What does the theory of two-sided markets tell us about competition in the interbank and card payment systems?

February 2016
What does the theory of two-sided markets tell us about competition in the interbank and card payment systems?

This paper explores the theory of two-sided markets and the common features that these markets exhibit. It then applies this theory in the context of payment systems to help rationalise the market structures in place, as well as providing some preliminary views on what may happen in the future.

What is special about two-sided markets?

Interdependence – two-sided markets are markets characterised by two distinct end user groups, where one user group’s participation is in some way dependent on the other user group’s participation, and vice versa.¹ There are many examples of such markets: the newspaper market targets both advertisers and readers; the games consoles market attracts both game developers and gamers; and the payment cards market supports both merchants and customers. There are also examples of multi-sided markets with more than two types of end users attracted to a given platform. For example, music-streaming services, such as Spotify, attract music artists, advertisers and consumers. While similar analysis would apply in the context of such multi-sided markets, we focus the remainder of this paper on two-sided markets.

The key implication of the interdependence of the user groups in two-sided markets is that the price structure matters and not just the overall price level summed across the two sides of the market. This means that, for a given total price, the total volume of interactions on the platform will vary as the allocation of that price between the two sides of the market is varied. Rochet and Tirole (2003) explain that increasing the price by a small amount on one side of the market, while reducing the price by an equally small amount on the other side of the market, will result in a change in the overall volume of transactions. The diagram below shows that a slight decrease in price on ‘Side A’, increases quantity on side A, which causes an outward shift in the demand curve on ‘Side B’ (due to the interdependence), such that the demand increases on ‘Side B’ in spite of a slight price increase.

Single v multi-homing – some two-sided markets are characterised by two phases of decision-making: a membership decision, i.e. whether to join the platform, or not; and a usage decision, i.e. whether to

transact over that platform once you have joined. How these two phases of decision-making play out can be a key determinant of bargaining power and thus how price is shared between the two sides of the market. A user who joins only one platform is said to single-home, while a user who joins more than one platform is said to multi-home. Users who single-home would lose the ability to interact with users on the other side of the market who are not a member of that same platform. The literature suggests the following help determine the extent of multi-homing on any one side of a market.2

- **The extent of multi-homing on the other side of the market** – the greater the extent of multi-homing on one side, the less incentive there is to multi-home on the other side of the market. If, for example, all the same film producers had deals with both Netflix and Amazon Prime, then there would be little incentive for a consumer to have both Netflix and Amazon Prime accounts. This may, in part, explain the big drive towards exclusive content on these platforms, as a means of steering consumers to a preferred platform. Clearly, this raises a ‘chicken and egg’ problem, as the decision to multi-home on one side is affected by the decision to multi-home on the other, and vice versa.

- **The extent of platform differentiation** – the extent to which homing decisions vary with platform differentiation may depend on the type of platform differentiation observed. Vertical differentiation, i.e. by quality, may see users single-home on either a high- or low-quality platform, depending on how much they value that quality. However, horizontal differentiation, i.e. differentiation by precise product offering, may be more susceptible to multi-homing.3

- **The cost of joining the platform** – if the costs of joining the platform are high, then the more likely it is for users to single-home. If the costs of joining to one side of the market are very low relative to the other side, then, other things being equal, we would expect the former side to multi-home and the latter to single-home.

Overall, therefore, the extent of multi-homing on one-side of the market will reflect some combination of these (and other) factors. The implications for competition are that platforms tend to compete more intensively for the side of the market on which there is a higher degree of single-homing. This is because by attracting users on the single-homing side of the market, the platform gains market power in the other side of the market, by having a monopoly over access to its single-homing end users. This means that, while the single-homing side is likely to benefit from lower prices through competition, the (relatively) multi-homing side is likely to experience higher prices due to the platform exercising its market power.

**Network effects** – most two-sided markets exhibit network effects. Network effects arise where the participation of an additional user in a network changes the value of that network to existing users.4 In a two-sided context there are both direct and indirect network effects.

Indirect network effects arise where additional participation on one-side of the market affects the value to users on the other side of the market. An example of this would be the benefit of an additional music artist signing up to a music-streaming platform to the benefit of existing consumers of music streaming services. Similarly, existing music artists would benefit from additional consumers signing up to the music-streaming platform, as the music artists earn a small fee for each time one of their songs is played.

