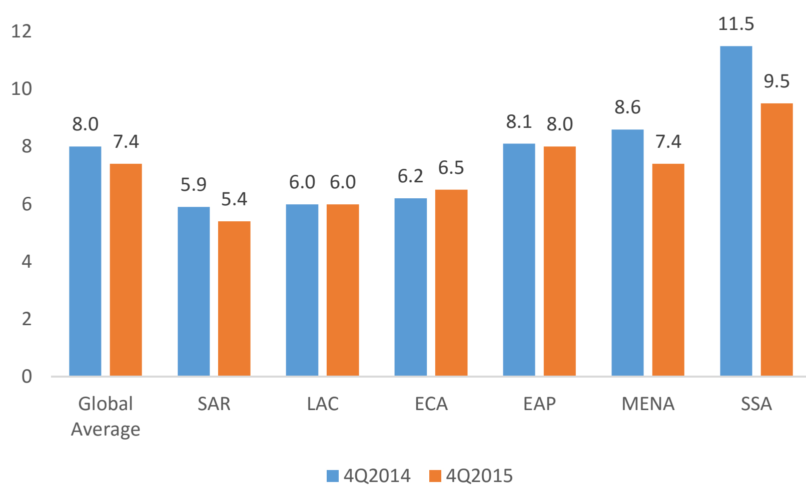
"The global payments industry is undergoing significant and rapid technological change, including […] distributed ledger and blockchain technologies, and as a result we expect new services and technologies to continue to emerge and evolve."

- *Visa 2016 10-K, under "Risk Factors"*

Payments systems, including those used for cross border payments, are a relatively inconspicuous but critical component of financial systems. This theoretical paper focuses on inefficiencies and barriers to technology adoption in cross-border remittances between individuals,[1] which are important for economic growth and development, financial inclusion, and stability. I use a microeconomic model to explain why remittance service providers continue to use costly, unreliable, and outdated (i.e., suboptimal) technology as well as to show how they will shift as a result of the disruptive potential of new technologies now available for peer-to-peer remittances.[2] Global remittances totaled $582 billion in 2015, of which $432 billion went to developing countries. Remittance flows to developing countries are larger than official development assistance and more stable than foreign direct investment and private borrowing. They are an important part of many of these countries' GDPs, for example comprising 10 percent of the Philippines' GDP and over 20 percent of the GDP of eight smaller countries.[3]

The average transaction cost of all remittances was 7.4 percent of the transferred value, corresponding to a total consumer cost of $43 billion, and the average cost to developing countries is higher still. As seen in Figure 1, the destination with the highest average cost of

Figure 1: The average cost of sending $200 is highest in Sub-Saharan Africa



[1] I use remittance to refer to the consumer-to-consumer funds transfer that is the focus of this paper, while I use cross-border payment in the more general sense, encompassing remittances and other cross-border transfers such as business-to-business payments. The other three categories of cross-border payments, as defined in http://www.earthport.com/wp-content/uploads/2014/02/Cross-Border-Payments-Perspectives-A-Glenbrook-Earthport-Research-Brief.pdf, are (1) supplier, or B2B, payments, (2) eCommerce purchasing, and (3) payroll, retirement, and benefits payments.

[2] E.g., the majority of interbank communications use the Society for Worldwide Interbank Financial Telecommunication (SWIFT) network when sending and receiving orders about financial transactions. This network predates the web and has been prone to hacks in recent years.

[3] The countries are Tonga, Tajikistan, Kyrgyz Republic, Moldova, Haiti, Nepal, Liberia, and The Gambia. The source for these statistics is the World Bank's 2016 "Migration and Development Brief 26".

receiving $200 is Sub-Saharan Africa. This region received $35 billion of remittances in 2015 at an average cost of 9.5 percent in the fourth quarter.

The UN Sustainable Development Goals target an average remittance fee below 3 per cent by 2030. A challenge for reaching this is that, at least until recently, remittance service providers have largely avoided competitive and regulatory pressures. Also, as part of their de-risking processes to comply with regulations issued since the 2008 financial crisis, large banks have closed correspondent banking accounts of money transfer operators[4] (MTOs) to limit exposure to money laundering and other financial crimes, increasing remittance costs and flows in rural and remote regions. (World Bank, 2016) In particular, "smaller banks located in jurisdictions perceived to be too risky, are especially affected by the reduction in the number of relationships." (CPMI, 2016)

Compounding the problem of increasing regulatory compliance costs, banks and MTOs have not been adequately incentivized to introduce the major technology changes that would lower costs, leaving the system in a suboptimal equilibrium as a result of coordination failure.[5] Incremental changes, including cloud computing, open protocols, and enhanced data management capabilities, have been introduced which have improved individual financial transactions, (Gifford & Cheng, 2016) but the potential for larger efficiencies exists.

However, as the Committee on Payments and Market Infrastructures[6] (CPMI) reports, correspondent banks[7] "may encounter a substantial coordination problem [with regard to implementing dedicated solutions], as there might be a first-mover disadvantage in implementing some of these measures," (CPMI, 2015). Distributed ledger technology (DLT), a form of which (blockchain) was introduced as a core component of the cryptocurrency Bitcoin in 2008, (Nakamoto, 2008) is an innovation that has the potential to significantly lower costs of remittances. The introduction of DLT and the resulting reduction in excess inertia has spurred research and pilots and might act as a catalyst for innovation/system change. The potential benefits are significant. According to CPMI's February 2017 "DLT in PCS: An analytical framework", DLT has the potential to transform financial services and markets by

i. Reducing complexity;
ii. Improving end-to-end processing speed and thus availability of assets and funds;
iii. Decreasing the need for reconciliation across multiple record-keeping infrastructures;
iv. Increasing transparency and immutability in transaction record keeping;
v. Improving network resilience through distributed data management; and

[4] Money transfer operators are non-bank firms that provide remittance services. The largest are Western Union, MoneyGram, and Ria.
[5] Rent-seeking, inefficient institutions, and underinvestment in R&D and training can each be explained as coordination failure, as described by Hoff and Stiglitz (2001).
[6] Formerly the Committee on Payment and Settlement Systems (CPSS).
[7] Correspondent banking is defined by the CPMI as "an arrangement under which one bank (correspondent) holds deposits owned by other banks (respondents) and provides payment and other services to those respondent banks."

vi. Reducing operational and financial risks.

In the case of correspondent banking, DLT has the potential to reduce or reverse the decline in correspondent banking services, which is critical "given the importance of correspondent banking for international payments. […] in the extreme case, the loss of access to such services can affect the functioning of local banking systems, create financial exclusion and drive some payment flows underground." (CPMI, 2016) Although DLT was partially designed to (and might) disintermediate banks (including correspondent banks), these institutions are trying to figure out how to use it to their advantage. Somewhat ironically, a technology designed to underpin a de-nationalized currency and drive payment flows underground might be used by the very institutions it was created to circumvent in order to maintain their relevance.

DLT also has the potential to introduce disruptive technical, business, and legal changes in cross-border payments. "Numerous firms have identified the slow, indirect, and expensive settlement of cross-border payments as a current point of friction that could be alleviated through the application of DLT arrangements." (Mills, et al., 2016) DLT's ability to reduce the network effects that support the current system might decrease the excess inertia that slows attempts to shift to a better equilibrium. One direct way in which DLT reduces excess inertia is by enabling disintermediation – reducing the number of intermediaries necessary to effect cross-border payments (Mills, et al., 2016) and potentially removing the role of correspondent banks in remittances. This will be good for some entities, such as regional banks, businesses, and consumers, and less so for others, such as correspondent banks and central counterparties. As a result correspondent banks are forced to attempt to innovate, however reluctantly. "The eventual gains will probably flow to customers rather than producers, because that is usually the way with leaps in technology. But the banks know that it is better to be first than to bring up the rear." (The Economist, 2017)

The potential savings that would result in greater consumer welfare are significant. As a result of increased efficiencies, banks estimate that DLT will reduce banks' infrastructure intermediation costs for cross-border payments, securities trading, and regulatory compliance by up to $20 billion annually.[8] This is a significant portion of the total cost to the finance industry of clearing and settling trades, estimated at $65-80 billion per year,[9] and it

> There is variance in the use of the term "distributed ledger technology", reflecting the novelty and nature of the technology. I use the one chosen by the Committee on Payments and Market Infrastructures in their report on the technology. (CPMI, 2017):
>
> "DLT refers to *the processes and related technologies that enable nodes in a network (or arrangement) to securely propose, validate and record state changes (or updates) to a synchronized ledger that is distributed across the network's nodes."*

[8] Santander. The Fintech 2.0 paper: Rebooting financial services.

[9] From Oliver Wyman's 2014 "The Capital Markets Industry," page 8: sum of Clearing, Securities services (Settlement, Custody, Collateral Management), and Post trade data and analytics. Cited in FT's "Big banks plan to coin new digital currency," among other articles. $40 billion goes to custodians (correspondent banks), and $20

does not include potential price cuts due to lower barriers to entry and greater competition. In response to the prospect of cost savings, as well as reduced latency and increased security, large banks have researched and initiated trials of DLT. The most visible examples of this response are industry consortia and pilot projects.[10] As explained by Santander's head of R&D: "Today trading between banks and institutions is difficult, time-consuming and costly, which is why we all have big back offices. [Transitioning to DLT] is about streamlining it and making it more efficient." (Arnold, 2016)

In addition to direct cost savings to service providers from switching to DLT, a portion of which may be passed along to consumers, some argue that there might be an economic benefit as total factor productivity may increase if thousands of former correspondent bank and other financial sector workers find employment in sectors of the real economy.[11]

In this paper I use a model introduced by Joseph Farrell and Garth Saloner in 1985 to model the coordination failure evident in banking infrastructure investment. While others, such as the CPMI, have noted this coordination failure, my contribution is to model it and also show how the introduction of a new technology (in this case DLT) might sufficiently shift the equilibrium to change the strategic behavior of various firms.

The rest of the paper is structured as follows: Section 2 provides a coordination failure literature review as well as an overview of cross-border payments infrastructure, Section 3 summarizes DLT in order to explain why this technology might change the status quo, Section 4 applies the Farrell and Saloner model to remittances, Section 5 describes some costs of the coordination failure, Section 6 provides some policy suggestions based on the model results, and Section 7 concludes and provides directions for further research.

