Credit Card Schemes in Australia

A response to the Reserve Bank of Australia and Australian Competition and Consumer Commission Joint Study

Visa International Service Association

VISA

Prepared by
Network Economics Consulting Group Pty Limited

◐necg

January 2001

Executive Overview

The Joint Study into Debit and Credit Card Schemes in Australia *(A Study of Interchange Fees and Access)* discusses many aspects of the 4-party credit card systems operating in Australia and around the world. It claims that these systems have several socially undesirable attributes.

In particular, the Joint Study asserts that 4-party card systems:

- are more socially costly than debit cards;
- have unnecessarily high interchange fees;
- result in a subsidy being paid to non-revolving users of credit cards from consumers using other payment methods; and
- have unnecessarily restrictive membership rules for their issuers and acquirers.

These conclusions fail to acknowledge, however, that:

- interchange fees help promote network membership and usage;
- proving the existence of a cross subsidy requires looking at the indirect as well as the direct costs and benefits (e.g. transaction costs and foregone benefits) faced by all parties; and that
- as evidenced by the large fall in the cost of credit since their introduction, open credit card systems actually bring competition to the market, both between payment systems and within each open credit card system.

The Joint Study acknowledges that, due to the presence of network externalities in credit card networks, interchange fees serve a useful balancing role, ensuring that growth occurs both on the issuing and on the acquiring side of credit card systems.

The question is then whether these fees are too high. One way to assess this is to look at whether the network is ''unbalanced''. There is, in this respect, nothing in the Joint Study to suggest that current fees are ''unbalancing'' or in any other way distorting the development of credit card networks.

However, the Joint Study argues that credit card networks in Australia are mature and that, as a result, the ''balancing role'' justification for setting the interchange fee to encourage network growth is not relevant.

The evidence presented in this paper shows that credit card networks in Australia are in fact not mature. Rather, they are still experiencing strong growth. No less importantly, this paper demonstrates that even if the networks were mature, the need for an interchange fee would persist. Moreover, the efficient level of that fee might be higher, not lower, than in a network at an initial stage of development.

The Joint Study puts forward two alternative approaches to the setting of interchange fees. Both are cost-based approaches. However, an interchange fee level based solely on costs will only ever be efficient by pure chance – and it is shown here that the approaches the Joint Study proposes are far more likely to hinder than to promote efficiency.

The exclusion of closed credit card networks (e.g. American Express and Diners Club) from the scope of the Joint Study is also a matter for serious concern. Any moves to regulate interchange for open card systems such as VISA would provide a significant competitive advantage for closed credit card networks such as American Express. Indeed, there is a strong argument that if the interchange fees set by the open systems are regulated, then there is a clear need for final price regulation of closed systems.

Furthermore, any move to regulate the setting of interchange fees would have serious international ramifications, placing Australia at odds with global practice.

Overall, this paper shows that efficient determination of the interchange fee involves a myriad of factors. Mechanical approaches, of the kind favoured by regulators and advanced by the Joint Study, cannot capture this complexity, and hence entail efficiency losses. In contrast, the current arrangements, in which these fees are determined by the members in the course of negotiation, will benefit from an understanding of commercial realities that regulators lack.

The Joint Study also argues that two other aspects of 4-party card networks should be changed. In particular, it proposes:

- abandonment of the no-surcharge rule; and

- relaxation of the rules regulating membership of the open card systems.

The proposal to abandon the no-surcharge rule is a consequence of the Joint Study's conclusion that current arrangements provide a cross-subsidy to non-revolving credit card users from consumers using other methods of payment. However, the Joint Study's analysis is mistaken as a matter of economics. Rather, it is shown here that the no-surcharge rule both promotes efficiency in the operation of credit card systems and protects consumers. The costs to the Australian community of removing the no-surcharge rule are conservatively estimated as being in excess of $4 billion.

Nor are the Joint Study's conclusions with respect to membership rules any better based. Well-established and widely accepted economic principles highlight the need for entry criteria of exactly the type used by VISA.

There are serious risks associated with any dilution of prudential standards for those seeking entry into open card networks. Members, be they card issuers or merchant acquirers, are required to bear very substantial liabilities towards consumers, merchants, system members and the system as a whole. The current arrangements rely on the RBA to ensure that members have the capacity to discharge the obligations this entails. Weakening the membership criteria could jeopardise the financial soundness of the systems in Australia and overseas.

In addition to their impact on financial soundness, changes to the access rules would damage the functioning of the systems, including by raising the costs involved in decision-making and reducing the effectiveness of governance arrangements. There is consequently no basis for thinking that such changes would be in the public interest and hence consistent with any of the current legislative instruments.

The issues raised by the Joint Study are of great importance. They are important to the institutions directly involved and to the Australian community.

Currently, the Australian payments system is strong, secure, dynamic and technologically sophisticated.

It has evolved in a way that has allowed the economic development of credit card systems that are sought after by consumers and merchants.

Regulatory intervention in a successful, privately run and competitive joint venture would set an undesirable precedent. It may well have destabilising effects on the Australian Payments System, isolating it from global best practice. It should only be envisaged if compelling grounds for intervention can be advanced.

The Joint Study seems to favour regulation, but the arguments it puts simply do not resist scrutiny. Rather, experience and analysis of open credit card systems suggest that regulation is likely to do more harm than good. The RBA and the ACCC should carefully reconsider the views they have put.

Table of Contents

1. Summary

The joint RBA/ACCC study (henceforth the "Joint Study") examines several important aspects of electronic payment systems. It reaches strong conclusions, and advances recommendations that could have far-reaching implications. However, the Joint Study errs in numerous respects, seriously compromising the quality and reliability of its analysis.

At the most general level, the Joint Study is notable in lacking any kind of economic foundation for the views it advances. No attempt is made in the Study to systematically assess, much less quantify, the effects of the current arrangements on aggregate welfare. As a result, the study fails to establish that there are any genuine grounds for concern with the operation of the payment card industry in Australia. Without carrying out this test, it is simply impossible to demonstrate that any interventions which alter the status-quo will improve welfare.

Where commercial arrangements arise through private sector innovation and initiative and prove to be successful, a principled, well-supported case must be made before regulatory intervention can be considered justified. Before changing a vigorously successful and growing system from which many segments of the economy clearly derive great benefits, it should be demonstrated that the changes will be welfare-enhancing.

The onus for establishing the superiority of the proposed alternatives therefore lies with the RBA and the ACCC. This response identifies a significant number of conceptual and empirical flaws in the Joint Study. Collectively, these errors and omissions demonstrate that the Joint Study falls well short of satisfying the necessary burden of proof.

Turning to the specifics of the analysis, this response shows that the Joint Study:

i.   overly focuses on the resource costs of alternative payments systems, without giving adequate consideration to the benefits each of these systems brings;

ii.   relies on an incorrect assumption that interchange fees are the source of bank margins;

iii. does not present an alternative basis for determining interchange fees which is based on sound economic analysis;

iv.   takes insufficient account of the reasons merchants accept cards and the benefits of card acceptance;

v.   presents an internally inconsistent assessment as to what would happen if the "no-surcharge" rule were prohibited;

vi.   in some cases suffers from a lack of clear evidence, e.g. as to the relative costs of cash and payment cards to merchants, while in other instances it misinterprets the evidence that is presented – for example, the data which appears to contradict its stated view that payment networks in Australia are mature; and

vii. does not pay sufficient regard to the implications for the competitive provision of credit cards vis-à-vis store cards and charge cards of adopting the recommendations advanced.

These are not minor points. Rather, they represent material and fundamental weaknesses in the analysis underlying the Joint Study. There is therefore an appreciable risk that implementation of the Joint Study's proposals will cause more harm than good.

This paper provides a detailed consideration of the issues raised by the Joint Study. It begins in section 2 by setting out some basic economic concepts related to networks, an understanding of

which is necessary for what follows. Moving to the substantive issues canvassed by the Joint Study, section 3 sets out those aspects of the Joint Study's approach that can be accepted – that is, which fall outside of the area of controversy.   The subsequent sections then address the three main aspects of card systems discussed in the report, namely:

i.   the level of interchange fees,

ii.   the welfare effects of the no-surcharge rule, and

iii.   the membership rules.

In each case, we:

i.   identify the errors in the assumptions and analysis adopted by the Joint Study and the way these errors contribute to flawed conclusions;

ii.   demonstrate that existing practices are consistent with standard principles of network economics; and hence

iii.   show that a case for regulatory change or intervention has not been established.

2. Network Economics: An Overview of Basic Principles

The purpose of this section is to briefly introduce several concepts that are essential to understanding the economic forces at work within payment card networks.

It is useful to start from the premise that ideally, economic activity ought to be organised as *efficiently* as possible. "Efficiency" here refers to a state of affairs in which it is *not possible* to make one person better off, for example by increasing their disposable income, without making another person worse off. A weaker form of the same idea requires that if some change results in one person gaining more than another loses, that change increases efficiency because it is possible for the "loser" to be compensated while still leaving the "winner" better off. This is the efficiency notion that underlies most micro-economic policy analysis, including of competition issues. It is concerned exclusively with aggregate welfare, leaving the distribution of that welfare as between individuals to other policy instruments, and in particular to tax policy and to social expenditure.[1] It is important to distinguish between changes that redistribute welfare between agents without creating any additional surplus, and those that do actually increase aggregate welfare. The former are merely *transfers* between agents; they do not increase the total available surplus.

As a general matter, efficient outcomes are achieved by ensuring that the incentives economic agents face reflect the costs and benefits their decisions cause. However, specific challenges are involved in securing efficient outcomes in network activities.

Three concepts are crucial here: these are network effects; network externalities; and network pricing. Underlying all three concepts is the idea of a network, which can be defined as a set of compatible goods and/or services, such as the set of services provided under the VISA brand for example.

A *network effect* exists when, other things being equal, economic agents prefer to join a larger network. For example, the fact that more merchants accept VISA than AMEX means that consumers facing the same terms and conditions of use for each card would prefer to carry a VISA card than an AMEX card. Similarly, the fact that there are more VISA cardholders makes accepting VISA cards more attractive to merchants than accepting AMEX cards, even if the terms and conditions of accepting these cards were identical.

Network effects can also arise indirectly, when users anticipate that joining a larger network will given them access to lower prices for one or more services, or to more services at the same price. For example, if people believe that installing the Windows operating system will allow them to choose from a larger range of application software, an indirect network effect is generated.

Network externalities are generated from direct network effects. A network externality is an unpriced spillover effect that arises when agents join or leave a network. Suppose, for example, that a merchant decides to stop accepting a particular credit card. All consumers carrying that card suffer an uncompensated drop in wealth (broadly defined) because they no longer have the option of using their card to make purchases from that merchant. This is what economists call a negative externality.

---

[1] *This is not to suggest that distribution is unimportant. Rather, it merely recognises that micro-economic policy, and notably competition policy, is a poor way of achieving distributional goals. The Trade Practices Act 1974 states (at s.2) that its object is "to enhance the welfare of Australians": not of a particular group within the community. Reflecting this, it is important to clearly distinguish a concern with aggregate welfare from a concern over its distribution. The former is directed at curing inefficiencies; the second at effecting transfers.*

T.1

A similar effect operates in reverse when new agents join a credit card network: in this case, all agents on the "other" side of the network receive an increase in wealth which they do not pay for directly. All merchants that accept the card are immediately better off when an additional cardholder joins, and all cardholders are immediately better off when an additional merchant joins.