Direct effects arise where the presence of an additional user on one side of the market affects the value to users on the same side of the market. This could either be as a result of increased possibilities of

---


3 In this latter case, the question of what constitutes horizontal differentiation and what constitutes an entirely different market arises.

4 Katz and Shapiro (1994) give the analogy of a “hardware/software” paradigm. Two-sided markets often involve the adoption of some hardware, e.g. a mobile phone, which in turn lets them access software, e.g. mobile apps. Thus in making a decision about whether to adopt the hardware, a rational user would be forming expectations about the subsequent availability of hardware, with the extent of software provision itself dependent on how many users adopt the hardware. These indirect network effects cause rational expectations to become self-fulfilling. See: Katz, M. L. and C. Shapiro (1994), “Systems Competition and Network Effects”. *Journal of Economic Perspectives*, vol. 8(2), pp. 93-115.
interaction on that side of the market, for instance we benefit from more of our friends signing up to the same social media site – or as a result of the additional users they attract on the other side of the market. A good example of the latter would be the benefits that new app store users exert on existing app store users, by the former encouraging new app developers to enter the market.

The presence of these network effects tends to lead to relatively concentrated markets. This is because of the positive feedback effect whereby the value of the network increases as the network grows in size, thereby further encouraging growth in the network, which increases the network value, and so forth. The upshot is that in a two-sided market with two platforms which are ex ante identical, a small temporary advantage in the market share for one could, through network effects, be enough for that platform to become the dominant force in the market. The concern, therefore, is that the outcome achieved by the market may not be socially optimal, as a lower quality platform may drive out a higher quality platform, due to mutually reinforcing network effects – and some initial luck.

The extent of market concentration would depend, among other factors, on how the strength of network effects vary with the size of the network. In some markets, indirect network effects may be positive but diminishing, for example due to overcrowding or due to increasing costs of search. However, for other markets, it is possible that they may be close to constant (i.e. non-diminishing). Two-sided markets characterised by the latter may experience a natural tendency towards a monopoly platform.

Economies of scale – a common feature of two-sided market that may compound platform concentration is the high fixed costs of platform development relative to ongoing variable costs. This cost structure means that incumbent platforms can benefit greatly from economies of scale, whereby the cost per transaction falls as the number of transactions increase (as the fixed costs are spread increasingly thinly). Again, a positive feedback loop can develop, whereby more transactions lower the costs per transaction, which attract more transactions and so on. As a result, network effects and economies of scale can be mutually reinforcing in driving platform growth/concentration.

Mitigating factors – there may be some factors that mitigate the drive to platform concentration caused by network effects and economies of scale. First of all, if there were cross-compatibility between platforms then this would reduce the strength of network effects on any one platform. This would be the case if the infrastructure/technology that an end user needed to access one platform, was the same as that needed to access another platform.

Another mitigating factor may be the ability of a firm to leverage its position in an existing market in order to establish itself in a new market, in spite of the presence of a large incumbent platform. The logic is that if a consumer is happy with the service they receive from that organisation in a complementary market, then they would more readily switch to this organisation if it entered a new market. A good example of this is Google establishing itself in the web browser market with Google Chrome (despite established players like Microsoft Internet Explorer and Mozilla Firefox), by leveraging its strong position in Internet search and email markets.

---

5 There can, of course, be negative network effects, both directly, e.g. overcrowding, and indirectly, e.g. the additional search costs created by advertising, but here we focus on the positive network effects.
7 Of course, some platforms exploit these effects by charging premiums to end users to overcome them. Uber, for example, operates surge pricing in times of high demand, which helps correct for the overcrowding by pricing those currently value the service less out the market at that point in time. Similarly, in terms of increased search costs, some platforms, e.g. Spotify, charge premiums to customers who wish to access their platform without advertising.
8 Roson (2005) makes the point that indirect network effects may be (close to) constant where a two-sided market is characterised by multiple interactions (e.g. payment systems) as opposed to single interactions (e.g. house buying). See: Roson, R. (2005), “Two-sided markets: a tentative survey”. Review of Network Economics, vol. 4(2), pp. 142-160.
What does the theory of two-sided markets tell us about competition in the interbank and card payment systems?