2 Coordination failure literature and cross-border payments infrastructure

When technologies exhibit increasing returns based on number of users (network effects), multiple outcomes and equilibria are often possible. In many cases, inertia and path dependence result in the persistence of an inefficient equilibrium. The ubiquitous QWERTY keyboard is perhaps the best known example of a suboptimal standardization equilibrium. David (1985) describes why QWERTY is the standard despite earlier efforts to switch to the rival DSK (Dvorak Simplified Keyboard), which apparently allows users to type 20-40 percent faster. The US government could have mandated that

---

billion goes to third parties including data & technology providers for post trade data and analytics. Note that these data are not specific to remittances; I use them to give a sense for the potential impact of the technology.

[10] The best known consortium is R3CEV, and an example of a high-profile trial project is the "utility settlement coin" being tested by UBS, Deutsche Bank, Santander, BNY Mellon, and ICAP.

[11] This argument is, however, based on the optimistic assumptions that there will be more productive jobs available and that the employees liberated would be able to fill them.

companies use the more efficient standard, but the National Bureau of Standards did not write interface standards for the computer industry because it did not want to impede innovation (Hemenway, 1975).

However, the QWERTY keyboard is also an example of the desire to "construct" coordination problems where they do not exist. (Liebowitz & Margolis, 1990) argue that the QWERTY keyboard story is false, and that many other claims of coordination problems are of "a highly conjectural nature." They state that *"it is essential that the literature present real examples of demonstrable market failure if the concept of network externality is to have any relevance."* In addition, (Liebowitz & Margolis, 1994) attempt to limit the implied scope of network externality, writing that "While network effects are common and important, network externalities as market failures, we will argue, are theoretically fragile and empirically undocumented." In this paper I attempt to show that the high costs of remittances are the result of a market failure.

(Oren & Smith, 1981) research positive demand externalities in communications services, where users derive a mutual benefit from being connected to other users. (Katz & Shapiro, 1985) extend Oren and Smith's monopoly model and develop an oligopoly model with network externalities. They analyze the social and private incentives for firms to produce compatible products and find evidence that public policy might have an important impact in certain instances. For example, a dominant firm may choose to remain incompatible with a rival since it will lose market share if it becomes compatible. As an example, correspondent banks are less likely to fully support a move to a peer-to-peer cross-border payment system, as this shift would decrease or remove a source of revenue. In many cases it is profitable for dominant firms to remain incompatible with rivals: as (Arthur, 1989) finds, technologies that gain an early lead might "lock out" later technologies, regardless of which one is superior. Arthur extends a simple model of increasing returns with adoption by introducing an expectations case, where agents' returns are affected by future agents' choices. This happens in the case of *standards*, which are important for financial services firms that must adapt to technical requirements designed by other firms and regulatory standards set by regulators. This situation contains parallels to correspondent banking decisions, where "individual actors may face a considerable degree of uncertainty and high investment costs with regard to implementing dedicated solutions." (CPMI, 2015)

(Farrell & Saloner, 1985) provide examples of the benefits of standardization, which enable competing television channels to broadcast on the same set, subscribers of different telephone companies to call one another, and computer programs from rival organizations to generally (but not always) communicate with each other. They produce a model that describes situations in which standardization benefits can "trap" an industry in an inferior standard when a better alternative is available. This trap does not occur in situations with complete information and identical preferences among firms, but when information is incomplete, *excess inertia* can occur. Two types of excess inertia occur. In the first type ("symmetric inertia"), the firms unanimously prefer the new technology but do not change, thus maintaining the *status* quo. In the second ("asymmetric inertia"), the firms differ in their technological preferences, but the total benefits from the switch would exceed the total costs. *This situation is similar to what we see in financial services infrastructure, where* *"the specific interests of banks of different sizes or which specialize in different types of business may not coincide, making it difficult to agree on a detailed design."* (CPSS, 2005)

Additionally, Farrell and Saloner explore how converters between otherwise incompatible technologies can hinder the shift to a more efficient equilibrium. They "find that the existence of converters can actually lead to less compatibility than would occur in their absence – and that this happens precisely when compatibility is most important!" (Farrell & Saloner, 1992) Correspondent banks might thus be considered converters that enable, at a cost, the continuing use of conflicting technologies and thus hinder the shift to a more efficient equilibrium. Additionally, the correspondent banking industry is highly concentrated: the four largest account for more than half of the $170 trillion in assets under administration. This concentration reduces innovation incentives: "while all firms including monopolies have incentives to adopt cost reducing innovations, the relative strength of this incentive is greater in competitive markets." (Milne, 2005)

Other papers that present explanations for inefficient equilibria include (Banerjee, 1992) and (Bikhchandani, Hirshleifer, & Welch, 1992). These differ from Farrell and Saloner's papers in that they do not assume "strong complementarities", i.e., they find that inefficient equilibria can be reached even in situations where network effects are not present. Banerjee explains that a "herd externality" created by this "clustering behavior," and (Bikhchandani, Hirshleifer, & Welch, 1992) explain this "uniform social behavior" as a result of "informational cascades" – situations in which "it is optimal for an individual, having observed the actions of those ahead of him, to follow the behavior of the preceding individual without regard to his own information" (prior probabilities in the form of signals are transformed into posterior probabilities based on what others have done). (Bikhchandani, Hirshleifer, & Welch, 1992) emphasize that cascades are vulnerable to shocks and thus fragile: they explain not only uniform behavior but also instances of drastic change such as fads.

2.1 Coordination failure in remittances

Financial services for cross border payments are significantly more expensive than domestic payment services. As the SEPA project, which encompasses all EU member states indicates, the cost of currency exchange does not justify the cost difference. Moving to a more efficient equilibrium, however, is difficult. "Intuitively, it is plausible that the industry, once firmly bound together by the benefits of compatibility or standardization, will be inclined to move extremely reluctantly to a new and better standard because of the coordination problems involved." (Farrell & Saloner, 1985)

Groups such as the Committee on Payments and Markets Infrastructures have been formed with the goal of improving payments infrastructures.[12] The CPMI's mandate is to promote "the safety and efficiency of

[12] The CPMI, formerly the Committee on Payment and Settlement Systems (CPSS) (the name change occurred in 2014 with an updated mandate), traces its roots to the Group of Experts on Payment Systems, formed in 1980 by the Governors of the G10 countries. The CPSS was set up as a permanent central bank committee by the G10 Governors in 1990.

payment, clearing, settlement and related arrangements" improving regulation, policy, and related practices worldwide. Farrell and Saloner model such committees as a many-period game where players meet and talk repeatedly to agree on a joint action. Their game has the basic structure of a "battle-of-the-sexes" game where the overall game is a "war of attrition." (Farrell & Saloner, 1988) However, although we can sometimes nominate "government" as a universal authority figure, "even that breaks down in international problems," (Farrell & Saloner, 1988) as is the case with cross-border payments.

Where there are significant legal, technical, operational, regulatory and other barriers to adopting a new technology standard firms might be too reluctant to switch. This "excess inertia" can occur when early adopters bear a disproportionate share of transient incompatibility costs. Farrell and Saloner call it the *penguin effect*. "Penguins who must enter the water to find food often delay doing so because they fear the presence of predators. Each would prefer some other penguin to test the waters first." (Farrell & Saloner, 1986) Additionally, the existence of correspondent banks acting as "converters", as in the case of cross-border payments, might decrease incentives to standardize, thus reducing "the realized degree of compatibility." (Farrell & Saloner, 1992)

3 Distributed ledgers (and blockchains)

The purpose of this section is to provide support for why DLT (*Y*) is a lower-cost and more efficient alternative to the current state (*X)* and explain how it has created a shift in incentives for firms deciding whether to adopt a new technology.

The best-known type of DLT, the blockchain, has been described as a *foundational* technology (rather than merely disruptive) since it "has the potential to create new foundations for our economic and social systems." (Iansiti & Lakhani, 2017). It has been likened to the internet and will challenge financial firms in a manner similar to how the internet disrupted media companies and advertising firms. "Such a fundamental restructuring of a core part of the economy is a big challenge to incumbent firms that make their living from it. Preparing for these changes means investing in research and experimentation." (Ito, Narula, & Ali, 2017)

Blockchain was the main technical innovation of Bitcoin, introduced in 2008 by an unknown group of pseudonymous developers (Nakamoto, 2008). Bitcoin's (and blockchain's) novelty involves melding well-established technologies from the fields of software engineering, cryptography science, and game theory. Its fundamental element is the enablement of an electronic peer-to-peer money transfer system, without the presence of intermediaries. Bitcoin removes the need for intermediaries by using digital signatures and, critically, solving the "double-spend" problem. To eliminate double spending, the Bitcoin network "timestamps transactions by hashing them into an ongoing chain of hash-based proof-of-work, forming a record that cannot be changed without redoing the proof-of-work." (Nakamoto, 2008) In other words, blockchain is a shared public database that permanently records transactions after verifying them using a protocol.[13] Transactions are added to a "block" once verified, and that block is then added to a "chain" (history) of transactions.

Blockchain, and DLT more generally, might reduce excess inertia maintaining inefficient legacy technologies in the financial system because it has a lower critical mass of other participating banks and a higher stand-alone value than earlier alternatives. (Keser, Suleymanova, & Wey, 2012) show that the technology with a lower critical mass is risk-dominant and chosen by the maximin criterion. In the paper they use the critical masses and stand-alone values of technologies as proxies for relative riskiness.

While DLT will continue to be researched, developed, and probably adopted within financial services, large obstacles remain before the technology gains widespread acceptance. Current problems include low awareness and understanding of the technology, but DLT also "has genuine drawbacks." (Pauget, 2016) The primary challenges once the benefits of blockchain become more widely known will relate to business processes, regulatory/legal hurdles, and technical shortcomings.

In the short term, technical shortcomings are the most salient – encompassing questions of scalability, privacy, security, and an underdeveloped system infrastructure. Comparisons are made to the Internet, such as the observation that we are still "designing and refining the Internet's own scalability." (Mougayar, 2016) However, what complicates attempts to scale DLT is that it is the first such attempt for a "decentralized network with an economic model that is tied to its security." (Mougayar, 2016) Although DLT provides a potential opportunity to replace clearing and settlement systems with a faster, cheaper, and more secure technology, it remains unclear whether it will be an improvement over existing, centralized alternatives, and for which specific use cases. For example, Bitcoin can only handle seven transactions per second, compared to thousands for Visa.