Network pricing is designed to compensate for, or equivalently to "internalise", network externalities. Direct network effects, which are the only ones that create externalities,[2] arise from the production of what might be called a "two-sided" service. In all such cases, two different types of economic agent are involved in the production of a unit of service. To appreciate this, consider the examples presented in Table 1 of services in which direct network effects are generated.

In each of these cases, the value of the service to Type 1 agents depends on the number of Type 2 agents that are also using the service (i.e. connected to the network). As a result, there are direct network effects, and network externalities, at work in each of these cases. It is standard practice to set different service prices to each agent type for all of these services. Moreover, it is not unusual for type 2 agents to receive the service entirely free of charge.

Table 1:
Network Services and Agent Types

Service	Agent Type 1	Agent Type 2
Telephony	Calling party	Called party
Car Fair	Car seller	Car buyer
Credit Card	Merchant	Cardholder
Business Exchange	Sellers	Buyers

In most modern economies, all of the services in Table 1 are provided in highly competitive markets. As a result, inefficient pricing structures are unlikely to survive. Suppose, for example, that an established car fair operator, competing against several others, decided to start charging an entry fee to buyers. This will drive price sensitive buyers to competing car fairs, even if these have no more sellers. As the stock of buyers gets reallocated to competing car fairs, some sellers will also switch to alternative market operators. The size of the starting network (the initial car fair) will shrink, and it will therefore provide less value to all of its users. This value reduction may set off a further round of switching, in which case the pricing inefficiency can generate a downward spiral of value for the initial supplier.

---

[2] *Indirect network effects do not generate externalities, because their impact is mediated through the price system. Windows users receive tangible benefits, in the form of greater software choice, as a result of the indirect network effect associated with computer operating systems. There is no unpriced spillover in this case.*

To avoid such an outcome, providers of network services in competitive markets need to set prices that adhere to a few simple principles, none of which will be unfamiliar to anyone who has studied the problem of setting efficient prices for a firm that produces multiple outputs.

First, the total revenue earned under the pricing structure cannot be materially greater than the total cost of providing service to both sides of the network. If this were not the case, a competitor would be able to capture market share by offering a similar service at proportionately lower prices.

Second, no agent can face a price that exceeds the value it receives from the service. There is nothing remarkable about this principle, which applies to any competitively supplied service, whether based on a network or not. No agent will freely purchase a service for which the price violates this principle.

Third, each agent type should contribute towards the common costs of the system in inverse proportion to their demand sensitivity. The common costs of the system are those that cannot be directly attributed to the actions of any particular agent, and must therefore be shared across users in some way. Sharing these charges on the basis of demand sensitivity – a principle long established in economics[3] — is efficient because it minimises the extent to which individual consumption decisions are distorted by the need to cover common costs.[4]

These principles together define a way of dealing with a situation in which directly aligning the incentives individuals face with *private* costs and benefits will not yield an efficient outcome. Rather, these incentives need to take account of the wider consequences each individual's actions have for the welfare of others: that is, with *social* costs and benefits. Here, "private" refers to the consequences that pertain directly to the individual taking the decision, while "social" refers to the sum of the consequences for that individual and for the others affected by the decision.

In these situations, the "first best" solution – which corresponds to the outcome that would be mandated by an omnipotent and omniscient social planner – is usually unattainable. Rather, what must be sought is a "second best" solution that, accepting the constraints that bear on decision makers (including the fact that they do not have access to perfect information), seeks a pragmatic approximation to efficient outcomes. So too, in credit card networks, systems such as VISA must seek to align the incentives network participants face with the wider impact of their actions on the network.

Thus, the cost of making any single credit-card purchase cannot be directly attributable to any single agent. Rather, two different agents need to have made a total of three distinct decisions before a credit card transaction can be effected. The buyer and seller must both have decided to join the card network, and the buyer must also decide to use the credit card for the transaction, rather than any one of (possibly) several alternatives.

---

[3] *This principle is often referred to as Ramsey-Boiteaux pricing. It was first systematically elaborated by the French economist, Jules Dupuit in the 1840s.*

[4] *If the common costs involved in producing a service were loaded on to those with a low willingness to pay, they would significantly reduce their consumption of that service – even though they valued it at more than the marginal cost of its supply. Their welfare would diminish by more than the increase in welfare of those now bearing a lower share of the common costs. The outcome would consequently be inefficient, in that sense that a reallocation of the common costs could yield an increase in welfare overall.*

The general principle for sharing the costs of the transaction between the merchant and the cardholder is that (all else being equal) each share should be larger, the less price sensitive the agent is. Further, if the demand of each agent depends on choices made by the other, the efficient cost-sharing rule should take this dependency into account. In the case of credit cards, this dependency operates primarily through the decision to join the network, since buyers are less likely to join if very few merchants accept cards. Nonetheless, this membership decision has a strong and obvious effect on card usage as well, since a card cannot be used in any given transaction unless both cardholder and merchant have joined the network.

Demand for card services by merchants and cardholders are therefore strong complements. If the price to merchants rises enough, card acceptance will fall and this will make cardholding less attractive, other things being constant. In this complementary demand case, it may well be efficient to set negative prices for the most price sensitive agent: the cardholder. This, of course, is a close approximation to reality in many credit card networks, including those owned by a unitary company (such as AMEX).

Zero or negative pricing is also quite common for "type 2" agents in the other networks presented in Table 1. For example, it may be in the interests of sellers in business exchanges to pay prices that exceed what could, outside of a network service environment, be interpreted as their "stand-alone" cost. The resulting funds would then be used to provide additional services to the buyers, conditional on their participation in the network.

In such a case, no subsidy is involved from type 1 to type 2 agents because the network service would not exist in its present form in the absence of the negative price to type 2 agents. In the context of two-sided network services, it is meaningless to speak of the "stand-alone" cost of serving one type of agent, since there is no service provided unless both types of agent are served.

Consider the implications of negative pricing to type 2 agents. Suppose that initially the entire network service is provided by a single firm, with different divisions of the firm serving each type of agent. Clearly, each of these divisions should have different targets. For example, it would clearly be foolish to require the division serving type 2 agents to earn revenue if negative pricing to type 2 agents were efficient. Rather, this division should be funded from the revenue earned by the type 1 agent division. This requires an explicit or implicit *transfer* of funds between divisions. Closed network payment cards, such as AMEX, operate in this manner.

A similar transfer is required when the network service is provided in a decentralised way by stand-alone firms serving one or both types of agent. This occurs in credit card associations, where the association's members compete for market share in the service of each type of agent. In this case, the transfer is referred to as an interchange fee.

The interchange fee performs exactly the same function as the internal transfer between the divisions of a single firm such as AMEX. Note also that this would still hold true even if AMEX chose not to organise itself into type 1 and type 2 divisions. Irrespective of the way a closed network credit card firm is organised, type 2 agents (cardholders) receive an implicit transfer from type 1 agents (merchants).

This brief discussion highlights several important points. First, credit card networks offer a joint service to two types of agents with different demand characteristics. Second, in common with other such services, it can be efficient for one type of agent to receive the service at a significantly lower direct price than the other. Third, when this is the case, it can be efficient to *transfer* some revenue earned from serving one type of agent to the supply of services to the other – in the case of card networks, from the supply of service to merchants to the supply of services to cardholders. Fourth, such transfers are indeed present in closed credit card services operated by single companies as well as in open credit card associations.

6

3. The Starting Propositions

The Joint Study identifies three important features of four party payment card systems. These are: first, the relevance of network externalities; second, the central role of interchange; and third, the need for an "honour all cards" rule.

Recognition of these features is indeed necessary to form a reasonable perspective on the issues considered by the Joint Study. These features then have important implications for other issues that are critically examined by the Joint Study, namely:

i.   the process and methodology for determining interchange;

ii.   the no-surcharge rule; and

iii. membership rules.

3.1 Card payment systems subject to network externalities

The Joint Study recognises the presence of network externalities in credit card payment systems. This is an important concept and one that is fundamental to the determination of issues relating to the appropriate level of the interchange fee.

As the Joint Study puts it:

> *'A defining economic characteristic of most networks, at least in their formative period, is that the value of the network expands more than the value of connection to a new user.'*

A clear example of this in the case of credit card networks is that the more participating merchants there are in a credit card scheme, the more valuable membership in the scheme becomes to the individual consumer, thus inducing more consumers to subscribe. Similarly, the more consumers that subscribe to the scheme, the more valuable membership becomes to the individual merchant, thus inducing more merchants to subscribe. However, in making a decision about whether or not to subscribe, a consumer or merchant is only likely to take into account his or her private benefits, and will not consider the increased value which his or her membership of the network brings to other consumers or merchants.

The qualification "…*at least in their formative period...*" in the above quote is not supported by analysis in the Joint Study, and cannot be so supported in the case of credit card systems.[5] This qualification raises the possibility that at some level of membership of such a system, marginal changes in membership have no external effects. To understand why this is not so, consider the effect of one merchant leaving a fully subscribed network. This action would affect the value that cardholders as a whole derive from the network, yet their welfare is not considered by the merchant in deciding to leave. The result is the creation of a (negative) network externality. There is no reason to suppose that these effects vanish as network membership increases.

---

[5] *It is possible when some networks reach a certain size for there to be no externality associated with the marginal member. For example, the marginal member of an Internet chat group may not bring any benefits to a chat group with a large enough subscriber base, if the positive externality of a member is solely due to the diversity they bring. As demonstrated below, this is not the case for credit card networks.*

3.2 Interchange is necessary

Given these network effects, the Joint Study acknowledges that the interchange fees paid by one participating financial institution to another can allow a payment network to grow beyond a size it would attain if the participating institutions involved earned revenues solely from direct charges on customers. This is because the fees can serve as a redistribution mechanism: through the fees, the benefits received by one of the parties, which are more than sufficient to induce it to join the network, can be shared with other participants to create adequate incentives for more of them to participate. This sharing process, which depends on the existence of an interchange fee, can increase the value of the resulting network to **all** participants beyond the level that would otherwise be achieved. For that reason, it is fully consistent with economic efficiency. As noted in section 2, transfers between different types of users are a very common feature of network services supplied in highly competitive environments.

The direction and magnitude of the interchange fee required to achieve the efficient outcome is, however, a complex matter. It depends on a myriad of factors including systematic differences in the cost and demand elasticities between issuers and acquirers, and the nature of competition in the issuing and acquiring functions. Nonetheless, based on an examination of the facts, *ex post facto* hypotheses can be offered regarding why, in the case of credit card networks, the interchange fee is paid by the acquirer to the issuer.

The Joint Study suggests one possibility – that the balance of benefits is on the side of the merchants, such that it would be in the merchants' collective best interests to pay a higher merchant service fee to acquirers, some of which is then passed on in the form of the interchange fee paid by acquirers to issuers.

The Joint Study goes on to note that the benefits accruing to merchants by the creation of a credit card network of the optimal size can only be obtained if issuers agree to incur several types of cost. The first is the funding cost of the interest free period, which leads to higher usage by cardholders and thereby benefits participating merchants. The second is the cost incurred by the issuer in guaranteeing payment for goods to merchants; this benefits merchants because it allows them to avoid the costs of fraud and fraud prevention, credit losses and risk control costs. The third is the processing cost incurred by the issuer. It is argued below that the cost of loyalty programs is also a legitimate cost of issuing: it leads to higher usage by cardholders and thereby also benefits participating merchants.

The Joint Study acknowledges that the interchange fee allows the total system costs, which yield net benefits to all of the end-users, to be recovered efficiently.