We now consider the implications of the two-sided market theory in the context of interbank and card payment systems. It becomes clear through the discussion that payment systems have some key nuances that complicate the standard picture of two-sided markets. We focus our analysis on two areas – interchange fees in card payment systems and platform competition in the interbank space.

Interchange fees in card payment systems – at first sight, payment systems appear to be like any other two-sided markets: they support two users groups whose participation is dependent on one another – the payers and the payees. In card payment systems, the payers are the consumers (the cardholders) and the payees are businesses/merchants. Therefore, as we saw above, the price allocation on the two sides of the market matters to the total volume. This means that the price on each side of the market should not necessarily reflect marginal cost, but rather be allocated in such a way so as to maximise the number of interactions on the platform.

However, an additional feature of payment systems, which is not true of all two-sided markets, is that the number of transactions must be equal on both sides of the market – a payer must always be matched with a payee for a transaction to occur. The key implication is that the desired transaction volumes on both sides of the market should match, as otherwise the size of the market will be constrained by whichever side demands fewer transactions. The diagram below shows how in these circumstances it is optimal to introduce a side payment from one side to the other, in order to maximise the transaction volume.

In two separate markets, equilibria would be achieved at point x (with $Q_{\text{Payer}}$ transactions) and point a (with $Q_{\text{Payee}}$ transactions). However, in the payment systems setting, the number of transactions would be constrained at $Q_{\text{Payer}}$, at which point the payees would be willing to pay more ($d$) for additional transactions than the price that the supplier requires to offer the transaction. In this scenario, transaction volume could be increased by a subsidy of $bc$ from the suppliers of payees to the suppliers of payers, such that the supply curves shift from $S_1$ to $S_2$ in each market. Thus, while in (the vast majority of) one-sided markets a higher price would lead to a lower volume of transactions, in a two-sided setting this may not be the case, as a higher price on one-side of the market can be used to lower the price on the other side of the market and thus lead to a higher volume of transactions on the side that was previously constraining the market.
This suggests that, in the absence of identical demand conditions on the two sides of the market, some level of interchange fee is socially optimal in payment systems. Of course, in card payment systems, as there are clearly two distinct end user groups (merchants and consumers), then demand conditions should differ and, therefore, a non-zero interchange fee appears optimal.\textsuperscript{9}

As some non-zero interchange appears justified in card payment systems, by extension we can say that, while pricing above marginal cost may be indicative of market power in a one-sided context, pricing above marginal cost in a two-sided context may actually be the sign of the market functioning well. By the same logic, neither should pricing below marginal cost (and in some cases below zero)\textsuperscript{10} necessarily be seen as a predatory technique of incumbents. Instead, what really matters in a two-side setting is how the sum of prices across the two sides of the market compares with the total marginal cost across the two sides.

While some level of interchange fee appears optimal, the implications of single- and multi-homing help provide insight on what is practically optimal for the payment system platforms. We make three key observations of payment systems, in line with the factors affecting multi-homing discussed earlier:

- There is widespread multi-homing by merchants in card payment systems – (almost) all merchants accept both MasterCard and Visa, which may incentivise very few consumers to multi-home (hold both a MasterCard and Visa).\textsuperscript{11} This thought is supported by the fact that a large number of merchants choose not to accept American Express cards because few consumers choose to hold this card exclusively (instead holding an American Express in addition to either a MasterCard or Visa).
- As MasterCard and Visa now operate on the same technology, the marginal cost to a merchant of accepting Visa if it already accepts MasterCard, or vice versa, is effectively zero. Consumers may face, in relative terms, higher costs of joining an additional card platform, as it may involve opening a current account with a new bank which issues this card type or requesting an additional card from their existing bank. Therefore, while not high, these adoption costs may be sufficient to discourage a consumer from multi-homing (especially given the circumstances described in the first bullet).
- There is limited platform differentiation between MasterCard and Visa, which is likely to have reduced the incentives for consumers to multi-home. Where consumer multi-homing does occur, it is usually because the consumer holds an American Express card alongside either a MasterCard or Visa, as American Express differentiates itself as a high quality card payment platform.