Business obstacles include a dearth of qualified/knowledgeable individuals,[14] the sheer scale of projects that would be required to move assets to the blockchain, and the need to reach a critical mass of users (different from the critical mass of banks that must adopt it, which is lower, for example for a system such as Bitcoin's which requires no banks to function as intermediaries). Compliance requirements and unclear regulations are a potential hurdle, although DLT may eventually provide a regulatory and compliance benefit.[15]

It seems likely, however, that the technical and operational issues preventing robust, wide-scale implementation of DLT will be solved. Coordination problems will pose the greater difficulty. "Although technology has been one contributing factor in determining the design of a particular PCS arrangement, the fundamental need for coordination has often required joint action through new or existing legal

[13] A protocol is a set of automated rules. An example is SMTP, the protocol used for sending email.
[14] 5,000 developers who have worked with blockchain vs. about 9 million with Java/JavaScript.
[15] For example, in a May 24, 2016 report "Blockchain: Putting Theory into Practice", Goldman Sachs estimates $3-$5 billion in savings for anti-money laundering and "know your customer" compliance, in addition to up to $12 billion in savings from streamlined clearing and settlement of cash securities.

entities to, at a minimum, provide organization and governance." (Mills, et al., 2016) As a result, banks and fintech firms are working on solutions together: "Many models may eliminate some roles of current intermediaries in payments, clearing, and settlement but may not necessarily eliminate the need for coordination or centralization of central functions by trusted intermediaries. These trusted intermediaries could still be needed to play important roles in addressing frictions beyond what DLT may be able to accomplish or may be able to use DLT arrangements to improve or evolve how they accomplish their respective missions." (Mills, et al., 2016)

The challenges to near-term large-scale adoption do not preclude continued investment in DLT and a gradual shift to the new technology. As Bill Gates has mentioned, "We always overestimate the change that will occur in the next two years and underestimate the change that will occur in the next ten." The fact that blockchain is open source and peer-to-peer simply lowers the costs of experimenting and the critical mass of participating banks necessary, which might explain why increasing amounts of investment have been directed toward firms that use the technology. As Visa writes in its 2016 10-K filing, "Technolgoy and innovation is shifting consumer habits and driving growth opportunities in ecommerce, mobile payments, block chain technology and digital currencies. These advances are enabling new entrants, many of which depart from traditional network payment models."

The model presented below provides an explanation for why the nature of DLT is increasing correspondent banks' willingness to experiment with the new technology: the downside risk of failed experiments with the technology is (initially) inexpensive, and the probability of a bank being disintermediated if it fails to keep up with the new technology is much greater.

4 The model

I use a model presented by Joseph Farrell and Garth Saloner in a 1985 paper to provide an explanation for why excess inertia helps explain the persistence of inferior technology and processes in cross-border payments services.[16] I selected this model for its focus on complementarities: a bank that cannot transact with other banks is very limited in the services it can offer clients; therefore using technologies/systems that are compatible with other banks is critical.

I begin with an introduction to Farrell and Saloner's model with complete information (with a simple example of how it might apply to cross-border payments infrastructures), before switching to the more interesting and relevant incomplete-information case. In both cases I make adjustments to reflect the situation in remittances where we have a group of long-established incumbents using a dated system and new entrants attempting to use digital technologies such as DLT to establish more efficient processes and challenge the incumbents.

[16] I change some inequalities from strict to weak to show that there might come a time when B(1,Y) is equivalent to B(2,Y), but otherwise the model I use is the same as Farrell and Saloner's original.

In the incomplete-information case, I model two types of excess inertia – symmetric and asymmetric – and describe how DLT might have shifted critical levels of model parameters. In the case of symmetric excess inertia, I model a decrease in type required to switch to a new technology for a firm that is choosing whether to 1) switch immediately or 2) switch only if the other firm switches (two correspondent banks play this game). In this case i* decreases via an increase in B(1,Y). In the case of asymmetric excess inertia, I model a decrease in the level required to potentially switch to a new technology for a correspondent bank that is choosing whether to 1) switch only if the other firm switches or 2) not switch at all (the correspondent bank plays this game against a money transfer operator). In this case i_bar decreases via a decrease in B(1,X).

Before the introduction of DLT, firms decide to not switch to a new technology. However, the introduction of DLT decreases the risk of a unilateral switch to the new technology (e.g., through lower fixed costs) and increases the risk of remaining alone with the old technology (e.g., peer-to-peer technology could circumvent the correspondent banks). I argue that DLT reduces the firm "type" required to switch, thus incentivizing correspondent banks to innovate.

4.1 A model with sequential decisions and complete information

In the first Farrell and Saloner model firms' decisions are taken sequentially and payoffs are common knowledge. If, allowing for transition costs, all firms would prefer the industry to switch, then the only subgame perfect equilibrium is that they all do. Let $N = \{1,2, \dots, n\}$ denote the set of firms in the industry. For any $j \in N$ and any $S \subseteq N$ containing j, $B_j(S, Y)$ is defined as the net benefit to firm j from switching, together with the other firms in S, from the old standard (X) to the new one (Y), relative to its benefit if all firms stick with X. By normalizing so that each firm gets zero benefit in the *status quo*, we have $B_j(N, X) = 0$.

Figures 2 and 3 represent standards *X* and *Y*, respectively.

Figure 2: Technology X: Cross-border payment using traditional correspondent banking network[17]

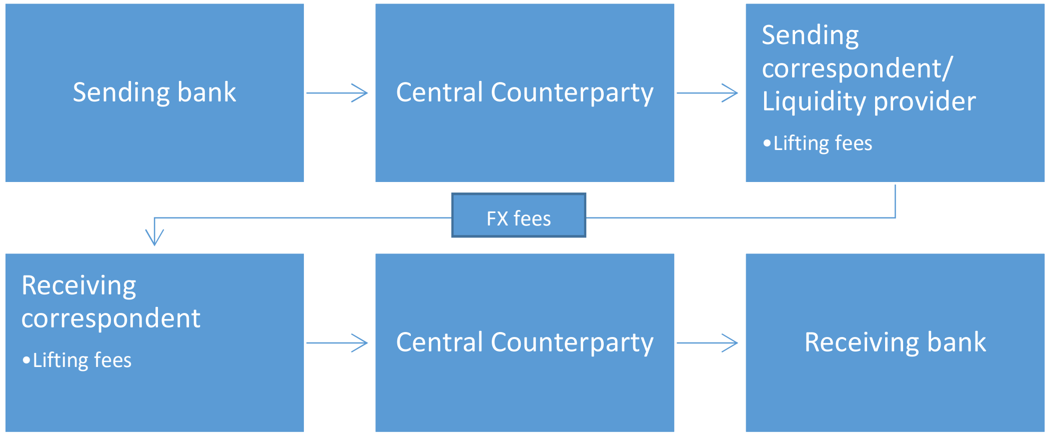
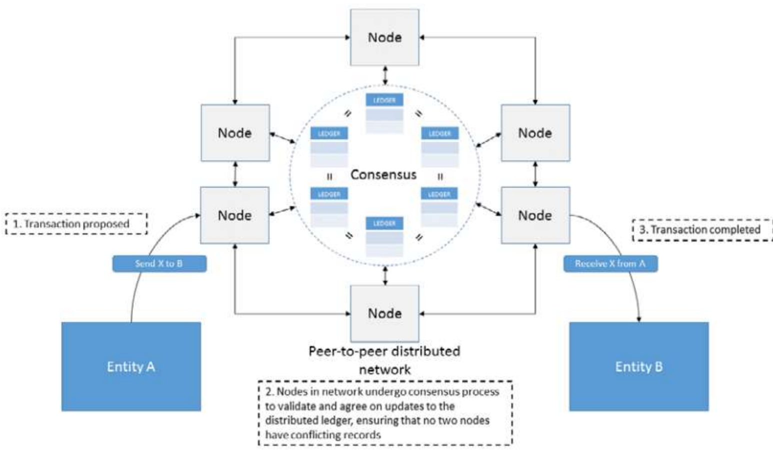


Figure 3: Technology Y: Cross-border payment using distributed ledger protocol[18]



$B_j(S, Y)$ is the value to j of switching and having the other members of S switch. This is a present value, net of any transition costs. Thus, firm j would favor a change by the entire industry if and only if $B_j(N, Y) > 0$.

Also define $B_j(S, X)$ for subsets S containing j, as j's payoff if j and the other members of S stay with X, while the members of $N \setminus S$ switch to Y.

[17] This figure is based on a chart in (Gifford & Cheng, 2016). The arrows represent funds moving from sending to receiving bank. The "process is sequential and fragmented. Multiple steps involving manual processes due to the lack of interoperability between payment systems add delay, cost and risk to the transaction." A lifting fee is defined as "the transaction fee charged to the recipient of cash transferred through a wire transfer, which the recipient's bank or an intermediary bank imposes for handling the transaction."

[18] This image is taken from (CPMI, 2017). (Gifford & Cheng, 2016) explain that in a system such as this, payments "are either fully and simultaneously settled in real-time or they do not occur at all – a process called atomic payments. This eliminates or reduces many of the risks that plague today's reliance on intermediaries for cross-border payments, including credit and operational risk."

The fundamental assumption of positive network effects is as follows:

Assumption 1. If $j \in S \subseteq S'$ and $k = X$ or Y, then $B_j(S,k) \leq B_j(S',k)$.

I.e., whatever j's choice, he prefers that others make the same choice. This dynamic introduces the coordination considerations.

- **Symmetric case.** When assuming that $B_j(S,X)$ depends only on the number of firms in S, and likewise for $B_j(S,Y)$, we can write the benefit functions as $B_j(m,k)$, where *m* is the total *number* of firms in *S*, i.e., the number making the choice that *j* makes. Moreover, we then assume the function $B_j(\cdot,\cdot)$ is the same for all *j* and can thus drop the *j* subscript.
- **The model.** The set *N* of firms is given, as are the alternative standards *X* and *Y*. All firms are initially at standard *X*. There are *n* periods in this game, which has perfect and complete information. Since one firm has a decision each period, the number of periods is equal to the number of firms. In period *j*, firm *j* decides whether to switch to *Y*. If *S* denotes the set of firms that do switch, then the payoffs are
    - $B_j(S,Y)$ for $j \in S$
    - $B_j(N \setminus S, X)$ for $j \notin S$

Proposition 1. Suppose that, for each *j*,

$$B_j(N,Y) > B_j(\{j, j+1, \dots, n\}, X) \tag{1}$$

Then the unique subgame perfect equilibrium involves all firms switching.