3.3 Honour all cards rule is necessary

One of the most important requirements for any successful payment system is that the means of payment be widely accepted. Because open credit card networks result in cards being issued by numerous members around the world, the integrity of the system – the value of the "brand" associated with the network – is best served by ensuring that all cards on issue are treated equally. This is achieved through the "honour all cards" rule. Under this rule, any merchant that participates in a particular system, say VISA, must accept any VISA card, irrespective of which bank issued it. As a result of this rule, merchants cannot choose to accept VISA cards when tendered by rich-looking people but not by people who look poor or disadvantaged in some way, when tendered by people who do not have debit cards but not when tendered by people who do have debit cards, and so on.

This rule applies not only domestically but also internationally: it ensures that VISA cardholders from (say) Finland can use their VISA cards at any VISA-accepting merchant anywhere in the world.

Similarly, it ensures that Australians can be confident of being able to use their cards in the next suburb or town, interstate and overseas – even if the card is not issued by a major Australian bank, but by their small credit co-operative or a small local deposit taking institution.

This powerful rule thereby helps give the open credit card systems the international ubiquity which has been so vital to their existence and growth and enables even the smallest issuer to compete on equal terms with the largest.

The Joint Study notes the fundamental importance of the "honour all cards" rule to a transnational, networked payment system. On page 56, the report notes:

> *"The study recognises that there are good reasons why card issuers should have financial standing. Credit card networks can only operate if all cards are accepted. Under the 'honour all cards' rule, merchants signing up to MasterCard, for example, must accept all MasterCard credit cards; they cannot refuse to accept a MasterCard because the card is issued by a bank they have not heard of."*

The existence of the "honour all cards" rule has important implications for some alternative methods of determining interchange rates. In particular, any bilateral bargaining approach to setting these rates would need to confront the fact that any issuer (even a very small one) would have hold-up power in bilateral negotiations as a direct consequence of this rule.[6] This illustrates the fact that the design of card networks is a holistic task in which each element must mesh with all others.

3.4 Implications of the starting propositions

While these starting propositions, which the Joint Study accepts, are relatively straightforward, they have far-reaching implications. Most importantly, when properly applied in the light of the available evidence, they are inconsistent with the Joint Study's apparent conclusions, namely that:

i.   interchange fees are too high;

ii.   credit card networks are mature;

iii.   as a result, reducing the interchange fee will increase welfare;

iv.   the no-surcharge rule leads to over use of credit cards;

v.   welfare will therefore increase if the no-surcharge rule is abolished; and

vi.   all else being equal, merchant costs will fall if non-deposit taking institutions are permitted to act as merchant acquirers in credit card systems.

In the following sections these inconsistencies and their sources are demonstrated.

---

[6] *If a bilateral bargaining approach were adopted instead of centralised interchange agreements, then the honour all cards rule would give even small issuers inordinate bargaining power over large acquirers. This power to "hold up" acquirers arises because each acquirer's merchants are obliged by the honour all cards rule to accept credit cards issued by any issuer. If the terms of interchange must be established bilaterally between each issuer and each acquirer, then any issuer can drive a hard bargain with each acquirer. An acquirer that fails to agree with any single issuer is unable to meet the honour all cards requirement, and therefore would be unable to participate in the payment system at all. Faced with this prospect, an acquirer will be willing to accept a very high interchange fee since the alternative is no profit at all. The level of such a fee can confidently be expected to be well in excess of current credit card interchange fees. These issues are explored further in Small and Wright (2000), "Decentralised Interchange Fees in Open Payment Networks: An Economic Analysis", NECG and University of Auckland, reproduced here as Attachment 1.*

4. Are Interchange Fees Too High?

The claim that interchange fees are "too high" forms a central element in the Joint Study. However, the reasoning underpinning this claim is not well articulated. A range of arguments, some seemingly inconsistent, are deployed, but no overall logic is set out. Addressing the Joint Study's argument therefore requires some restatement of the Study's chain of reasoning.

At its simplest, the Study's analysis runs along the following lines.

To begin with, the Study accepts the fact that credit card systems are characterised by network externalities. It also states that at least in the early phases of these systems' development, capturing these externalities can require a mechanism that balances the incentives for system participants to expand the network on the issuing and acquiring sides of the market. As a result, the Study accepts that – again in the development phase of the system – the efficient interchange fee may well be above resource cost.

However, the Joint Study then asserts that the development phase is now over: the systems are, using its word, "mature". Given that they are mature, the Study says, interchange fees should now be based on resource costs. The Study then attempts to quantify these resource costs and concludes that the fees are significantly above this benchmark. The Study sees the high level of fees as distorting: most notably, of competition as between credit card systems on the one hand, and debit cards on the other. (The "no-surcharge" rule, discussed in the next section of this document, is cast by the Study as a significant factor allowing this distortion to arise and persist). The result of this distortion, the Study asserts, is that credit cards are over-used and debit cards under-used as compared to the economically efficient outcome.

How can it be known that credit cards are over-used and debit cards under-used relative to the requirements of economic efficiency? The answer, the Joint Study suggests, is simple: debit cards have lower resource costs than credit cards; hence, efficiency requires that they be used more (though the Study never defines quite what "used more" actually means). The Joint Study then notes that credit card use has expanded relative to the use of debit cards — and hence infers that the market for payment systems is working poorly.

Set against this inefficiency, the Study argues that reducing interchange fees would not undermine the credit card systems — because they are "mature", such a reduction would not cause undue shrink-age. However, it would yield benefits as the incentives for further, presumably inefficient, expansion would be curbed. The Study therefore outlines two alternative approaches that could be used to set interchange fees at levels that it believes would be more appropriate.

This line of argument raises a number of obvious questions.

To begin with, are the credit card systems indeed mature? And even if they are mature, does the Joint Study's inference that this justifies setting interchange fees at resource cost follow as a matter of economics?

Second, the Study seems to believe that the expansion of credit card use relative to debit card use is evidence that interchange fees are inefficiently high: again, can any such inference be validly drawn, even if it is the case that the resource costs of debit cards are lower than those of credit cards?

Third, given the answers to the questions above, are the alternative approaches to determining the interchange fee that the Joint Study sets out economically defensible?

This section will show that the Joint Study has provided the wrong answer to each of these questions. The remainder of this introduction provides an outline of the analysis, which is developed more fully in the following subsection.

10

Starting with the state of the industry, it is clearly inaccurate to characterise the credit card markets in Australia as mature. Compelling contrary indications are provided by empirical evidence and discussed in section 4.1. But even if the systems were mature, the inference the Joint Study draws that this requires setting the interchange fee at resource costs is simply mistaken as a matter of economics.

In comparing the relative usage of credit and debit cards, it is essential to recognise that social welfare, rather than resource costs, should form the basis of policy prescriptions for payment systems. The Joint Study implies that minimising resource costs is equivalent to maximising welfare, but this is only the case when the alternatives being costed provide equivalent benefits. The pitfalls with this approach are demonstrated in section 4.2 below with an example in which welfare benefits of credit card versus debit card payment systems are compared.

More fundamentally, the Joint Study fails to define precisely what is meant by the claim that current incentives for card users are encouraging "over-usage" of credit cards. A proper definition of over-usage needs to define what is being over-used (is it the number of times cards are used or is it the number of cardholders?) and the criteria by which over-usage is judged. It is explained in section 4.3 below that even if some cardholders use credit cards when debit cards would be more socially efficient, it does not necessarily follow that welfare would be enhanced by lowering the interchange fee so that cardholders face higher costs and merchants lower costs of usage. Because any reduction in interchange fees will also affect the membership and usage decision of all other cardholders and merchants, any welfare analysis must also take these changes into account.

Turning to the policy recommendations, the Joint Study's presumption that little harm would be done to the open credit card systems by a regulated reduction in interchange fees does not withstand scrutiny, as shown in section 4.4 below. That conclusion relies on several assumptions that simply do not hold, and overlooks the important dynamic of competition between closed (3-party) credit, charge and store card systems and open (4-party) credit card systems.

None of this explicitly addresses one of the few empirical statements the Joint Study makes: namely that interchange fees are high relative to resource costs. As will be clear from the discussion in section 2, from an economic point of view, there is no reason for interchange fees to equal resource costs if efficiency is the goal being pursued. In focussing on what it claims to be the resource cost of issuing, the Joint Study has completely overlooked the fact that the credit card service is a network service provided simultaneously to two different types of agent. As discussed in section 2, it is economically meaningless to suggest that the cost of serving cardholders can be considered separately from the cost of serving merchants.

Even ignoring this fundamental point, however, and assuming (quite incorrectly) that the resource cost of issuing is separately identifiable, the Joint Study remains in error in that it has overlooked some important resource costs. It thereby greatly over-states the degree to which interchange fees could, under even the most generous interpretation, be said to exceed resource costs. Section 4.5 sets out the errors the Joint Study makes in this respect.

T.1

4.1 Network maturity

The Joint Study asserts that credit card networks in Australia are now mature systems. As a result, the Study suggests, it is no longer necessary to set interchange fees at a level which promotes further growth in network membership and usage. This view is expressed in two passages.

> *"The intent of an interchange fee is to ensure that network benefits are taken into account by overriding the usual market mechanisms under which buyers and sellers compare private costs and benefits. While this may be justified in the early stages of development of a payment network, the weakening of normal price signals in a mature network can lead to higher interchange fees than are necessary to establish and maintain the viability of the network."*

And:

> *"In the early stages of credit card networks, consumers may well have responded to merchant surcharges by not taking up and using credit cards; at that time, consumers may have needed incentives to test the benefits of using these networks. However, the credit card networks in Australia are now mature systems."*

The question of whether a network is mature is inherently difficult to assess. The Joint Study's rationale for the view it adopts consists of several qualitative and subjective observations contained in a single paragraph at pages 54-55:

> *"Credit cards are second only to cheques as a payment instrument and the number of credit card payments has doubled over the past four years. Unlike a new instrument, consumers are comfortable using credit cards and well aware of their benefits and convenience. Merchant acceptance has also grown. Credit cards are being used for new classes of payments, such as grocery purchases and utility bills, and they have become a predominant form of payment over the phone and Internet."*

These observations fall well short of establishing that credit card networks are mature. A more detailed assessment is therefore needed of what is meant by maturity, and what implications do or do not flow from a finding that a network is mature.

4.1.1 Does the evidence suggest maturity?

As noted above, there is no well-defined interpretation of what it means for a card network to be "mature". To address the empirical validity of the Joint Study's view, it is therefore appropriate to consider a range of indicators that could possibly form a basis for the maturity hypothesis. None of these indicators support the view that credit card networks are mature.

To begin with, consider the product life cycle approach to product maturity. Within this approach, "maturity" refers to a stage where sales growth has stopped or slowed significantly. This will occur when only limited scope remains for the product to expand within its existing applications, and the opportunities for new applications have been largely exhausted.

By this product life cycle definition, credit card networks are clearly not mature. To begin with, as evidence presented in the Joint Study shows, credit card networks in Australia are currently in a strong growth phase, which has shown no signs of tapering off as would be required to provide any support for the maturity hypothesis. Second, it is apparent that new applications of credit cards, such as those involving grocery, utility, telephone, and Internet payments, are being developed. Within this group of new applications, Internet payments alone are a very important source of new growth in card

usage. Despite the rapid uptake of e-commerce, only 31% of Internet users have made an online purchase (BCG 1999 Consumer Survey), and usage of the Internet more generally is still growing very rapidly (the growth rate is estimated to be between 46% and 125% per annum by the OECD[7]). These figures indicate that there will be substantial growth in Internet-mediated buying over the next few years. Indeed, current forecasts predict international revenues in e-commerce will be between US$1.2 and US$4.6 trillion by 2003.[8] Since credit cards are the major form of payment over the Internet, the projected rates of growth in Internet buying are likely to further increase the usage of credit cards. Consequently, whether one looks at existing usage in conventional environments (where the data show strong growth) or at the potential growth of new applications, the product life-cycle view of credit cards must lead to the conclusion that credit cards are not a mature product.