The above evidence would suggest limited incentives to multi-home on the consumer side, and this is reflected by the high degree of multi-homing on the merchant side relative to the consumer side.\textsuperscript{12} This may incentivise card platforms to compete more intensely for single-homing consumers, and then exploit any market power they gain through higher prices on the merchant side of the market. Platforms can effectively exploit the fact that merchants’ multi-home and by doing set a more skewed pricing structure for the market. It could be argued that this may go beyond the social optimum (where the sum of consumer and merchant surplus are maximised) to a point at which the consumer surplus is maximised subject to the constraint that merchant surplus is non-negative. By this logic it could be argued that the privately optimal interchange fees exceed the socially optimal interchange fees and hence that an interchange fee cap is justified.

However, there is academic literature, as referenced in Gonçalves (2003), which finds that the socially optimal price \textit{structure} is consistent with the privately optimal price structure, but that the price \textit{level} on

\textsuperscript{9} In interbank payment systems, as the same users can be on both sides of the market, it is unclear to what extent demand conditions may differ and thus whether an interchange fee is justified.

\textsuperscript{10} Think credit card holders, who receive lucrative rewards for using their credit cards for payment transactions.

\textsuperscript{11} What we mean by this is that not many consumers hold a MasterCard and Visa of the same card type, i.e. a MasterCard debit card and a Visa debit card, or a MasterCard credit card or a Visa credit card. It is, of course, more common to hold a MasterCard and Visa, if one is a debit card and one is a credit card as they are serving different functions.

\textsuperscript{12} The incentives for multi-homing and the prevalence of multi-homing are, of course, mutually reinforcing as well.
What does the theory of two-sided markets tell us about competition in the interbank and card payment systems?

Each side of the market will be above the social optimal. This would suggest that an interchange fee cap is not justified, as this only addresses the price structure and not the price level. There are also some other limitations of the above logic that would question the imposition of interchange fee caps:

- **Impact on card v cash usage** – lower interchange fees would discourage card transactions and increase cash transactions, which may be inefficient as studies have found transactions in the latter to be more costly to merchants. Therefore, it may be expected that lower interchange fees to acquiring banks would not be passed through to consumers in lower prices, as such lower interchange fees may increase the number of cash transactions which are of higher cost to the merchant.

- **Impact on nascent card markets and prospective entrants** – lower interchange fees would make it difficult for nascent card markets to grow, or for prospective entrants to mature card markets establish themselves. Such growth is dependent on the market, or entrant, attracting sufficient issuers (and cardholders), such that it becomes attractive for acquirers (and merchants) to join the market or platform. One way in which this can be achieved is through higher interchange fees in the early period of development.

The recent EU Interchange Fee Regulation (IFR) has imposed a 0.2 per cent interchange fee cap on debit card transactions and a 0.3 per cent cap for credit card transactions. There is an exemption to this regulation for those platforms whose market share is less than 3 per cent, but this exemption only holds until December 2018. This regulation could therefore raise concerns about whether more efficient (less costly) card usage could be displaced by less efficient cash usage, and about whether the development of card markets with currently low penetration rates, and the establishment of potential competitor platforms, could be hindered by an inability to set higher interchange fees in the early stages. Such concerns could therefore question the suitability of interchange fee caps.

### 1.1.1 Platform competition

We analyse the implications for platform competition in payment systems using the theory on network effects discussed earlier. Firstly, it is clear that end users in payment systems exhibit positive indirect network effects on one another. As more consumers (payers) adopt a given payment system, the value of that payment system to a given merchant (a payee) increases, and vice versa, as there are more possible consumers/merchants to transact with.

On the payer side there are also positive direct network effects because, as more payers adopt a payment system, this encourages greater uptake by payees, which in turn is of benefit to the original payers on the platform. On the payee side, there may be positive direct network effect (for the same logic just described), but also negative direct network effects, as a result of additional merchants stealing customers from existing merchants (assuming they offer competing goods).

However, a key feature of payment systems which complicates the traditional two-sided market thinking is that end users do not interact directly with the platform, but rather operate indirectly through payment service providers, i.e. banks, as shown in the diagram below.

---


What does the theory of two-sided markets tell us about competition in the interbank and card payment systems?