Proof. The condition (1) ensures that, for each *j*, if $1, \dots, j-1$ have already switched, then *j* prefers to switch (if he believes all the rest would follow) rather than to stay (*whatever* his beliefs about how many others would switch). Since *j* knows this is true for $j+1, \dots, n$, he knows they will switch if he does; and so he will switch.

Proposition 1 does not use Assumption 1. Using Assumption 1 yields the following result.

Corollary. If

$$B_j(N,Y) > B_j(N,X) \text{ for all } j, \tag{2}$$

Then the unique subgame perfect equilibrium involves all firms' switching. Therefore, *in this model there can be no excess inertia in the symmetric sense that each firm prefers an overall industry switch but it fails to happen.*

Condition (1) is weaker than unanimity (2), however. So Proposition 1 tells us that players *j*, late in the game, sometimes switch, even though $B_j(N,Y) < B_j(N,X)$. Moreover, it is clear that there is no necessary relationship between $\sum_j [B_j(N,Y) - B_j(N,X)]$ and the outcome of the game: we can find excess inertia or its opposite, excess momentum, if we make judgments based on adding benefits.

Being late in the game is a strategic disadvantage because of the assumption that each agent has only one chance to choose his standard; thus, early movers are able to commit. In a game of complete information, there is no countervailing value to waiting to see how things evolve. This is expressed by the following result.

Proposition 2. Given the preferences of all agents, each agent is better off (not necessarily strictly) moving earlier than later ("the New Hampshire theorem").

Proposition 2 is proved in the Appendix using only Assumption 1 → the presence of network externalities. The essence of the proof is that having an earlier position gives power over later movers, and hence even earlier movers are obliged to treat one's preferences with more respect.

Intuitively there is a benefit of commitment from moving early. In a general game there can be a countervailing "regret" factor: once a von Stackelberg follower has moved, the leader would like to change his move if possible. In this game that does not happen as every sequential equilibrium would also be an equilibrium if firms decided simultaneously on their choices. The other factor that sometimes makes it desirable to move later in certain games, information flowing in, is also absent from this model, but is addressed in the following model with incomplete information.

A simple example in which Proposition 2 holds strictly is provided by the following two-firm case:

Firm A

	$B_A(m,X)$	$B_A(m,Y)$
$m = 1$	-2	-1
$m = 2$	0	1

Firm B

	$B_B(m,X)$	$B_B(m,Y)$
$m = 1$	-2	-3
$m = 2$	0	-1

If *A* moves first, then he will switch and *B* will follow. If *B* moves first, however, he will not switch, and *A* will then not switch. It is easy to check the claim of Proposition 2 that each firm prefers the outcome that results from its moving first.

There are instances, however, when Firm A will move first and B will not follow, as shown below:

Firm A: Composite incumbent

	$B_A(m, X)$	$B_A(m, Y)$
$m = 1$	-2	-5
$m = 2$	0	-4

Firm B: Composite new entrant

	$B_B(m, X)$	$B_B(m, Y)$
$m = 1$	-2	1
$m = 2$	-1	2

Sum of A's choice and B's choice:

- Regardless of who moves first, the outcome is $-2 + 1 = -1$
- If both had selected the old technology, $0 - 1 = -1$
- If both had selected the new, $-4 + 2 = -2$

Note that in this case firms are worse off in the aggregate with a move to the new technology, representing an increase in consumer surplus.

In these tables, *X* represents using the traditional correspondent banking system to conduct money transfer operations (as the traditional MTOs do currently), *Y* represents conducting money transfer operations using peer-to-peer technology (as, for example, Abra and TransferWise do)

Firm A, a composite of the existing incumbent firms (Western Union, MoneyGram, and Ria) that currently dominate the market, wants to maintain the status quo. Firm B, a composite of new entrants (e.g., Abra and TransferWise), benefits from using new technology but would benefit more from widespread acceptance (e.g., regulatory, legal, reputational support) of an efficient cross-border fiat currency settlement system.

If Firm A continues to operate on the current correspondent banking system along with its competitors, it continues to earn revenues and profits in line with historical experience, so $B_A(2, X) = 0.$ However, as new firms enter and adopt peer-to-peer technology, Firm A slowly loses customers, corresponding to $B_A(1, X) = -2.$ Firm A will never adopt peer-to-peer technology because if it does, consumers may benefit but A will face a steep decline in revenues and profits, corresponding to $B_A(1, Y) = -5$ or $B_A(2, Y) = -4.$ These firms will not benefit from technological innovation; maintaining the status quo is ideal.

For new entrants, however, there is no point in adopting the old technology. Money transfer services is a saturated industry, and setting up a firm to compete with Western Union, MoneyGram, and Ria using their business and operational models will result in an investment loss, regardless of other firms' technological choices. This outcome is represented by $B_B(1, X) = -2$ and $B_B(2, X) = -1$. I assume that new entrants must invest 1 unit, and if they provide services using new technology their net return is either $B_B(1, Y) = 1$ (taking 2 of Firm A's profits, corresponding to $B_A(1, X) = -2$) or $B_B(2, Y) = 2$, taking 3 of Firm A's profits, with consumers benefitting with one unit.

Consumers benefit if a change to the new technology happens: "because of emulation and competition, innovators capture only a small part of the benefit of their innovation."[19]

4.2 A model with incomplete information

Recognizing that the complete knowledge assumption is somewhat unrealistic (as is the unanimity one), Farrell and Saloner then shift focus to the more relevant and interesting incomplete-information models. In these models incomplete information about the other firm's preferences is used as a proxy for uncertainty about whether a firm's technological choice would be followed if it switched.

They analyze a model with incomplete information about the eagerness of each firm to switch to the new technology. The resulting equilibria resemble bandwagons, where firms that strongly favor the change switch early and hope that the remaining firms at least moderately favor the change and will join the bandwagon if it gets rolling. One of two types of excess inertia arises in the model with incomplete information. The first, symmetric inertia, occurs when firms are unanimous in their preferences for the new technology but do not make the change since no firm is sufficiently motivated to start the bandwagon rolling (this might be the scenario if, e.g., both players are correspondent banks or MTOs). This is the first type of excess inertia I model, in which both players as correspondent banks.

In the second type of excess inertia, asymmetric inertia, firms have heterogeneous technological preferences, but the total benefits of switching exceed the total costs. As in the case of symmetric inertia, those in favor are not sufficiently in favor to start the bandwagon rolling. Symmetric inertia is purely a problem of coordination, and as a result nonbinding communication of preferences and intentions may eliminate the inertia (Farrell, 1982). If this was the case in cross border financial services, committees such as the CPMI and indications of intentions such as the "utility settlement coin"[20] and R3CEV consortium might solve these problems quite quickly. However, *the cross border financial services infrastructure resembles a case of asymmetric inertia, with institutions such as correspondent banks and central counterparty clearing houses benefiting from*

[19] Larry Summers, Washington Post, https://www.washingtonpost.com/opinions/picking-on-robots-wont-deal-with-job-destruction/2017/03/05/32091f08-004b-11e7-8ebe-6e0dbe4f2bca_story.html.

[20] Four banks (UBS, Deutsche, Santander, and BNY Mellon) and a broker (ICAP) are developing a form of digital cash to clear and settle financial trades over blockchain. The move "is one of the most concrete examples of banks co-operating on a specific blockchain technology to harness the power of decentralized computer networks and improve the efficiency of financial market plumbing." https://www.ft.com/content/1a962c16-6952-11e6-ae5b-a7cc5dd5a28c.

the inefficient/non peer-to-peer technology, and entities such as regional banks, businesses, and consumers benefiting from a more efficient payments system that would allow peer to peer payments (or more efficient payments generally). Additionally, firms do not have complete information about other firms' preferences, leading to uncertainty about whether they will be followed after switching to a new technology.

Farrell and Saloner continue to allow for endogenous timing of moves, but find that in conjunction with the incomplete information assumption this yields a richer set of possibilities than Proposition 3 would suggest.

4.2.1 The games' setup

Since we are explicit about incomplete information and differences among firms, we can write the benefit function as $B^i(\cdot,\cdot)$, where i denotes a firm's type. There is no need for subscripts because differences are captured in different values of i, higher values of which are taken to indicate stronger preference for change to DLT (e.g., a correspondent bank would have a lower i value than a startup since it probably has less to gain and more to lose from a shift to DLT).

The lost revenues of correspondent banks, central clearinghouses, and technology providers, and the costs saved by regional banks would make their i' (level of preference for change to a new technology) different. I use this model because it might be a fair simplification of the financial services infrastructure, where new systems undergo thorough testing (e.g., regulatory hurdles in addition to technical requirements) and firms' preferences for standards differ. Notably, "most of the costs involved in correspondent banking arise not from the actual payments processing but from compliance and IT work on system modifications." (CPMI, 2015) Additionally, banks that specialize in the for-profit provision of correspondent banking services (as opposed to banks that maintain these services to support the cross-selling of other products) might prefer an inefficient technological equilibrium to a technology such as DLT, which circumvents their core function. As a result, this model might be a useful framework for analyzing the recent heightened level of activity and investment in financial technology, which began with digitalization of services and accelerated with the hype surrounding distributed ledger technology.

I assume that there are two periods (1 and 2) and each firm can switch at time 1, 2, or never. I rule out re-switching (and based on Footnote 9 in (Farrell & Saloner, Standardization, compatibility, and innovation, 1985), such switching would never occur[21]).

S denotes the action "switch" and *D* denotes "do not switch," so a strategy for player j is described by

- $\sigma_1^j: [0,1] \rightarrow \{S, D\}$,

[21] Assuming that reswitching might occur, a sufficient condition is $\frac{i^*-\bar{\iota}}{i^*}B^{i^*}(2,Y) + \frac{\bar{\iota}}{i^*}B^{i^*}(1,Y) \geq \frac{i^*-\bar{\iota}}{i^*}B^{i^*}(1,X) + \frac{\bar{\iota}}{i^*}B^{i^*}(2,X)$. Using the definition of i* and the fact that $B^{i^*}(2,X) = 0 \geq B^{i^*}(1,X)$, this condition is always satisfied. Farrell and Saloner note that they "do not fully understand this remarkable conclusion."

- $\sigma_2^j: [0,1] \times \{S, D\} \to \{S, D\}$,

i.e., the second-period move is based on the player's own type and the opponents' first-period moves.