Alternatively, a statistical view of "maturity" could be used to argue that credit card networks in Australia have reached saturation, or 100 per cent of the feasible market penetration. Again, this interpretation appears inconsistent with the evidence. The vigorous growth in cardholder numbers, merchant outlet numbers, transaction numbers, and transaction value clearly indicates that the market is not saturated. Moreover, the Joint Study reports that around one-half of adult Australians do not have credit cards, suggesting that the market is far from saturated.

The concept of a "mature" product can also be used, in a marketing context, to distinguish between well-established products and those which are novel, untested or unfamiliar to consumers. The unfamiliarity of a new product to the base of potential customers may make it difficult to sell, and once this unfamiliarity can be overcome promotional strategies are likely to change. In at least one instance the Joint Study suggests that this is the meaning it intends:

> *"Unlike a new instrument, consumers are comfortable using credit cards and well aware of their benefits and convenience."*

However, although it is probably true that current cardholders are comfortable using credit cards and well aware of the benefits they derive by doing so, it is not clear that consumers who are not credit card holders are so well aware, and it could certainly not be said that non-credit card holders are comfortable using credit cards. So while "*cardholders* are comfortable using credit cards and well aware of their benefits and convenience" this statement is really a tautology: it could be said of any product at any stage in its life-cycle, from the first unit sold to the last, that customers who use the product are likely to be comfortable using it and to be well aware of the benefits of doing so. This, however, is of no practical relevance in the present circumstances since it does not in any way bear on the issue of what might happen if the existing arrangements were altered. As a result, this marketing based definition of maturity does not assist in assessing the issue at hand.

---

[7] *Coppel, J., 2000, E-Commerce: impacts and policy challenges, OECD Economics Working Paper No 252*

[8] *It is worth noting that without credit cards, the simultaneous development of an online payment system and of e-commerce applications that make use of it would have been a major obstacle to the development of e-commerce. Only now that e-commerce has taken-off, have new Internet based payment systems started to become viable. Major organisations such as AOL, AT&T, Citibank, Microsoft, and Yahoo have recently developed new payment instruments such as Internet bill payments, Internet debit payments, person-to-person payments, and pre-paid web accounts for micropayments.*

Finally, maturity could be assessed by reference to some optimality criterion, in which case it may be possible to determine whether the credit card networks in Australia have grown to, or beyond, their optimal size. This approach is inherently difficult, and would require a clear specification of the criterion by which optimality is to be assessed. The Joint Study has not undertaken any analysis of this kind to support its view that absolute over-expansion has occurred or may occur in the absence of corrective measures.[9]

In short, although "maturity" seems important to the Joint Study's analysis, the Study itself neither defines the concept nor provides any evidence to suggest that it is indeed relevant to Australian circumstances.

4.1.2 Resource cost interchange fees in a mature network

Even had the Joint Study established that the networks were mature, would the inference that interchange fees should be set at resource cost follow?

Here the Study's reasoning is hampered by its failure to use any analytical model of efficient interchange fees. This is not because models of this kind do not exist – rather, it is because the Joint Study has chosen to ignore them.

To correct the Joint Study's error, it is therefore useful and important to start from such a model. Consider for instance the model of interchange developed by Professor Richard Schmalensee, Dean of the Sloan School of Management at MIT,[10] in which a specific expression for optimising interchange fees is derived. According to this model, all else being equal, the higher the elasticity of demand by cardholders relative to merchants, the higher will be the optimal interchange fee flowing from acquirers to issuers. This result is likely to be quite general. Intuitively, the network will be optimised when it raises the largest contribution to system costs from the most inelastic demand and uses this to promote the more elastic demand. The greater the *difference* in demand elasticity the greater the share of system costs that is optimally allocated to the acquiring side of the market.

Is it likely that "maturity" would be associated with a pattern of demand elasticities that warranted setting interchange fees at resource cost?

Far from it: rather, maturity (at least as it is defined by the Joint Study) seems quite inconsistent with a view that interchange fees should be regulated to cost. Thus, according to the Joint Study (p.15),

> "*Nearly 60 per cent of respondents with income below average weekly earnings do not have a credit card, and the proportion falls as the income of the main earner rises.*"

---

[9] *The Joint Study does argue that credit card use has over-expanded relative to debit card use. However, the argument it puts in this respect would, if it were valid (and as shown below, it is not) apply at any absolute level of use of credit cards. It is in no way relevant to the issue of whether the credit card networks are "mature". Even less can the mere assertion of over-use be regarded as evidence of maturity. Equally, it is completely circular to argue, as the Study seems to, that credit cards are mature because they are over-used – when the evidence of over-use depends on a finding of maturity.*

[10] *Schmalensee, R., 1999, Payment systems and interchange fees, mineo, Sloan School of Management, MIT.*

This observation suggests that new cardholders are increasingly from lower income groups; that is, people with a lower willingness to pay and a generally more elastic demand profile. The Joint Study also claims (p.48)

> *"But as credit card networks become more widespread, accepting a credit card becomes a condition of doing business rather than a means of gaining an advantage over competitors."*

This claim ignores the alternatives to credit card acceptance that are available to merchants, including the provision of alternative credit facilities and the targetting of non-cardholders with lower prices or other marketing initiatives. More fundamentally, it suggests that credit card acceptance is a necessity rather than a choice for merchants. If this really is the case, merchant demand must have become less elastic over time.

The combination of increasing elasticity of demand on the cardholder side and decreasing elasticity of demand on the merchant side has implications for the optimal interchange fee. As was noted above, each agent type should contribute towards the common costs of the system in inverse proportion to its demand sensitivity. If merchants are really "locked-in" to accepting credit cards, then this suggests a high interchange fee would be desirable. As a result, the efficient interchange fee in such a network may well be **higher** than that which would be efficient at lower levels of card penetration.

The fundamental flaw in the Joint Study's argument is that there is nothing in the theory of interchange that suggests that just because a market is mature it will no longer be subject to network externalities. Adding a cardholder to the network provides benefits to merchants, whether it is the first cardholder or the last cardholder to join the system. Similarly, when an additional merchant decides to accept credit cards this provides additional benefits to cardholders, even if there are already a large proportion of merchants in the network. The standard theory of interchange in which the fee is optimally set to align marginal network benefits with marginal network costs still applies in a mature market. The Study's inference that maturity means that the fee should fall is consequently simply incorrect.

4.1.3 Summary of theory and evidence on maturity

In summary, there is no evidence to suggest that credit card networks in Australia are mature. Even if card networks were now mature in the sense described by the Joint Study, the established theory of interchange does not support the proposition that interchange fees should now be reduced.

4.2 Will welfare increase if the interchange fee falls?

It should be noted that the Joint Study makes no explicit claims about the aggregate welfare changes expected to result from any given policy. (Indeed, the term "welfare" does not appear in the Study). Nonetheless, the Joint Study appears to conclude that Australian society would be better off if interchange fees were lower.

This conclusion is based primarily on a comparison between the resource costs of issuing credit cards relative to debit cards. The Study asserts that the resource costs of using debit cards are lower than those of using credit cards; and on that basis seeks a way of encouraging greater use of the former relative to the latter. Reducing the interchange fee is the mechanism by which this goal is to be achieved.

This is because lower interchange fees would reduce the level of resources that issuers can devote to customer acquisition and retention activities. In particular, if the interchange fee were low enough, issuers would be unable to offer loyalty program credits to cardholders.

This would reduce the incentive to hold and use credit cards, with the result that cardholders on the margin would choose either not to hold a card, or to use one less frequently. The Joint Study anticipates this to have a negligible impact on merchant acceptance because merchants are viewed as having no real choice in whether they accept cards or not. Moreover, the Joint Study does not expect substantial reductions in the current levels of card usage.[11] Rather, it seems to anticipate that the main effect of reducing the interchange fee will be felt through future growth in the payment card market, which it sees as being directed more towards debit cards. In this way, debit cards could grow over time relative to credit cards without the usage of credit cards falling much in nominal terms.

Irrespective of the plausibility of this mechanism, the motivation for using it stems from the proposition that society would be better off if debit cards were more popular than credit cards, because debit cards have lower resource costs.

This proposition is only true under extreme conditions, which do not apply to the case at issue. More specifically, the claim that resource costs comparisons are equivalent to a welfare comparison ignores the benefit side of the social welfare equation. Even if it is accepted that resource **costs** are higher with credit card usage, it is certainly also the case that that **benefits**, to both cardholders and merchants, are larger too.[12]

---

[11] This could be the case if relatively infrequent users of cards were the ones on the margin of the card-holding decision. This seems a reasonable assumption during normal network growth periods – relatively low value users are likely to be the last to join the network. If, however, existing cardholders perceive a sudden drop in the value they receive from the card relative to the terms offered by competing means of payment (such as 3 party card networks), the first to switch may well be the high-usage cardholders, since for these people the loss of usage related benefits will be substantial. There are, in other words, sound reasons to question the assumption that card usage will not be greatly affected by this change.

[12] There are many attributes of credit cards that support the view that they provide greater benefits to consumers than debit cards. Credit cards provide ready access to credit, an interest-free grace period, the ability to pay only once a month thus reducing the time spent sorting out finances, the ability to provide security for future payments, wide acceptance internationally with low transaction costs, and the ability to earn loyalty benefits. For this reason, even if one finds that after subtracting from the total costs of credit transactions the costs of providing transfers from merchants to cardholders and banks that the remaining costs of credit transactions were higher than debit card transactions, it does not necessarily follow that debit is more efficient than credit. Credit cards still deliver several other benefits over debit cards; most obviously, the credit functionality that increases consumer utility through the ability to better time purchases, and increases merchant profit through increased sales.

[13] These diagrams implicitly assume that the producer surplus (i.e. bank profit) arising from the provision of each of these services is similar. If there were more producer surplus from one service than the other, this should be taken into account in a full welfare analysis. While eschewing formal welfare analysis, the Joint Study, in concentrating solely on end-user effects, implicitly places a zero weight on producer surplus.

Figure 1

Welfare from different products
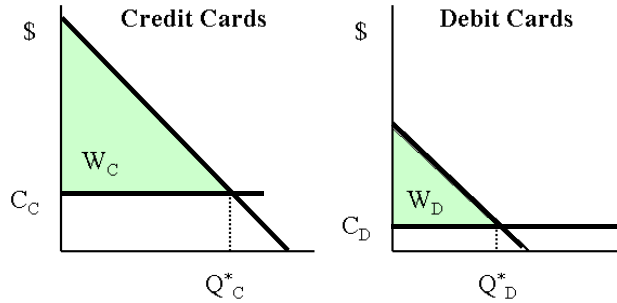
A reasonable framework for thinking about the net social benefits (welfare) of these services is presented in Figure 1, which reflects the fact that credit cards and debit cards are different products, and as such their demand curves are different.[13] This diagram illustrates the simple proposition that a service with higher unit costs can also produce higher aggregate welfare.

The horizontal lines in Figure 1 represent the unit cost of supplying each product while the downward sloping demand curves represent the total demand for each product arising from cardholders and merchants together. A credit card service is generally more costly to provide than a debit card service, so C_C is drawn as being higher than C_D. This is partly because it includes additional benefits such as an interest-free period, cardholder rebates and loyalty programs – all features that likely allow merchants to make more sales. On the valuation side, it is clear that consumers much prefer credit cards to debit cards when there is no material difference in the price of these services. This implies that the demand curve for credit cards is shifted further out than the demand curve for debit cards.