This means that the decision of whether, or not, to join the platform is effectively devolved from the payers and the payees to their respective banks. Nevertheless, we expect similar network effects to play out between the payer (issuing) and payee (acquiring) banks, i.e.:

- **Positive indirect network effects** – as payee banks benefit from the presence of more payer banks (who send them payments), and vice versa.
- **Positive direct network effects** – existing payee banks benefit from new payee banks joining the platform, insofar as it encourages adoption by new payer banks.
- **Negative direct network effects** – existing payee banks will be made worse off by the arrival of new payee banks due to the increased competition this creates, providing this bank represents a new source of competition.¹⁶

The benefit to an existing bank of another bank joining the platform therefore depends on the relative strength of these effects. Once the platform achieves a high-level of market coverage, then we might expect that the positive indirect and direct network effects (resulting from market expansion) may start to be outweighed by the negative direct network effects caused by increased competition (for existing end users in the market).

In practice, this may mean that the large banks that are partly in control of these payment platforms (Bacs, CHAPS etc.) have little incentive to admit new banks to the scheme. This is because the benefits of increased coverage can be offset by the detriment of increased competition. It may also explain why the major banks, which directly access payment systems, are reluctant to offer indirect access to these systems, as the money the major banks earn through indirect access provision may be offset by the competition that new banks could exert if they gain indirect access. There is clearly a trade-off between ubiquity and competition in payment systems. With an established platform which all of the major banks belong to, it becomes very difficult for a prospective platform to enter the market and compete.

The high cost of developing payment systems infrastructure, and the large economies of scale that result, are likely to have compounded the network effects in creating very dominant platforms in payment systems. Therefore, there is a natural tendency towards platform concentration in payment systems, but is this in fact a problem and, if so, what does theory suggest can be done about it?

Our first observation is that, while significant infrastructure investments are required to offer secure, resilient and high capacity platforms, operational roles (such as developing platform rules and membership

¹⁶ If, for example, a bank which already operates on the Visa platform also joins the MasterCard platform, then this does not really represent a new form of competition to existing banks, as they already compete with this bank on the Visa network and end users perceive these platforms as sufficiently similar. However, if a bank which does not currently operate on either card platform, were to join the MasterCard platform, then this could be seen to represent a new source of competition to existing banks.
criteria) are much less costly. This suggests that economies of scale primarily accrue in relation to the infrastructure costs. As a result, a decoupling of scheme operation from scheme infrastructure provision could be beneficial to payment systems, i.e.:

- In terms of *infrastructure provision*, given the high costs and hence large economies of scale that can be realised, it may be socially optimal to have only one infrastructure provider (as competition may otherwise prevent the full extent of economies of scale being realised).
- In terms of *platform operation*, given the limited economies of scale, it may be socially optimal to have different operators competing to provide banks with access to a common infrastructure (as seen in many utilities sectors, where there is a common shared infrastructure, but competition over the distribution and retail of the utility over that infrastructure).

As well as the platforms incurring high infrastructure costs, so too will the banks in developing their own infrastructure to ‘tap into’ the platform infrastructure. As a result, the cost to a bank of switching payment platform is likely to be very high, and so the incentive to switch from one platform to another is limited. This can be seen to further strengthen the argument for decoupling scheme operation and infrastructure, as this would significantly reduce bank switching costs and thus promote greater competition for the banks.

An alternative to this approach may be to ensure inter-operability of infrastructure. Although this may limit the economies of scale realised in developing one central infrastructure, it should help to promote switching and thus encourage platform competition and efficiency savings. However, given such inter-operability is costly to implement and beneficial to a prospective entrant at the expense of the incumbent, the incumbent would have little incentive to implement cross-compatibility solutions, unless obliged to do so through regulation.

Overall, we have reached the following findings by applying the theory two-sided markets in the setting of payment systems:

- *The rationale for an interchange fee cap may be misguided* – while some academics argue that the private optimal lies above the social optimal, others suggest that the private optimal and social optimal are consistent, and rather that there is only a distortion of the price levels which interchange fees do not correct for. There are also dynamic concerns that interchange fee caps can prevent development of nascent markets or viable competitors in established markets and that they may displace efficient card usage with more costly cash usage.
- *A decoupling of platform infrastructure provision and platform operation may be welfare improving* – there is a natural tendency towards platform concentration in interbank payment systems, given the high infrastructure costs and positive network effects. In this setting, a decoupling of platform infrastructure provision and platform operation may allow for the best of both worlds, i.e. achieving high economies of scale in platform infrastructure provision, while introducing competition and efficiency savings in platform operation.