Here σ_t describes the strategy for period *t* and maps the set of player types and history of play to date into the possible actions the firm can take. (Strictly speaking σ_t^j should be conditioned also on whether player j switched at time $t-1$. A player who did switch at time $t-1$ has no further decisions to make, however, and hence σ_t^j can be simplified as above without ambiguity.)

Assumptions are as follows:

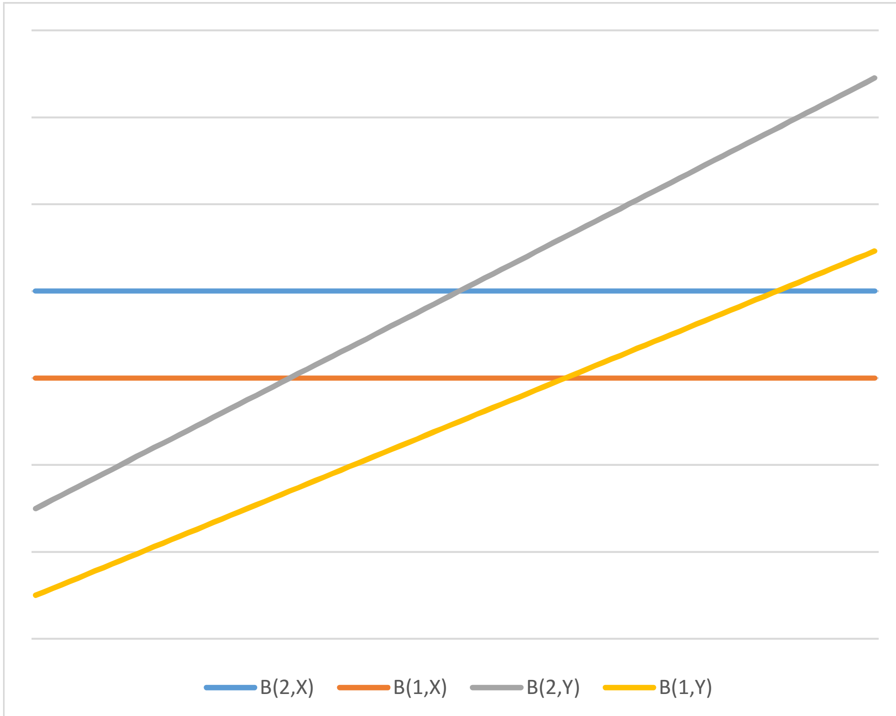
Assumption 1: Networks are beneficial. As noted above, this is the basic assumption of positive network effects:

$$B^i(2,k) > B^i(1,k), \qquad k = X \text{ and } Y.$$

Assumption 2: Higher "types" (higher i) are more eager to switch to the new technology, both unilaterally and if the other firm also switches:

$B^i(2,Y)$ and $B^i(1,Y)$ are continuous and strictly increasing in i.

Figure 4: Examples of benefit functions that satisfy the assumptions of the model



Assumption 3: Unilateral switching is worthwhile for at least one possible type of firm, and (at the other end of the spectrum) there are some firms who would rather remain alone with the old technology than join the other firms with the new:

$$B^1(1, Y) > B^1(2, X) = 0 \text{ and } B^0(2, Y) < B^0(1, X).$$

Note that as mentioned earlier $B^i(2, X) = 0$ (at every level of i) by normalization.

Assumption 4: If a firm of type i' prefers a combined switch to Y to remaining alone with technology X, then so do all firms with $i > i'$, i.e., if i' would follow a lead, then so would $i > i'$:

$$B^i(2, Y) - B^i(1, X) \text{ is monotone in } i.$$

Farrell and Saloner make an analogy to the political "bandwagon" effect. They describe how politicians will be concerned not only with how strongly they feel about an issue, but the likelihood that their stand will become the majority view. Vigorous opponents and staunch supporters will stake their positions without waiting to see whether their view will become popular, but a more "political" middle group may wait a while to test the political waters ("penguin effect"), declaring themselves to be "for" the measure if the bandwagon begins to roll and "against" otherwise. The situation is similar in banking infrastructure, where following the "herd" is generally considered to be the safe and practical move although there are exceptions.

A *bandwagon strategy* is defined by a pair $(i^*, \bar{\imath})$, with $i^* > \bar{\imath}$ such that:

i. if $i \geq i^*$, the firm switches at time 1;
ii. if $i^* > i \geq \bar{\imath}$, the firm does not switch at time 1 and switches at time two if and only if the other firm switched at time 1; and
iii. if $i < \bar{\imath}$, the firm never switches.

A *bandwagon equilibrium* is defined as a perfect Bayesian Nash equilibrium in which each firm plays a bandwagon strategy. (The specifications of the model I use focus on symmetric bandwagon equilibria, i.e., $(i^*, \bar{\imath})$ same for each player.)

Farrell and Saloner then show that a unique symmetric bandwagon equilibrium exists and that there are no equilibria that are not bandwagon equilibria.

First, let $\bar{\imath}$ be defined by $B^{\bar{\imath}}(1, X) = B^{\bar{\imath}}(2, Y)$, so any firm with type $i < \bar{\imath}$ would prefer remaining with the old technology to switching to the new, even if the other firm switched. On the other hand, a firm with $i > \bar{\imath}$ would switch in the second period if the other firm had already switched (assuming that switching back is precluded). This describes behavior in the second period, and using this we now analyze the first period:

Define $f(i) = iB^i(2, Y) - \bar{\imath}[B^i(2, Y) - B^i(1, Y)]$. Let $I = \{i: f(i) = 0\}$.

This function can be explained as follows:

For a player of type $i > \bar{\imath}$, the two first-period actions are either (a_1) switch at time 1 or (a_2) switch at time 2 if the other player switched at time 1. At the point of indifference i^* between these two actions, the utilities are the same. Note that expected utilities are derived using probabilities assuming that firm "types" are uniformly distributed on the interval $(0,1)$:

$$u^i(a_1) = B^i(2, Y)(1 - \bar{\imath}) + B^i(1, Y)\bar{\imath}$$

$$u^i(a_2) = B^i(2, X)i^* + (1 - i^*)B^i(2, Y) = (1 - i^*)B^i(2, Y)$$

$$u^i(a_1) - u^i(a_2) = i^*B^i(2, Y) - \bar{\imath}\left[B^i(2, Y) - B^i(1, Y)\right] = f(i)$$

Lemma 1:

(a) $f(i) < 0 \; \forall \; i < \bar{\imath}$; [I switch $i < \bar{\imath}$ to a strict inequality to apply to cases where B1 curve = B2 curve, e.g., when "going it alone" and building an app using peer-to-peer technology is just as profitable as choosing the same technology that other firms pick (and maybe even more so, if the p2p technology has a broad enough reach when used alone; there may even be a negative network effect as more competitors use it)]

(b) $f(i)$ is strictly increasing in $i \; \forall \; i > \bar{\imath}$;

(c) $f(1) > 0$;

(d) I contains exactly one point (which we call i^*), and $i^* \in (\bar{\imath}, 1)$.

Proof:

(a) For $i \leq \bar{\imath}$, $\bar{\imath}B^i(2,Y) \geq iB^i(2,Y) \geq iB^i(2,Y)$. Also $\bar{\imath}B^i(1,Y) < \bar{\imath}B^{\bar{\imath}}(1,Y) < \bar{\imath}B^{\bar{\imath}}(2,Y) \equiv \bar{\imath}B^{\bar{\imath}}(1,X) < B^{\bar{\imath}}(2,X) \equiv 0$. So $iB^i(2,Y) - \bar{\imath}B^i(2,Y) + \bar{\imath}B^i(1,Y) < 0 \; \forall \; i \leq \bar{\imath}$.

(b) Immediate since $(i - \bar{\imath}) > 0$ for $i > \bar{\imath}$ and since $B^i(2,Y)$ and $B^i(1,Y)$ are strictly increasing.

(c) $f(1) = B^1(2,Y)[1 - \bar{\imath}] + \bar{\imath}B^1(1,Y)$. But $\bar{\imath} < 1$ (since $B^1 > 0$) and $B^1(2,Y) \geq B^1(1,Y) > 0$

(d) Since $f(i)$ is strictly increasing and continuous on $(\bar{\imath}, 1]$ with $f(\bar{\imath}) < 0$ and $f(1) > 0$, there exists exactly one $\bar{\imath} < i^* < 1$ for which $f(i^*) = 0$.

Lemma 2:

(a) $B^{i^*}(1,Y) \leq 0$ and

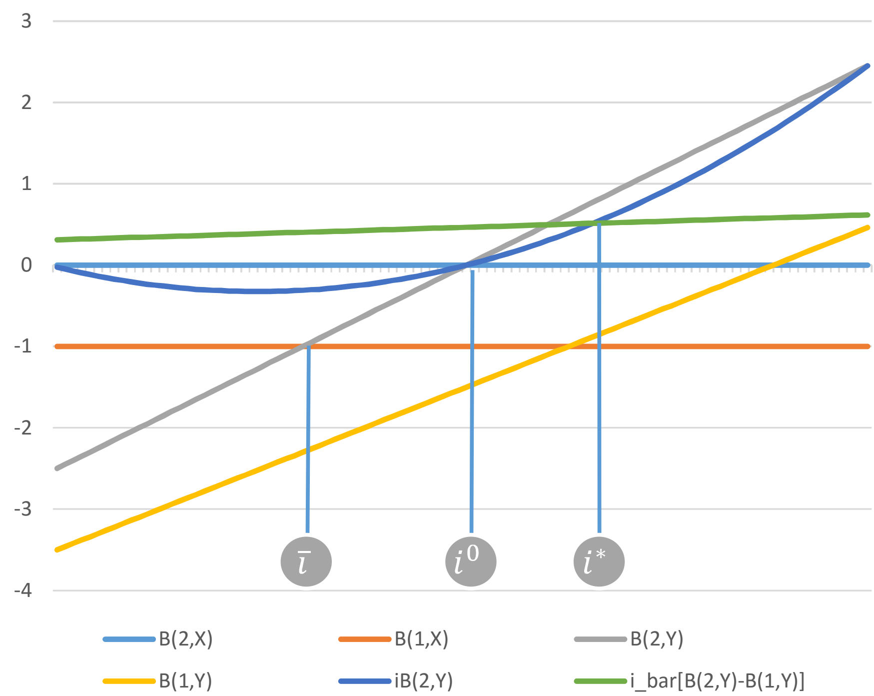
(b) $B^{i^*}(2,Y) > 0$.