One measure of the total welfare available from these services can be found by comparing the size of the shaded triangles that lie below demand and above cost in Figure 1. The actual relative size of these triangles is an empirical matter. Figure 1 is consistent with the view expressed in the Joint Study that credit cards are both more costly and more popular than debit cards, and in this case it seems likely that the credit card welfare triangle is larger than that of debit cards.[14]

Now consider the effect of the proposal to reduce interchange fees, with the express intention of reducing the benefits offered to cardholders. This will have two effects in the credit card market. First, to the extent that it involves a redefinition of the product to one that is less desirable to users in

---

[14]   *This is not known for certain, of course. A careful empirical study would be required to estimate welfare from these services. The Joint Study clearly is not based on such an analysis, as it provides no evidence whatsoever with respect to benefits. In the absence of such work, however, the most reasonable inference from the observation that credit cards are more popular despite having higher costs, is that credit cards provide more welfare than debit cards. It is also implicit in this analysis that the current interchange fee is set in a way that maximises the combined value of the network across its participants.*

[15]   *It is worth noting that the term "market" is not being used here in its anti-trust sense: that is, as the smallest grouping of sales over which a hypothetical monopolist could profitably exercise a small but significant, non-transient increase in price. Rather, it merely connotes an area of commercial supply.*

aggregate, this change will shift the demand curve in the credit card market to the left.[15] Second, the cost of supply as the Joint Study defines it will be reduced. The combined effect will be to make the credit card market look much more like the debit card market. But since the benefit derived from the credit card market exceeds that from the debit card market, this is more likely to *reduce* welfare than to increase it.

This outcome differs from a more straightforward situation in which the costs and benefits perceived by consumers are the same as the costs and benefits to society. In the presence of network externalities, cardholders are not able, and have no incentive, to balance marginal social costs and benefits, because part of the benefit of their consumption accrues to the merchants. Hence, as discussed in section 2, it is inappropriate for prices *to cardholders* to fully reflect the costs of serving cardholders alone. The interchange fee provides a means of partially correcting this market failure.

To summarise, it does not follow from the fact that the cost of producing product A is less than the cost of producing product B that society is better off if everyone buys A instead of B. What matters is the consumer and producer surplus generated as a result of those choices. The cost of supply is only half of the story.

However, the Joint Study's perspective is different. Rather than concentrate on the surplus or **net benefit** society derives from alternative payment systems, the Study seems to be focussing on **transfers** – that is on the benefits derived by selected participants rather than the market as a whole. (It is not always easy to know whether this is indeed the case, as the Study fails to distinguish between transfers and welfare changes in its discussion). It may well be the case that some participants in a credit card system would benefit from lower interchange fees: but this in no way suffices to show that a policy of reducing credit card use would benefit society as a whole.

4.3 Relevance of 'over-usage' for interchange fees

In the previous section it was shown why aggregate welfare, not merchant costs, is the relevant criteria to judge whether interchange fees are too high. This criterion does not rule out the possibility that some cardholders use their credit cards even when it would be socially efficient for them not to. The present section considers this possibility and shows that even if this were the case, it does not imply that a reduction in interchange fees will necessarily increase welfare.

Suppose that in certain situations credit cards provide cardholders and merchants (users) with increased benefits over debit cards that fall short of their higher costs. In this sense credit cards could be said to be "over-used".[16] However, suppose there are also other situations where the additional net benefits obtained by users through credit card transactions are in excess of the alternative of debit card transactions. In this case it is not clear that a reduction in interchange fees will increase welfare. Any change in interchange fees will affect the marginal cardholder's decision about holding a card and about when to use it. While a lower interchange fee may curtail some excessive usage, it will also tend to reduce card membership below the socially optimum level and increase merchant card acceptance above the socially optimal level. To the extent that consumers who do not hold credit cards will miss transactions that would have been welfare enhancing, social welfare will be lower.

Unless there exists an all-knowing and all-powerful central planner that can dictate who holds a credit card and when, as well as the circumstances in which they are to be used and accepted, it will gener-

---

[16] *A full analysis would require that producer surplus is also counted in any welfare analysis; that is, the profits of member banks of the card association. It is assumed that the producer surplus arising from debit and credit transactions is the same at the margin, so that this element can be ignored in the following comparison.*

18

ally be impossible to obtain the first-best world where credit cards are used only when it is efficient to do so, and yet consumers hold cards up to the point where the marginal network benefit of doing so equals the marginal network cost. Instead, the efficient outcome in practice is likely to involve a trade-off between network membership and card usage, with existing cardholders using cards in some cases excessively, but with the marginal cardholder still providing a positive contribution to the net benefits of the system.

The argument that the socially optimal interchange fee may be set such that it induces some excessive usage while trying to promote greater card membership has been shown formally in Wright (2000).[17] The model he presents compares credit cards with cash, and assumes cash is costless to provide. Based on the calibrated model of the Australian economy presented in that paper, the welfare maximizing interchange fee is found to be actually slightly higher than the existing one even on the restrictive assumptions made. Relative to a central planner that can pick which consumers join the card systems and when they use their cards, the optimal interchange fee is such that it trades-off some excessive use of cards with a lower than optimal number of cardholders.

This analysis confirms what is true more generally - that even if one can show there are situations where a particular payment mechanism does not coincide with the first-best outcome in one dimension, this does not imply that any forced adjustment in pricing can improve social welfare. Given that consumers and merchants are in practice heterogeneous, to achieve the first-best pricing outcome usually requires that firms perfectly price discriminate (among other things). Just because firms in practice cannot perfectly price discriminate, does not in itself imply there are grounds for regulatory intervention. If it did, regulation of all prices in the economy would be required.

4.4 Harmful effects of proposals

The Joint Study's diagnosis of excessive interchange fees is based on a characterisation of credit card networks as mature, and on an analysis of resource costs in place of a welfare analysis. It has been demonstrated that the maturity claim is empirically incorrect, insofar as evidence permits a judgement, and that the use of resource costs in place of welfare analysis is theoretically unsound.

The Study's diagnosis is therefore wrong – so there must be grave doubts as to its recommended cure.

This subsection considers the range of harmful welfare effects likely to result from the solution the Joint Study advances: administratively reduced interchange fees.

However, before considering the precise impacts reducing interchange fees could have, it is important to note one outcome lower interchange fees will not likely achieve: namely, reduced bank margins.

The Joint Study expresses the view that interchange fees contribute to the overall margins earned by issuing banks. It implies that these earnings are excessive and that reducing the interchange fee would help correct matters.

Seen from an economic point of view, the Study's attitude here is curious. Reducing the earnings banks make is not an end in itself. Merely redistributing income – from one set of income earners to

---

[17] Julian Wright, (2000) "An economic analysis of a card payment network", NECG and University of Auckland, reproduced here as Attachment 2.

another – does nothing to expand the income available to the community as a whole. The Joint Study's failure to distinguish income transfers from changes in welfare means that it sets itself a policy goal that bears no clear link to the efficiency objectives economic policy ought to pursue.

However, even putting this aside, the Study's claim that cutting interchange fees would indeed reduce bank earnings is contradicted by evidence and argument the Study itself sets out.

The interchange fee is a source of revenue for issuers. The larger this revenue stream, the greater the potential earnings of a bank which succeeded in monopolising the issuing function. In fact, the issuing function is not a monopoly, nor is it even particularly concentrated in structure. Rather, there is open competition for cardholders, and the effect of this competition is that most of the revenue earned through the interchange fee is likely to flow through to serving cardholders. In fact, the Joint Study's view that interchange fees determine the level of cardholding and use relies on precisely such a pass-through.[18] As a result, changing the fee does little to alter the profitability of banks: rather, it affects the size and balance of the network.

This reflects the fact that interchange fees are the primary method available to an open card association to ensure a balanced emphasis on attracting both cardholders and merchants. It is in terms of this role of the fees that the Joint Study's proposals must be assessed. In the sections that follow each of the Joint Study's proposals for setting interchange fees is considered. Following this, the little-mentioned but fundamental issue of competition between 3-party credit, charge and store card systems and 4-party credit card systems is discussed.

4.4.1 Joint Study alternative 1

The first approach to regulating interchange that the Study proposes is to attribute costs among cardholders and merchants in relation to the value these parties obtain from transactions, and to set the interchange fee so that issuers recover those costs that arise from services provided to acquiring banks. This approach can be challenged in at least three respects.

First, a credit card network is inherently a joint service - it would have no value to merchants without cardholders, and vice-versa. It therefore makes no sense simply to divide costs in the way the Joint Study does.[19] To say that merchants receive only half of the benefit of the interest free period, and so should only pay half the cost, is to ignore the fact that without cardholders, merchants would receive no benefit. From an economic point of view, the approach the Study adopts is entirely arbitrary.

Second, setting interchange fees based purely on the cost of issuing ignores the fact that issuer costs are only one part of the complex endeavour of balancing network development across both cardholder and merchant users of the system. The purpose of the interchange fee is not to match **private** marginal benefits with **private** marginal costs, but rather to match **network** marginal benefits with **net-**

---

[18] *A more highly concentrated banking sector, or one that was for whatever reason less competitive, may have a lower degree of pass-through. To that extent, the fee could affect margins. But if this is the vice, the remedy should be to reduce the barriers to competition in banking – not to undermine the role of the interchange fee as an equilibrating factor in credit card networks.*

[19] *At page 50, for example, the Joint Study refers to the Cruickshank Report into UK banking. The Cruickshank approach was to disallow entirely the recovery of an issuer's interest-free period costs through interchange fees. The Joint Study acknowledged the unreasonableness of that approach and suggested instead a 50-50 sharing of the cost of the interest-free period between issuers and acquirers as both the cardholder and merchant derived some benefit from this credit card feature. Unfortunately, even a 50-50 split is arbitrary, and would result in a welfare-maximising interchange fee only by chance.*

work marginal costs. If interchange fees are set based purely on private cost there will be too little usage of cards and too few cardholders, as cardholders ignore the benefits they generate for merchants. These benefits arise when people with credit cards make more purchases, larger purchases and in some cases, new types of purchases. In order for merchants to capture these benefits, consumers have to be encouraged to obtain and carry credit cards in the first place. An interchange fee that is used to fund interest free benefits and loyalty programs can help to do this in an efficient way.

Finally, cost-determined interchange does not provide the incentives for efficient usage or provision of cards that the Joint Study seeks. To see this consider a situation in which issuance costs decrease, creating an opportunity for society to benefit from an increase in issuance activity (more cardholders and more card usage) relative to other activities. However, if the interchange fee is regulated at the cost of issuing, it will fall and there will be no change in the gross income that banks earn on each transaction, and so no particular reason for cardholder fees or rebates to change. As a result, there will be no new cardholders and no increase in card usage. To the extent the decrease in interchange fees reduces average merchant service fees, it will if anything shift the balance of network growth towards an increase in the number of merchants rather than an increase in the number of, and usage by, cardholders. This may indirectly increase cardholder numbers, but not in any way likely to be optimal when compared with the situation in which a system-wide decision is made about the level of interchange fees.

Overall, the Joint Study's first approach has little to commend it. It seems plainly inconsistent with economic efficiency, and hence is likely to impose net social costs.

4.4.2 Joint Study alternative 2

The Joint Study's second approach for regulating the interchange fee involves setting the fee at the level at which issuers are just willing to participate in the network. The Joint Study calculates that issuing banks have a shortfall of $0.19 per transaction, which implies an interchange fee of $0.19 for issuing banks to break-even.

However, as noted above, the purpose of redistributing funds from one side of the market to the other is not to make sure both issuing and acquiring banks break-even.[20] Rather, it is to ensure an economically efficient balance between promotion to cardholders on the one hand and promotion to merchants on the other. The break-even approach disregards the need for promotion on the issuing side, by overlooking the costs of promotion, of marketing and of providing loyalty programs to cardholders. Without the right level of such promotional programs, open card networks would not be able to compete with closed networks, and more generally would not maximise the value of their networks to users.

Previous academic and regulatory studies of interchange fees (Baxter, W.F., 1983, Schmalensee, R., 1999) have concluded that the appropriate level of interchange fees depends on systematic differences in costs and demand responsiveness to price between issuers and acquirers, and the nature of competition on both the issuer and acquirer side of the market. The Joint Study's second alternative simply ignores these studies, and proposes that fees be set on a basis that seems quite unrelated to achieving efficient outcomes.

---

[20] *Regardless of the level of the interchange fee, issuers and acquirers will adjust the scale of their activity to break-even. The purpose of the interchange fee is not to provide for cost-coverage (as this will occur regardless) but rather to set incentives for issuer and acquirer activity to be taken to the level efficient for the system as a whole.*

This proposal is not well defined, and explication of the details may well reveal further problems. For example, the breadth of issuers' activity considered under this proposal will clearly be important. If the supply of non-revolving credit services to cardholders were included in what is effectively an "issuing profitability" test, the resulting interchange fee could provide issuers with insurance against credit risk such that the losses associated with bad loans were socialised among the open loop system users. This would create obvious moral hazard problems, and the resulting equilibrium, though poorly defined, is unlikely to be anywhere close to economic efficiency.

Fundamentally, both of the Joint Study's proposals suffer from the same flaws. Because they fail to recognise the two-sided nature of the service being provided, the resulting interchange fees are unlikely to perform their crucial balancing role in an efficient manner.

4.4.3 Comparison with 3-party credit card systems

Open credit card systems, such as VISA, MasterCard, and Bankcard, involve four parties: the cardholder, the card-issuer, the merchant, and the merchant's transaction-acquiring bank. Such 4-party systems necessarily involve transactions between the two sets of intermediaries: the issuers and the acquirers.

Open credit card systems compete with two other types of "closed" credit card systems, which are generally operated by non-banks: store credit and 3-party credit and charge card systems. In both of these competing types of credit card systems the "interchange" between the card issuing function and the merchant transaction-acquiring function takes place within a single organisation. In some store credit arrangements that organisation is the merchant itself. In the case of 3-party payment systems, such as American Express or Diners Club and the majority of store cards, the same organisation is the issuer and the acquirer, so any "interchange" payment would simply be an internal transfer payment. The Joint Study's authors appear to have chosen to exclude these credit card competitors from the scope of their study.

This exclusion is unfortunate since almost all of the issues raised by the Joint Study relating to 4-party card systems could have been usefully addressed by examining the same issues under the simpler arrangements of a 3-party card system. The particular pricing that 3-party systems adopt, such as positive merchant fees, no cardholder transaction fees, annual fixed fees, and cardholder rebates, are properties that for these systems clearly have nothing to do with interchange fees as such (since these systems do not have explicit interchange mechanisms). However, in open networks an explicit mechanism – namely, interchange fees – is required to achieve similar pricing outcomes.

Now, it is plainly the case that end-users have a direct influence on the cardholder and merchant fees that a closed card system will want to set. By exactly the same logic, it follows that the claim that end-users in open systems have no influence on pricing because they are one step removed from the setting of the interchange fee cannot be true. Rather, end-users have a similar influence on the cardholder and merchant fees that are optimal for an open credit card system as they have with respect to a closed system, and thus on the interchange fee an open credit card system will want to set.

The exclusion of closed card systems from the Joint Study's consideration not only confuses the analysis – it also creates a seeming bias in the Study's coverage and in the views it expresses. For example, while the report criticises the eligibility rules of the open credit card systems, it ignores the fact that 3-party card systems such as American Express are generally closed to entry by any other institutions as either issuers or acquirers.

Similarly, the level and method of setting interchange fees in the open credit card systems is heavily

criticised in the report. However, there is no criticism of the level or method of establishing the transfer payments between issuing and acquiring functions in the 3-party payment systems, even though these methods are far less transparent, and lead to higher merchant service fees than do the interchange arrangements in open credit card networks.

Even more seriously, the exclusion has important implications in terms of competitive neutrality. The Joint Study acknowledges the risks involved on page 5:

> *"Although they have many similarities with credit cards, store cards and charge cards such as American Express and Diners Club are usually not issued by financial institutions and do not have interchange fee arrangements. The operation of these card schemes are outside the scope of this study. Nonetheless, the study is mindful that its findings on credit card schemes may have implications for the competitive position of credit cards vis-à-vis store cards and charge cards."*

Despite this recognition, some of the Joint Study's recommendations would have such a differential impact on open credit card systems versus their closed counterparts.

For instance, if authorities were to regulate the interchange fee of a 4-party card system, they would limit the ability of a 4-party card system to offer the benefits to cardholders which would maximise its network value. However, a 3-party card system, which by definition has no explicit interchange fee, would still be free to offer benefits to cardholders in any way it sees fit. To maintain competitive neutrality, interference in the setting of the interchange fee for a 4-party system would require that authorities also directly regulate the pricing and loyalty programs offered by a 3-party card system. Competitive neutrality would, in other words, require that the final prices of the closed systems – the prices they charge consumers and merchants – be set by the regulatory authorities.[21] If this is what the Joint Study is advocating, it ought to say it explicitly.

In short, the exclusion of closed credit card networks from the scope of the Joint Study has the potential to harm the open credit card networks in a manner which is not competitively neutral. Here too, the Joint Study's approach is more likely to harm than to promote economic efficiency.

4.5 Resource cost inferences

As a final matter, it is necessary to turn to one of the few empirically-based assessments in the Joint Study: namely the claim that interchange fees are above resource costs. In this section, it is shown that the Joint Study, in carrying out its assessment, has overlooked some important resource costs, thereby overstating the degree to which interchange fees actually do exceed resource costs properly defined.

4.5.1 The Study's resource cost benchmark

The Joint Study makes frequent reference to the resource cost of supplying card services, so it is important to begin by considering just what this term means. Most economists would define resource cost as the minimum cost of producing a given product or service. For example, the resource cost of producing a particular model of Holden car would include the cost efficiently incurred for making each of its components (engine, wheels, air conditioner etc), for assembling these components, and

---

[21] *Of course, the closed systems could be required to set the relativities between their charges in such a way as to define an implicit interchange fee that was (say) resource cost based. It seems unlikely that such a requirement would be capable of precise definition, much less efficient enforcement.*

T.1

for sales and distribution of the assembled vehicle. Clearly, the resource cost of production cannot be established before the product is defined – the top of the range Holden presumably has a higher cost than its entry-level counterpart.

In the case of credit card systems, the Joint Study appears to define the product quite narrowly. For example, the cost of loyalty program points awarded to cardholders is apparently not regarded as part of the product,[22] since it is explicitly excluded from the resource cost of card issuers. While loyalty points may not be crucial to the existence of a credit card service, neither is a car air conditioner crucial to the mobility service provided by cars. However, it would be foolish to claim that the cost of air conditioners, which are an important element in the value consumers place on cars, ought to be excluded from the calculation of the resource cost of Holdens. In exactly the same way, market behaviour shows that loyalty points are a valued part of the credit card product from the perspective of cardholders.[23]

By excluding the cost of loyalty programs, the Joint Study effectively redefines the credit card service provided to cardholders and calculates a "resource cost" of this new service, which does not actually exist in the marketplace. From an economic point of view, this approach is entirely arbitrary.

4.5.2 Gap between interchange fees and resource costs exaggerated

The Joint Study's under-inclusive approach to assessing the cost of providing credit card services leads directly to the conclusion that the service providers are significantly over-charging. However, on the Joint Study's own figures, once the cost of loyalty programs is included on the issuing side ($0.46) and the Joint Study's recommended contribution to capital costs is added ($0.30), the identified mark-up per transaction is reduced dramatically.

Accordingly, the claim that interchange fees overcompensate financial institutions for the costs they incur vastly overstates the case. Once the costs of the service actually provided are used in place of the costs of the hypothetical service which the Joint Study defines, the margin component of the fee appears modest. The Joint Study places great emphasis on its claim that interchange fees are <u>substantially</u> above resource costs; but even this claim is not well supported by the data provided by the Study.

This is not to suggest that interchange fees ought to be based on resource costs – as a general matter, there is no reason to believe that resource costs provide the right point of reference. Rather, the fees should be set at a level that results in an efficient balance as between the card-issuing and merchant-acquiring activities: resource costs are only one of the factors that jointly determine this level.

Despite this, the Joint Study's error highlights the difficulty any regulator would face: first in defining and implementing resource costs; and then, and even more so, in accurately assessing the other factors which economic analysis shows must be taken into account. In these circumstances, the claim that regulation would increase welfare naturally induces some scepticism.

---

[22] *Whether loyalty programmes are treated as product features or as promotional strategies, the cost of providing them is an important and legitimate cost of providing a competitive credit card payment system. Store card and 3-party card systems also provide loyalty and promotional benefits to their users. It would be unreasonable to expect these services to be priced in such a way as they never recovered the costs of these schemes. It would also be unreasonable to impose a policy regime on four party card networks that limited their ability to compete with these alternatives. Moreover, the fact that loyalty schemes are observed in other highly competitive markets demonstrates that firms in these markets could not win market share by abandoning loyalty programs and thereby reducing cost.*

[23] *Loyalty schemes are a competitive mechanism for transferring surplus to consumers, just as lower prices are. It might be claimed that in this case loyalty programs are inefficient as compared with simply offering lower prices. However, this is not consistent with a view that what is occurring is a transfer of surplus. The same competitive forces pressing firms to pass rents to consumers would ensure the mechanism for doing this was efficient. Alternatively, it might argued that the loyalty schemes are evidence of profit enhancing price discrimination* and *that this is welfare-reducing (an undemonstrated empirical possibility). However, this is also implausible given the same behaviour is observed in the smaller closed systems that very likely lack market power.*

4.6   Summary of interchange issues

The Joint Study considers that regulating the interchange fees in open (4-party) credit card networks will

i.   make debit cards more attractive than credit cards; thereby

ii.   saving resources; and

iii.   reducing the margins of issuing banks.

These assertions are all incorrect. Moreover, it has been shown above that the Joint Study's proposals are very likely to:

i.   have no effect on issuing bank margins;

ii.   increase the use of closed (3-party) credit-card networks, with minimal effect on the use of debit cards; and most importantly,

iii.   reduce aggregate welfare.

Aside from these fundamental concerns with the details of the proposals, the process which gave rise to them ignores most of the reasonable guidelines for the analysis of market interventions. Since open credit card networks are quite obviously a thriving and growing industry, intervention is only justified if:

i.   it can be established that credit card networks have grown past their socially optimal size;

ii.   an appropriate intervention can be identified and the welfare benefits of this intervention can be roughly quantified; and

iii.   the costs of intervention, especially the potential harm caused by regulatory error, can be shown to be less than the expected benefits.

The Joint Study has applied none of these tests. Rather, starting from a false premise, the Study arrives at flawed conclusions.

The most fundamental weakness in the Joint Study's analysis of interchange is to identify the benefits of intervention as being the savings in "resource costs" as a result of greater use of debit cards. This ignores the value that participants derive from the network. By analogy, imagine if, by some administrative device, people were induced to substitute Lada cars for Holdens. Resource costs would fall. But is it the case that *therefore* welfare would increase?