Proof:

- $i^*B^{i^*}(2,Y) = \bar{\imath}B^{i^*}(2,Y) - \bar{\imath}B^{i^*}(1,Y)$ by the definition of i^*. Therefore $B^{i^*}(1,Y) = \frac{(\bar{\imath} - i^*)B^{i^*}(2,Y)}{\bar{\imath}}$. Now $\bar{\imath} > 0$ and $\bar{\imath} \leq i^*$ imply that $B^{i^*}(2,Y)$ and $B^{i^*}(1,Y)$ have opposite signs, and then $B^i(2,Y) \geq B^i(1,Y)$ gives the result.

These lemmas are illustrated in Figure 5.

*Figure 5: Illustration of the derivation of the critical levels ibar, i^0, and i**



As Lemma 2 shows, there is a region below i^* where $B^i(2,Y) > 0$, and if all firms' types fall in this region, the switch will not be made even though it would have been made in a world of complete information and although both firms would then be better off – a case of symmetric excess inertia.

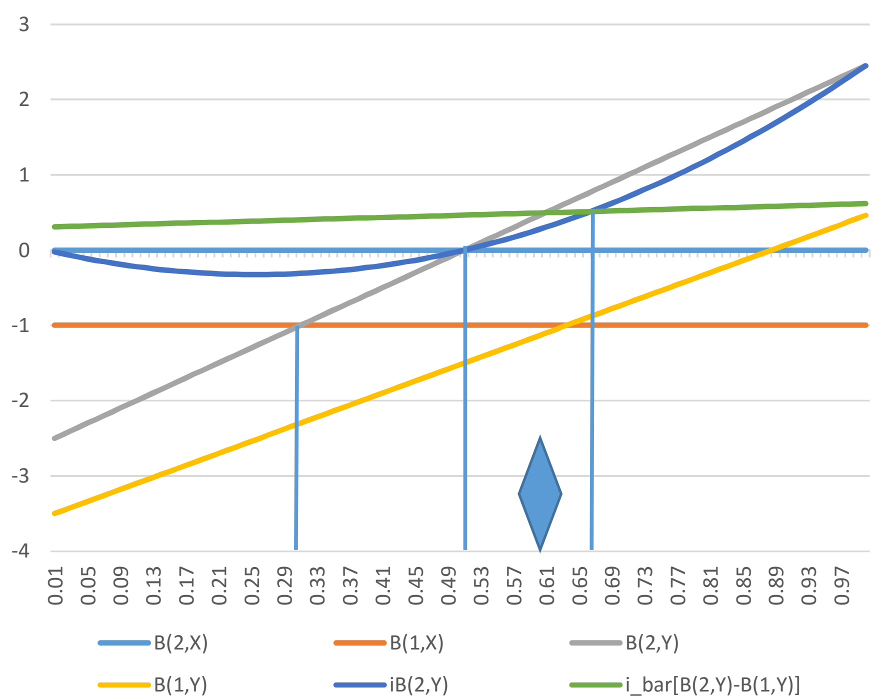
To summarize, excess inertia is problematic in the following two circumstances: cases with information or reaction lags, and cases in which firms have conflicting preferences as to choice of standard (even if network effects are still enjoyed). (Tirole, 1988)

4.2.1.1 *Two correspondent banks: symmetric excess inertia*

In the first game with incomplete information, we have two correspondent banks of "type" 0.6. Since they are both of a type in which they follow the strategy to adopt to the new technology only if the other firm does so, neither firm switches. Since net benefits of shifting to the new technology exceed the benefits of remaining with the current technology, this represents a case of excess inertia. Farrell and Saloner describe the situation of symmetric excess inertia in the two-firm case as follows: "Both firms are fence sitters, happy to jump on the bandwagon if it gets rolling but insufficiently keen to set it rolling themselves." Figure 6 below has the following levels of i:

$\bar{\imath}$	0.31
i^0	0.51
i^*	0.67

Figure 6: Symmetric excess inertia (diamond represents correspondent banks' "type")



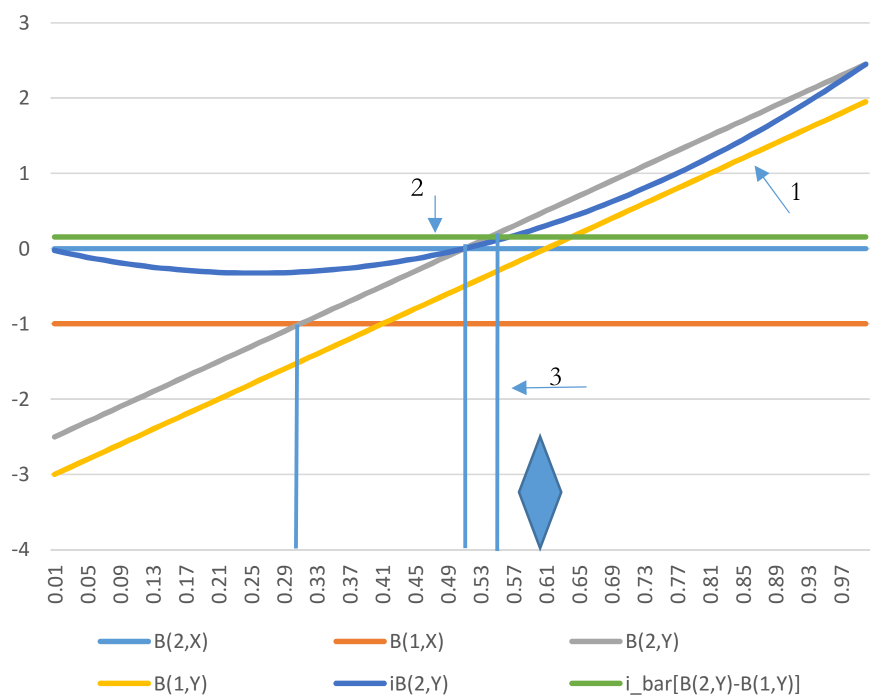
In the chart above, the two correspondent banks of type 0.6 will not switch, even though in this scenario there would be a net benefit of switching. This might describe the scenario from the 1970's until recently, where improvements in systems entailed significant fixed costs and were risky if peer institutions did not implement similar updates.

With recently introduced digital and peer-to-peer technologies (e.g., Bitcoin (Nakamoto, 2008)), the downside risk of switching to a new technology without peers adopting it as well decreases, represented by an increase in $B^i(1,Y)$. This ultimately decreases the curve $\bar{\imath}[B^i(2,Y) - B^i(1,Y)]$, which decreases the point at which that curve and $iB^i(2,Y)$ intersect, thus by definition decreasing i^*, ultimately reducing the space between i^0 and i^*, where symmetric excess inertia occurs.

With this increase in $B^i(1,Y)$ and the resultant shifts, the values of i and the figure describing the situation are as follows:

	Original	After shift
$\bar{\imath}$	0.31	0.31
i^0	0.51	0.51
i^*	0.67	0.56

Figure 7: Overcoming symmetric excess inertia



With our assumption of two correspondent banks of type $i = 0.6$, the switch would not occur under the old technological paradigm, even though both banks would have been better off switching (a case of symmetric excess inertia), while the decrease in downside risk created by the new technology decreases i^* sufficiently to eliminate the excess inertia and incentivize both banks to adopt the new technology. This might be the case in cross-border banking infrastructure, where investment has increased and correspondent banks have displayed an increased willingness to experiment with new technologies, such as the utility settlement coin.

4.2.1.2 *Correspondent banks and money transfer operators: asymmetric excess inertia*

In cases of symmetric excess inertia, however, communication between firms can overcome the inertia. Since correspondent banks are generally not precluded from communicating with each other, I present an alternative explanation for why banking infrastructure has remained at an inefficient equilibrium and the forces that might be catalysing a shift to a more efficient state.

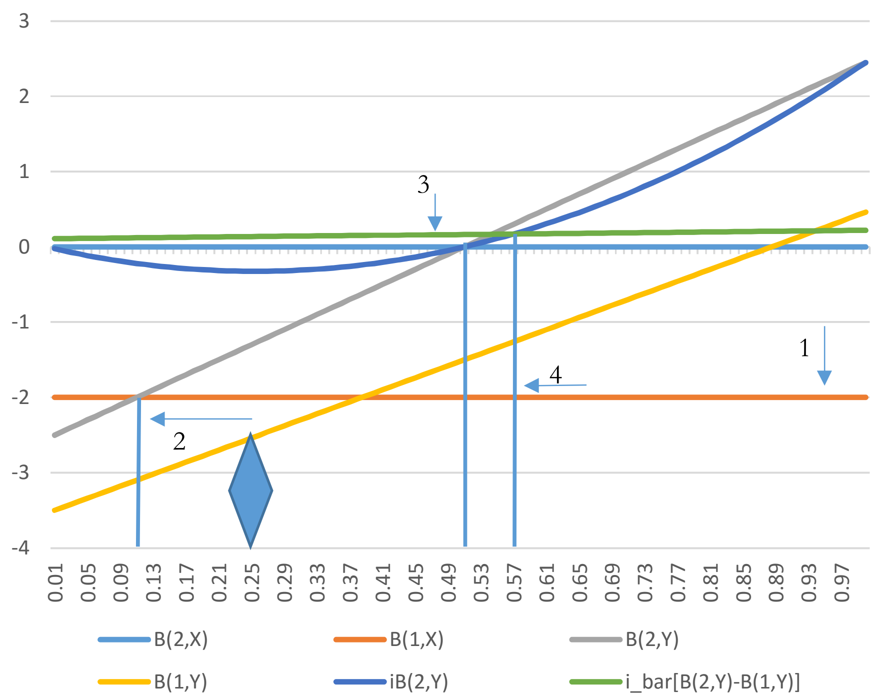
In this case where we have asymmetric excess inertia, correspondent banks are of type 0.25 and initially prefer to remain with the old technology regardless of competitors' choices. Regional banks, money transfer operators, and any other entities that provide remittance services might have preferred a system upgrade to a new technology but were powerless to make a move. As a result, the system has operated on legacy technology and costs of inefficiencies were passed along to consumers (and there was no risk of B(1,X), in other words being left alone with the old technology, as it was virtually impossible for a firm to unilaterally adopt a new technology). Now, firms are beginning to have alternatives. One example of a firm taking advantage of the blockchain is Abra, which uses the Bitcoin blockchain to store digital cash on a user's phone in order to provide near-free remittance services between the United States and the Philippines.[22]

The risk for a money transfer operator of not exploring the possibility of incorporating new technologies, regardless of the adoption decision of the correspondent banks, is greater now than before. Based on the shift in parameters listed in the table below, a correspondent bank originally would not switch to the new technology regardless of what its opponent did. After the shift in the old technology's stand-alone value, it will switch if its opponent switches.