The errors made in the Joint Study's own attempt to estimate resource costs show the pitfalls involved in attempting to determine interchange fees by administrative *dictat*. The setting of interchange fees is a complex matter that requires commercial judgement. In the current arrangements, this judgement is shaped by the realities of market-place competition: between VISA and its open credit card network competitors; between the open credit card networks and their closed counterparts; between credit cards and debit cards; and between cards and other means of payment. This judgement is then tested in the negotiating process over interchange between members, which elicits information about the likely outcomes associated with alternative possible fee levels.

It would be foolish to claim that this will lead to perfect outcomes. However, it would be even more foolish to assume that a regulator, insulated from competition and isolated from the realities of commercial judgement, could bring as wide a range of considerations to bear in determining the appropriate fee level. Rather, it is virtually inevitable that a regulator would be drawn to a mechanical formula, such as those proposed in the Joint Study. These, it has been shown, have little merit and are merely likely to distort efficiency. The need for caution before regulation is imposed is therefore all the greater.

5. Does the "No-surcharge" rule distort competition?

The Joint Study concludes that the effect of the no-surcharge rule is to distort competition between credit cards and debit cards in favour of the former. The underlying rationale is that cardholders do not perceive the true cost of their actions when merchants are prevented from passing on this cost at the point of sale. The Joint Study implies that this allows "cross subsidisation" to occur, in which credit card users are "subsidised" by consumers who use other means of payment. If cardholders were forced, through surcharging, to pay a greater contribution to the cost of credit card networks, this cross-subsidy, and the resulting present bias in favour of credit cards, would be reduced or eliminated.

This rationale is however subject to three potential criticisms. First, it is founded on a partial view of the credit card service, which disregards the value that merchants receive from the service.

Second, it assumes cost differentials between debit and credit will be passed on systematically and without further mark-up by merchants. Thus, it ignores the possibility of surcharges being used by merchants to expropriate consumer surplus. It also ignores the fact that setting multiple prices is costly for merchants and consumers.

Finally, the implications of surcharging have not been considered with respect to other payment instruments. In particular, while the Joint Study advances a general argument to the effect that merchants should be able to reflect the cost of payment in their prices, it restricts its attention to the effect of this in respect of open network credit cards. If the argument is valid then it must also apply to other means of payment, including cash.

Before addressing these issues in detail, an example may be helpful in illustrating some of the issues involved, particularly those related to cross-subsidies and surcharging.

5.1 The car-park analogy

Free car parking is commonly provided at shopping centres. Although parking spaces are provided free to shoppers,[24] these spaces are certainly not free to provide, involving the land and construction costs, the opportunity cost of the money tied up, as well as maintenance and security. These costs are recovered through the shop rentals that merchants pay to the shopping centre owner, and are thus ultimately passed back to consumers in the retail prices they pay. However, shoppers arriving at the centre by bus, or on foot, pay the same price as those who use the free car parking facilities. Because this situation is logically similar to the practice of not surcharging on credit cards, it is useful to consider whether the corresponding concerns are also present in this case. In particular, the validity of the following propositions can usefully be assessed:

i.   Shopping centres which do not offer free parking facilities are more efficient.

ii.   Shoppers who arrive by bus are cross subsidising those who drive.

iii. Since shoppers do not face the true cost of parking relative to alternatives, they will over-use the parking facility.

---

[24] *In some cases shoppers pay to park, but their price is discounted relative to the price paid by non-shoppers using the same car park. The same analysis applies in this case. The point is that shoppers receive car-parking at a price that is not fully cost-reflective.*

iv.   Individual merchants should be allowed to recover the costs of providing car parking by adding a surcharge for shoppers who have used the free car parking.

v.   Because free parking has become widespread, shopping centres are being forced to offer free parking.

vi.   The overall price of retail merchandise is higher because of the existence of shopping centres that offer free car parking.

These propositions do not survive analysis. First, even if a cheaper shopping centre exists which does not offer a free car park, this does not imply that such a centre is more efficient or more socially desirable. What determines the efficiency or otherwise of a shopping centre is whether the services it offers maximise the net benefits to society from the use of the resources the centre consumes. More specifically, whether providing a free car park is socially desirable or not depends on whether the provision of free car parking increases aggregate social benefits relative to social costs.

The services provided and prices charged are, of course, relevant to assessing this issue, but it cannot be determined by observing prices alone. In fact, the existence of free car parking may be an important factor in the success of shopping centres, since it enables centres to attract greater numbers of customers and thus, through economies of scale, to provide lower priced merchandise.

Second, there need be no cross-subsidy. In fact, shoppers arriving by bus would likely be worse off if there were no car park facility.[25] This will be the case if the free car parking facility is needed to attract a sufficient number of customers to the shopping centre. In that case, when the free parking is removed, individual shops, and ultimately the shopping centre, will have to either close down or increase prices to cover fixed costs. As a result, shoppers who arrive by bus will be worse off.

Now consider the third proposition: that because shoppers do not face the true cost of parking, they will over-use the facilities. This proposition is based on the false premise that car-shoppers are the only ones that benefit from the car park. In this sense, it fails to distinguish private costs and benefits from social costs and benefits. Even if shoppers' private benefit of being able to use the free parking is less than the full cost of providing it, this does not necessarily imply over-use, as the merchants also benefit from having attracted additional shoppers to the centre.

While there is no reason to believe car parking facilities will be over-used at the aggregate level, the possibility that some shoppers will over-use the free parking facilities cannot be ruled out. Where car-parking is offered free it may encourage some customers to use the facilities even though the social benefits of their usage falls short of the opportunity cost of the parking space they take. Equally, there may be other shoppers for whom the private benefits of shopping do not incent them to shop despite the benefits that would arise to merchants being significant. For such customers, even offering free parking may be insufficient to encourage them to shop. While there are some mechanisms shopping centres can use to discriminate between such customers (for instance, only offering free parking if a purchase is made at the shopping centre), centres are certainly not capable of engaging in perfect discrimination. Given this, the first-best solution will not be attainable. The second-best outcome will likely involve some trade-off; with over-usage of the parking facility by certain customers, while others do not use it at all, despite it being socially efficient for them to do so.

---

[25]   *It is shown below that for there to be a cross-subsidy then bus-shoppers must be better off without car-shoppers.*

T.1

From an economic point of view, this over-use is not properly considered as an inefficiency, as there is no practical means of correcting it without reducing welfare. Comparisons with first-best outcomes are no more relevant here than comparisons with Nirvana: if there is no way those outcomes can be attained, or (equivalently) the costs of attempting to approach them would exceed the benefits, they cannot provide a sensible basis for condemning the outcomes that are actually observed.[26] Rather, those outcomes, though departing from the first-best, may well reflect optimising behaviour within the constraints that reality imposes.

Fourth, consider the value of surcharging in this environment. Each individual storeowner will not voluntarily fund the cost of centre parking, and if they could, they may well want to levy an additional charge on those customers who use the parking facility, in order to offset that component of their rent. For any individual merchant, this strategy is likely to be profitable – once shoppers are at the shopping centre, merchants may have a certain (albeit transient) degree of market power over sales;[27] more importantly, each individual merchant may assume that the consequences of its individual decision as to whether to surcharge will have a small aggregate effect on the number of consumers. It should be apparent, however, that widespread surcharging would defeat the purpose of offering free car parking, which is to encourage shoppers to come to the centre. To solve this free rider problem, the centre owner will require that all the merchants contribute to the cost, and if the technology to surcharge customers for car parking existed, the centre owner would put in place a no-surcharge rule. Such a rule ensures both shoppers and merchants are better off.

The next proposition says that because free parking has become widespread, shopping centres are "forced" to offer free parking. Even if this were true, it does not follow that there is any conspiracy against the consumer, nor that the provision of free parking is socially inefficient. Shopping centres might generally prefer to not offer free car parking facilities if their customers did not expect them to do so, since these facilities are costly to provide. However, it is competitive pressures that force them to do so. If they do not offer free car parking they will lose customers to shopping centres that do. Merchants face many costs of doing business, but these costs are usually incurred because they make good business sense. The alleged "forcing" is therefore no more than the manifestation of competition at work – just as the fact that shopping centre owners must provide pleasant surroundings for consumers adds, rather than detracts, from social welfare.

The last proposition, that the overall price of retail merchandise is higher because of the existence of shopping centres that offer free car parking, equally does not necessarily follow for either of two reasons. Consider the case without free car parking. This may raise the total cost to the community of reaching the shops, and hence lower the amount of shopping that occurs. For example, it might lead to more (and likely purely wasteful) search for free parking spaces. Further, if this means a lower number of shoppers, the economies of scale provided by the shopping centre will not be exploited and the net result could well be higher prices. Alternatively, absent free parking facilities, the shopping centre might set up other arrangements to encourage shoppers to come to the centre, such as offering its own shuttle bus service. Such a system may well be a more expensive system to operate on a per-customer basis and so would lead to higher merchant prices. It is notable that the growth and success of shopping centres and supermarkets over individual retail outlets is usually associated with greater convenience for customers without any increase in retail prices.

---

[26] *The error the Joint Study seems to make, of using a comparison to an impracticable first-best option to condemn an existing practice, is commonly referred to in policy analysis as "Nirvana economics".*

[27] *The power at issue is transient as shoppers will learn that they risk being surcharged and will adjust their behaviour accordingly.*

This analogy demonstrates the fact that it can be socially efficient to provide some customers with additional services at no additional cost and that this need not involve any cross-subsidy or increase in overall prices.[28] A more detailed analysis of the no-surcharge rule as it applies in 4-party credit card networks is now presented.

5.2 The relevance of value

In suggesting that the no-surcharge rule distorts competition between credit and debit cards, the Joint Study implicitly assumes that merchants view these as equivalent means of payment.

Perhaps the primary difference between credit and debit cards is the automatic line of credit available with the former. This feature alone means that the average dollar value of purchases made on credit cards is likely to exceed that for debit cards. As a result, merchants who accept credit cards are able to make more sales and more high value sales than those who do not. Thus, if merchants were forced to choose between accepting credit cards and debit cards in a world where the same number of each type of card was on issue and the cost of acceptance was identical, they would always choose credit cards. Put simply, merchants place a higher value on credit cards than debit cards.

It follows that the proposition that competition between payment cards is distorted cannot be established without taking this value difference into account. The appropriate test is whether the ***net benefits*** merchants receive from accepting credit cards rather than debit cards are distorted in some way. This cannot be established through evidence that is confined to analysis of the ***gross cost*** to merchants of the two methods. In order to establish that competition is distorted by the no-surcharge rule it would be necessary to explicitly consider the value that merchants receive from credit card sales.

5.3 Would welfare rise if the "no-surcharge" rule was abolished?

The Joint Study appears to attribute inefficiencies in the payment system to the no-surcharge rule. If this were the case, it would be reasonable to infer that welfare would rise if this rule were abolished. However, there are several reasons to query whether removing the no-surcharge rule would increase welfare. These reasons are briefly outlined here and discussed in more detail in the following subsections.

First, the no-surcharge rule cannot be criticised for enforcing a cross-subsidy in favour of non-revolving credit card users if, as is demonstrated next, there is no such cross-subsidy.

Second, no merchant is forced to join any given card network. Many merchants choose not to accept American Express, presumably for the reason that the additional sales would not justify the significantly higher merchant service fees this card sets. Diners Club is also less popular with merchants, some of whom claim that this card is slower to pay than others. Thus, merchants will neither join a card network nor remain as members if they do not benefit from doing so under the rules as they currently exist.