	Original	After shift
$\bar{\iota}$	0.31	0.11
i^0	0.51	0.51
i^*	0.67	0.57

[22] Citigroup. "US Digital Banking: Could the Bitcoin Blockchain Disrupt Payments?" 30 June 2016.

Figure 8: Overcoming asymmetric excess inertia



This model shows both that an inefficient equilibrium might be maintained for quite a long time and that a large change in the technology alternative can create a significant change in the firm types that will switch. Blockchain, as a "foundational" technology (Iansiti & Lakhani, 2017), is likely to be a large enough improvement on current systems to have created a significant shift in the firm types that will adopt the new technology.

5 The costs of coordination failure in remittances

I describe two main types of costs that result from excess inertia in the financial services infrastructure. First is the direct cost of financial frictions, such as the $20 billion of potential cost savings mentioned by Santander. A key question is how much of these savings will be passed along to consumers. As an indicator, Oliver Wyman's "The Capital Markets Industry" indicates that correspondent banks have an 80 percent cost to income ratio, which might indicate that a significant portion of efficiency gains will be passed along to consumers. The second is the opportunity cost of unnecessarily high financial sector employment, an argument sometimes made based on the hopful assumptions that there will be more productive jobs available and that laid-off financial services employees would be able to fill them.

A third type of cost, which I do not describe further, is intangible costs such as the time/opportunity costs and frustration of individuals attempting to make an international transfer.

5.1.1.1 Direct benefits of financial technological improvement

Santander estimates $15 - $20 billion in savings for cross border payments, securities trading, and regulatory compliance, while Goldman Sachs estimates savings of $11 - $12 billion for the clearing and settlement of cash securities (equities, repo, and leveraged loans). Goldman Sachs estimates an additional $3 - $5 billion in cost savings by easing the process of complying with anti-money laundering (AML) and "know your customer" (KYC) regulations. These savings would be achieved through reductions in compliance personnel and AML regulatory penalties.

To get a sense for potential consumer savings, I start with a quick calculation for one category of cross-border payments: international remittances. As mentioned in the introduction, global remittances totaled $581.6 billion in 2015, of which $431.6 billion went to developing countries. The average cost of total remittances was 7.4 percent, corresponding to a total consumer cost of $43 billion. Additionally, large banks continue to close correspondent banking accounts of MTOs to limit exposure to money laundering and other financial crimes, adversely affecting remittance costs and flows in rural and remote regions. (World Bank, 2016)

Examples of firms using new technology to decrease the cost of cross-border financial services exist but are not yet sufficiently ubiquitous to benefit a majority of consumers. Potential consumer benefits are large if Abra and/or similar companies are successful in their expansion plans beyond targeted corridors..

5.1.1.2 Opportunity costs of financial sector overemployment

Financial sector size has an inverted U-shaped effect on productivity growth, and its growth tends to be a drag on productivity growth. (Cecchetti & Kharroubi, 2012) This is a policy and regulatory area in which there are opportunities to improve real growth potential by decreasing barriers to entry of innovative technology, such as complex financial regulation that in developed countries "result in the employment of several hundred thousand people. To employ such a large number of talented people to deal with complex regulation constitutes a large 'deadweight' cost to society." (King, 2016)

More Americans are employed in financial services than is optimal for maximizing economic growth according to the BIS. (Cecchetti & Kharroubi, 2012) In that paper 3.9 percent is cited as the turning point at which the share of financial intermediation in total employment starts being a drag on real GDP-per-worker growth. Currently, financial services comprises 4.1 percent of total US employment,[23] 256,000 more people than would theoretically be optimal for economic growth. Analysts are predicting large decreases in financial sector employment as a result of general technological/digitalization trends. [24] Short-term transition will probably be painful, although some economists believe that there will be longer-term benefits. As Larry Summers described the impact of financial sector job losses: "If a smaller share of

---

[23] Numbers from BLS for March 2017, accessed April 14, 2017. 6,223,100 financial sector employees out of total employment of 153,000,000.

[24] As an example of the impact of technological changes, Citigroup predicts 770,000 U.S. financial sector jobs (and over 1.8 million worldwide) will be lost in the next decade to digital disruption.

America's most talented people find themselves working in the financial sector, I think the nation will survive. And so I don't think that we should design our policies with the objective of maximizing employment in the financial sector. And I think the experience is that in general when you reduce frictions you make possible previously unimaginable combinations and that you tend to see benefits that you couldn't foresee at first before those frictions were reduced."

6 Policy suggestions

> *"I think that the Internet is going to be one of the major forces for reducing the role of government. The one thing that's missing but that will soon be developed is a reliable e-cash. A method whereby, on the Internet, you can transfer funds from A to B without A knowing B or B knowing A. […] Of course it has its negative side. It means that the gangsters, the people who are engaged in illegal transactions, will also have an easier way to carry on their business."*
>
> *Milton Friedman, 1999*

Past efforts to introduce new technology for remittances have confronted the excess inertia problems and dynamics found in many global coordination efforts: powerful and largely risk-averse incumbents (Farrell & Saloner, 1986), problematic attempts to improve the situation with committees (Farrell & Saloner, 1988), and existing converter institutions (correspondent banks) that link various existing systems/countries, however inefficiently (Farrell & Saloner, 1992). DLT is different: it reduces the very excess inertia that sustains the present network effects.

A form of DLT is likely to become a critical, perhaps dominant, transaction infrastructure with or without policy support. Although financial institutions have not yet implemented DLT on a large scale, in this case a slow and cautious adoption process may be optimal. While improving the efficiency of payment, clearing, and settlement systems is a significant aim of regulators, safety concerns are paramount. Since DLT is as yet untested at scale, understanding the risks it presents is critical. These risks include not only the credit, liquidity, general business, custody, investment, and operational risks that are a part of most payment, clearing, and settlement activities. (CPMI and IOSCO, 2012) DLT poses new or different risks including

i. Potential uncertainty about operational and security issues arising from the technology;
ii. The lack of interoperability with existing processes and infrastructures;
iii. Ambiguity relating to settlement finality;
iv. Questions regarding the soundness of the legal underpinning for DLT implementations;
v. The absence of an effective and robust governance framework; and
vi. Issues related to data integrity, immutability, and privacy. (CPMI, 2017)

This list, from the main international body concerned with the safety and efficiency of international payment systems and related financial stability and wider economic issues, indicates that, at least at the staff level, policy makers are grappling with the new risks created by DLT and working to identify ways to mitigate them. The CPMI focuses its work on four areas of questions:

1. Understanding the DLT arrangement,
2. Potential implications for efficiency,
3. Potential implications for safety, and
4. Potential broader financial market implications.

In attempting to maintain the stability of the system, policymakers may also need to address issues of moral hazard: "Market forces alone will not necessarily achieve fully the public policy objectives of safety and efficiency because FMIs and their participants do not necessarily bear all the risks and costs associated with their payment, clearing, settlement, and recording activities," (CPMI and IOSCO, 2012) partially as a result of the implicit subsidy provided to systemically important entities and activities.

In addition to the CPMI's analysis, the major central banks have researched several applications of DLT and potential impacts. An example is the Bank of England's (Barrdear & Kumhof, 2016), who study the macroeconomic consequences of issuing central bank digital currency ("a universally accessible and interest-bearing central bank liability") via distributed ledgers.

There is potential for a rapid scaling up of DLT as a distributed system, with or without policy support or barriers. If a "killer app" using the blockchain is developed and proves widely successful, change could happen quickly. Bitcoin has been called the killer app for blockchain, much as email was for the early internet. (Ito, Narula, & Ali, 2017) Governance arrangements might be similar to those for the internet, but determining the appropriate structure will require extensive research, as the internet is used to transfer information rather than value. Regardless of whether and which blockchain becomes widespread, policymakers should plan for that possibility while supporting incremental improvements to the current system.

The idea of a non-governmental money supply is not new. In 1976 Friedrich Hayek wrote about the denationalization of money, arguing that a competitive system of money creation would be preferable to the current system in which government retains a monopoly. "The further pursuit of the suggestion that government should be deprived of its monopoly of the issue of money opened the most fascinating theoretical vistas and showed the possibility of arrangements which have never been considered." (Hayek, 1976) At the time he lamented that "my proposals are at present impracticable," but the advent of DLT and cryptocurrencies such as Bitcoin provides an opportunity to empirically test his ideas. Reflecting the potential of alternative currencies, the market capitalization of cryptocurrencies has grown rapidly and is over $55 billion as of May 14, 2017.[25]

---

[25] The source for this statistic is https://coinmarketcap.com/. The largest cryptocurrency is Bitcoin, with market capitalization of $29.5 billion.

There are many potential applications of DLT. Policymakers may want to work with financial services firms to ensure that they improve systems incrementally (using a permissioned form of DLT) in an attempt to avoid the success of a non-permissioned (distributed) form of DLT, which might represent payment flows being driven underground and might lack a clear or reliable governance structure. If policymakers coordinate to ensure that DLT is incorporated in incremental improvements to the system, there is a better chance that some form of fiat currency will anchor the decentralized financial system of the future (as opposed to a distributed system such as Bitcoin), which would make it easier for central banks to maintain greater control.

7 Conclusion and directions for further research

In this paper I have applied Farrell and Saloner's model of coordination failure to the remittance industry in order to provide an explanation for why the banking sector has resisted certain technological improvements, and to analyse how new technologies (DLT) might reduce the excess inertia maintaining the current (suboptimal) equilibrium.

As a direction for further research, for a more nuanced analysis of the economic impact, it might be useful to develop an overlapping generations (OLG) model with financial frictions. In this model we might have "productive" and "unproductive" investors, and perhaps the financial infrastructure cost savings will be reflected in a larger portion of "productive" investors.[26] Additional directions for future research might include extending the game to consist of multiple players and using alternative models. As shown in the table below using data from Trefis[27], five banks dominate the global correspondent banking market. Also, the model developed by (Bikhchandani, Hirshleifer, & Welch, 1992) could be used to show that "informational cascades" can cause rapid fluctuations, which may be what is occurring in cross-border banking infrastructure with the large increases in investment in this area.

Table 1: Top five correspondent banks by assets under custody and/or administration

Assets under custody and/or administration	Q1 2016 (USD trillions)	Percent of Total
Bank of New York Mellon	29.1	17.1
State Street	26.9	15.8
JP Morgan	20.6	12.1
Citigroup	14.8	8.7
BNP Paribas	9.0	5.3
Top 5 Total	*100.5*	*59.1*
Industry Total	**170**	

Although it is probable that a distributed ledger will be more efficient than a centralized system (e.g., less reconciliation, quicker end-to-end processing) for certain transactions, safety is also a critical public policy objective and it is important that policymakers continue working to understand the technology that may replace

[26] http://crei.cat/files/filesActivitySummer/30/Lectures%202-3.pdf page 7.
[27] The data in the table is from the following analysis: http://www.nasdaq.com/article/what-is-the-market-share-of-the-5-largest-custody-banks-in-the-global-custody-banking-industry-cm635397. A recent article indicates that assets under custody and administration of the four largest banks increased by 6.8 percent as of Q1 2017: https://www.forbes.com/sites/greatspeculations/2017/05/08/largest-custody-banks-saw-asset-bases-swell-almost-7-in-last-year/#6acbddba8f5d. Since the more recent analysis did not include the overall industry estimate I show the 2016 data in the table.

current systems. Not only does a shift to a distributed (as opposed to decentralized) system make it potentially easier to conduct certain illegal activities, it introduces challenges to development and adoption, including in issues around business cases, technological hurdles, legal considerations, and risk management considerations. (Mills, et al., 2016) Awareness of DLT has increased concurrently with a decrease in correspondent banking networks resulting from stricter regulations. It is important that policymakers monitor developments and continue working to understand the implications of DLT in their attempt to make low-cost financial services available to the greatest possible number of people without sacrificing the stability and resilience of the system.

As the model shows, a technological upgrade using some form of DLT is likely to happen, and well-crafted rules will increase efficiency and promote financial inclusion. However, it is critical that the technology is better understood before it gains widespread use. Poorly-written (or a lack of) rules and poor coordination might decrease financial stability and, if monetary flows move to a distributed non-fiat system, weaken the ability of policymakers to manage future financial and economic crises.

8 Bibliography

Ackerberg, D. A., & Gowrisankaran, G. (2006). Quantifying Equilibrium Network Externalities in the ACH Banking Industry. RAND Journal of Economics, Volume 37, No. 3: pp. 738-761.

Arnold, M. (23 August 2016). Big Banks Plan to Coin New Digital Currency. Financial Times.

Arthur, W. (1989). Competing Technologies, Increasing Returns, and Lock-In by Historical Events. Economic Journal, Vol. 99, No. 394: pp. 116-131.

Banerjee, A. V. (1992). A Simple Model of Herd Behavior. Quarterly Journal of Economics, Vol. 107 No. 3: pp. 797-817.

Barrdear, J., & Kumhof, M. (2016). Staff Working Paper No. 605: The Macroeconomics of Central Bank Issued Digital Currencies. London: Bank of England.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. Journal of Political Economy, Vol. 100, No. 5: pp. 992-1026.

Carney, M. (2017). Chairman's Letter to G20 Finance Ministers and Central Bank Governors. Baden-Baden: FSB.

Cecchetti, S. G., & Kharroubi, E. (2012). Working Paper No. 381: Reassessing the Impact of Finance on Growth. Reserve Bank of India's Second International Research Conference. Mumbai, India: Bank for International Settlements.

CPMI. (2015). Consultative Report: Correspondent Banking. Basel: Bank for International Settlements.

CPMI. (2016). Correspondent Banking. Basel: Bank for International Settlements.

CPMI. (2017). Distributed Ledger Technology in Payment, Clearing and Settlement: An Analytical Framework. Basel: Bank for International Settlements.

CPMI and IOSCO. (2012). Principles for Financial Market Infrastructures. Basel: Bank for International Settlements.

CPSS. (2005). Central Bank Oversight of Payment and Settlement Systems. Bank for International Settlements.

El-Erian, M. A. (2016). *The Only Game in Town: Central Banks, Instability, and Avoiding the Next Collapse*. New York: Random House.

Farrell, J. (1982). Communication in Games. Mimeo.

Farrell, J., & Saloner, G. (1985). Standardization, Compatibility, and Innovation. RAND Journal of Economics. Vol. 16, No. 1: pp. 70-83

Farrell, J., & Saloner, G. (1986). Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation. American Economic Review, Vol. 76, No. 5: pp. 940-955.

Farrell, J., & Saloner, G. (1988). Coordination through Committees and Markets. RAND Journal of Economics, Vol. 19, No. 2: pp. 235-252.

Farrell, J., & Saloner, G. (1992). Converters, Compatibility, and the Control of Interfaces. Journal of Industrial Economics, Vol. 40, No. 1: pp. 9-35.

Federal Reserve. (2015). Strategies for Improving the U.S. Payment System.

Gifford, K., & Cheng, J. (2016). Implementation of Real-Time Settlement for Banks Using Decentralised Ledger Technology: Policy and Legal Implications. Banque de France Financial Stability Review, (pp. 143-151). Paris.

Hayek, F. A. (1976). *Denationalisation of Money -- The Argument Refined*. West Sussex: Goron Pro-Print Co Ltd.

Hemenway, D. (1975). *Industrywide Voluntary Product Standards*. Cambridge: Ballinger.

Iansiti, M., & Lakhani, K. R. (2017). The Truth About Blockchain. Harvard Business Review.

IMF. (2009). International Transactions in Remittances: Guide for Compilers and Users. Washington, D.C.: International Monetary Fund.

Ito, J., Narula, N., & Ali, R. (2017). The Blockchain Will Do to the Financial System What the Internet Did to Media. Harvard Business Review.

Johnson, S. (Spring, No. 40, 2016). The Financial System of the Future. Democracy: A Journal of Ideas.

Katz, M. L., & Shapiro, C. (1985). Network Externalities, Competition, and Compatibility. American Economic Review, Vol. 75, No. 3: pp. 424-440.

Keser, C., Suleymanova, I., & Wey, C. (2012). Technology Adoption in Markets with Network Effects: Theory and Experimental Evidence. Information Economics and Policy, Vol. 24, No. 3: pp. 262-276.

King, M. (2016). *The End of Alchemy: Money, Banking, and the Future of the Global Economy*. New York, NY: W.W. Norton & Company, Inc.

Liebowitz, S. J., & Margolis, S. E. (1990). The Fable of the Keys. Journal of Law and Economics, Vol. 33, No. 1: pp. 1-25.

Liebowitz, S. J., & Margolis, S. E. (1994). Network Externality: An Uncommon Tragedy. Journal of Economic Perspectives, Vol. 8, No. 2: pp. 133-150.

Mersch, Y. (2016). Making Europe's Financial Market Infrastructure a Bulwark of Financial Stability. Banque de France Financial Stability Review.

Mills, David; Wang, Kathy; Malone, Brendan; Ravi, Anjana; Marquardt, Jeff; Chen, Clinton; Badev, Anton; Brezinski, Timothy; Fahy, Linda; Kiao, Kimberley; Kargenian, Vanessa; Ellithorpe, Max; Ng, Wendy; and Baird, Maria. (2016). Distributed Ledger Technology in Payments, Clearing, and Settlement. Washington: Board of Governors of the Federal Reserve System.

Milne, A. (2005). What's in It for Us? Network Effects and Bank Payment Innovation. Helsinki: Bank of Finland.

Mougayar, W. (2016). *The Business Blockchain: Promise, Practice, and Application of the Next Internet Technology*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.

Oren, S. S., & Smith, S. A. (1981). Critical Mass and Tariff Structure in Electronic Communications Markets. Bell Journal of Economics, Volume 12, No. 2: 467-487.

Pauget, G. (2016). Systemic Risk in payments. Financial Stability Review: Financial Stability in the Digital Era (pp. 37-44). Paris: Banque de France.

The Economist. (6 May 2017). Special Report: International Banking. The Economist.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: The MIT Press.

World Bank. (2016). Migration and Development Brief 26. Washington, DC: KNOMAD: Global Knowledge Partnership on Migration and Development.

World Bank Group. (2015). Report on the G20 Survey on De-Risking in the Remittance Market. Washington, DC: International Bank for Reconstruction and Development.

9 Appendix

Proof of Proposition 2, as presented in (Farrell & Saloner, Standardization, compatibility, and innovation, 1985). They prove three lemmas to prove Proposition 2:

Lemma A. If $n = 2,$ each firm (nonstrictly) prefers to go first.

Proof. Firm A and B's decisions (X or Y) are k_A or k_B. Let (k_A, k_B) be an equilibrium when A goes first. Then, k_A is also A's best response to k_B. Therefore, B can achieve its payoff from (k_A, k_B) by moving first, simply by choosing k_B. However, B might be able to do better by making another choice when going first.

Lemma B. Whatever n may be, any firm would (nonstrictly) rather be #1 than #2.

Proof. This follows from Lemma A if we collapse the responses of firms 3, 4, … , n into the payoffs for firms A and B, which are trading places 1 and 2. All that needs to be checked is that the reduced game continues to satisfy Assumption 1, and that is clear.

Lemma C. For any n, and any $j = 1, 2, \ldots, (j - 1)$ a firm in position (j + 1) would (nonstrictly) like to trade places with the firm in position j.

Proof. Lemma B assures us that this would be true if we think of the actions of 1, 2, …, (j – 1) as not responding to the change. We must then show that any response by the early firms will be favourable to the firm, which has switched from (j + 1) to j.

The reason this is true is that the switch has made the consolidated response function of the players more in line with B's preferences (Lemma 1). Therefore, players 1, … , (j – 1), considered as playing a game with the responses of j, j + 1, …, n collapsed into the payoff functions, have had their preferences shifted in the direction of B's desires.

Proposition 2 now follows by repeated application of Lemma C. that is, to show that, given the order of the other (n – 1) firms, a firm prefers to be earlier in that sequence rather than later, one simply imagines the firm's repeatedly moving up one place and bumping its predecessor one place down. This proves Proposition 2.