---

[28] *Other examples of services that are commonly offered free by merchants to those customers who take advantage of them include free-call phone numbers, free gift wrap, free packaging and free delivery.*

T.1

Third, the inference that welfare would rise without the no-surcharge rule does not address the presence of externalities. While any given merchant would clearly prefer lower service fees, and accordingly might surcharge if it were permitted to do so, all merchants collectively are better off as a result of the cardholder promotion that is partially funded from service fees. The abolition of the no-surcharge rule would permit free-riding, which will reduce the value of the network to all participants.

Fourthly, formal modelling of the no-surcharge rule by Dr Julian Wright of NECG and the University of Auckland demonstrates the contrary proposition that the no-surcharge rule can be welfare enhancing. This modelling uses many of the properties assumed by the Joint Study, and looks at the full impact of a move to surcharging, incorporating the effects on those paying by credit cards and those using other instruments, as well as the effect on merchants and banks. In order to build a case that removing the no-surcharge rule will increase welfare, it would be necessary to show why the parameter settings that produce these welfare enhancing results are less likely than some other settings that support the contrary view.

Finally, international evidence suggests that the primary effect of removing no-surcharge is to permit the hold-up of captive consumers. This outcome is not desirable and is not likely to be efficient.[29]

5.3.1 Is there a cross-subsidy?

The Joint Study suggests that the net effect of the payments system is that (non-revolving) users of credit cards are subsidised by consumers using other payment methods.

Our earlier discussion of the provision of free parking at shopping centres broadly illustrated the concept of a cross-subsidy. To address the Joint Study's claim, however, it is useful to more explicitly examine what economists mean by a cross-subsidy. A cross-subsidy occurs when one user, or group of users pays less than the *incremental cost* of their usage while another user, or group of users, pays more than their *stand alone cost*. Incremental costs are the costs directly caused by the addition of an increment of output to the existing output level; the stand-alone costs are the costs that would be required to provide that increment if no other output was being provided. In this case, the relevant definition of cost is the total cost of providing retail service by any given merchant, including (in the case of the stand alone cost of serving a group of customers) the joint and common (overhead) costs.

The quantitative difference between incremental and stand-alone costs varies across industries and production processes. In some cases, such as the provision of a customer access network in telecommunications, there is a very considerable gap between incremental and stand-alone costs which arises from the fact that most costs are common. The greater is the proportion of common costs to total cost, the larger is the gap between incremental and stand-alone costs, and the greater is the scope for setting different prices to different groups of consumers without causing a cross-subsidy.

---

[29]   "Pure" hold-up may not be inefficient if it merely transfers surplus from consumers to merchants. However, it is unlikely to be the case that such fine discrimination could be practiced. More likely, the experience of hold-up would lead consumers to take protective action, with the cost this action involves being a straight income loss to society.

The definition of a cross-subsidy set out above implies a very natural test for whether cross-subsidisation is occurring: this is whether one group of users (those said to be paying more than their stand-alone costs) would experience a welfare increase if another group (those said to be paying less than their incremental cost) were eliminated in some way. In the present context, a cross-subsidy would only exist if retail shoppers other than convenience-only credit card users [30] are paying more than they would if this credit card service were not available.

There are several reasons why this is unlikely to be the case. If the credit card service did not exist, merchants who now accept credit cards would most likely make fewer and smaller value sales.[31] In order to cover their fixed operating costs, merchants may be forced to raise their prices. In this case the remaining customers would be worse off.

Alternatively, the merchant may offer its own charge card service in order to capture such sales. If the cost to merchants of so doing was lower than their current merchant service fee for general-purpose credit cards, then presumably store credit would be more widespread. The fact that store credit is not widespread suggests that for most merchants it is more costly than accepting general purpose credit cards.[32] This suggests that if store credit is used to fill the need for credit-assisted purchasing, customers who are not convenience users of general purpose credit cards are likely to be worse off in the absence of general purpose credit card usage. These conclusions are strengthened by two further observations.

First, the potential savings due to elimination of the alleged cross-subsidy that could be passed on to other consumers are at most equal to the *difference* between credit card merchant service fees and the (weighted) average cost of supplying the other payment methods that would be chosen if credit cards were not available, multiplied by the fraction of sales that are currently made using credit cards. This is much less than the merchant service fee itself, and there is no reason to think that it is not zero or negative. Certainly, the Joint Study, despite the strong claims it makes, provides no evidence that would suggest that the likely savings are material.[33]

Second, experience in cases that are acknowledged to involve a cross-subsidy (such as the provision of universal telephony service) suggests that such subsidies are rapidly unwound by competitive entry where this is feasible. There are no laws preventing a retailer from specialising in providing service to users of "low-cost" payment systems, just as there are no laws which require shopping centres to provide free car-parks. If a genuine cross-subsidy from non-credit card customers to credit card customers did exist, then some merchants could develop a profitable niche by passing to their non-credit card customers some of the costs they saved by refusing to accept credit cards.

Given these conditions, it can fairly be concluded that consumer prices would not fall in the absence of credit cards, and that there is no cross-subsidy.

---

[30] Convenience credit card users are those who do not make use of revolving credit facilities. Also called non-revolvers, they use a credit card like a charge card.

[31] Consumers are more inclined to make their preferred purchasing decisions when availability of funds is not an issue, when they can make purchases over the phone, fax, or on the Internet, and when they can make transactions overseas without incurring large transaction costs.

[32] This view is reinforced by the observation that, where store credit is available it is generally more expensive than that offered by general purpose credit cards. For example, the Myer/Grace Bros store credit interest rate is above 20%p.a.

[33] Even if the difference is positive, it is not necessarily the case that small reductions in cost will get passed onto customers. In practice, merchants will not want to adjust prices to very small variations in costs, especially as handling multiple prices or surcharges itself involves transaction costs, both to merchants and consumers.

T.1

5.3.2 Are merchants being "forced" to join?

If there is a cross-subsidy between card users and other consumers, then merchants who accept cards must, having regard to the definition of a cross-subsidy, have higher prices than those who do not. Thus, if some consumers want low prices rather than credit card mediated purchases, they will be able to obtain these by patronising the low-priced merchants. The market will tend to separate into merchants that accept cards and those that do not. This would seem to be a likely and efficient outcome.

The Joint Study, however, anticipates a different outcome: that all merchants accept credit cards and do so, not because they want to facilitate larger purchases, but because they have no alternative. In the words of the Joint Study (page 48),

> *"But as credit card networks become more widespread, accepting a credit card becomes a condition of doing business rather than a means of gaining an advantage over competitors."*

Taken as a purely empirical matter, this claim seems at best over-stated. How, for example, should the evidence which shows that not all merchants accept credit cards be interpreted? Further, of those that do accept cards, many choose not to accept some particular cards. One card with quite low penetration among merchants is American Express. There would seem to be two possible reasons for this low penetration rate.

The first is that the merchant service fees for American Express are too high, so merchants select lower cost methods, such as VISA. Alternatively, one might argue that merchants do not feel as much need to accept American Express because there are fewer cardholders for this card. Both of these reasons are likely to have some validity.[34] Indeed, they provide an explanation for VISA's success in the market, which is that it has found a low cost method of providing a service that is highly valued.

The fact that VISA is popular does not mean that joining is a "condition of doing business", however. Merchants can and do differentiate themselves by offering lower prices as a trade-off for less service in the form of not accepting credit cards. Indeed, there are numerous examples of similar strategies in other markets. Low cost "nuts and cola" airlines, for example, do just this. Similarly, discount chains compete with full service retailers using a "small profit, quick turnover" business model. Other merchants, notably some major chains in the UK, do not accept general purpose credit cards, preferring instead to promote their own store cards. Given these facts, to suggest that merchants have no choice is at best a serious exaggeration.

Having said that, the fact that some merchants do not accept VISA needs to be contrasted against the equally obvious fact that, at least within some market categories, most do. Given this, it is natural to ask whether there is some reason, unrelated to fair and open competition, that accounts for this high penetration rate. To address this question, the nature of fair and open competition needs to be discussed.

To a considerable extent, competition can be viewed as a process that forces firms to do things that they would otherwise not want to do. Retailers may prefer not to open on weekends, for example, but if most of their competitors are open on weekends they are likely to feel compelled to follow suit. Equally, tradespeople almost certainly would prefer to give "estimates" rather than "quotes" but insisting on doing so is likely to adversely affect their business when most of their competitors give quotes. In each of these cases, competition is indeed *compelling* firms to operate in ways that they would otherwise not choose.

---

[34] *Indeed, the two explanations are logically related. To say that merchant service fees for American Express are "too high" is another way of saying that because AMEX has fewer members, merchants are not willing to pay as much to take Amex as they are for larger cards.*

Seen in this light, the fact that most merchants feel the need to accept credit cards looks very much like competition at work. Accepting cards is not a regulatory rule or imposition; rather, it is part of the process by which firms try to match and better rivals.

In this sense, the distinction the Joint Study draws between conduct that is "*a condition of doing business*" and that which is "*a means of gaining an advantage over competitors*" is at best confused. The essence of competition is precisely that it constrains the ability of firms to provide consumers with other than what they want: it converts consumers' expectations into "*a condition of doing business*", and thereby protects and advances the consumer interest. The widespread acceptance of cards such as VISA is merely an illustration of this familiar but all-important process.

5.3.3   Would abolishing the rule give rise to free-riding?

The Joint Study expresses considerable concern about the potential for free-riding by non-revolving cardholders under present credit card schemes. This is said to arise from the assumed cross-subsidy between retail consumers. As discussed above, there are reasons to doubt the existence of a cross-subsidy. However the proposal to abandon the no-surcharge rule may itself create a cross-subsidy.

In the absence of the rule, merchants could choose to surcharge. Based on international evidence, the effect of this could be that merchants serving captive customers would surcharge, while those facing a more contestable base of customers would not, since doing so would likely lose the sale. For example, it is now quite common in the UK for a 10% surcharge to be levied on telephone sales of theatre tickets. Buyers of such tickets have very limited alternative purchase options. In this case, there would be some drop-off in cardholding as consumers on the margin realised that the card was less useful. This would reduce the value of the network in aggregate but the effect on the surcharging merchant is always smaller than the total loss (due to the existence of network externalities).

Thus, in aggregate merchants (and cardholders) would be better off without surcharging, but some merchants will rationally choose to free-ride on the network by surcharging if that is permitted. This has a direct parallel with the shopping centre car-park analogy described above, in which each merchant preferred to contribute nothing to the cost of car-parking, though all merchants benefit from its collective provision.

This is a genuine case of a cross-subsidy. The non-surcharging merchants are the source of the cross-subsidy and the surcharging merchant is the recipient. The non-surchargers would be better off if the surcharger did not accept credit cards at all, because then no cardholders would have cause to doubt the integrity of the network.

5.3.4 Lessons from modelling the "no-surcharge" rule

Aside from suggesting the no-surcharge rule is responsible for generating cross-subsidies from users of "low cost" payment instruments to credit card users, the Joint Study also suggests that the no-surcharge rule may be harmful to society in general. The argument is that if merchants had the discretion to impose surcharges, they would surcharge depending on the cost of the payment instrument. Because, it is said, merchants set prices to reflect costs, including a profit margin, cardholders would bear the true cost of the services they enjoy, ensuring that they face correct price signals to hold and use cards.

However, in credit card systems where there are network externalities, "correct price signals" will often mean prices that deviate from private costs. If this is accepted, as it appears to be by the Joint Study, it follows that the no-surcharge rule combined with an appropriate interchange fee may be needed to ensure that marginal social benefits and costs are equated.

However, the Joint Study views credit card networks as mature and considers (as has been shown, inaccurately) that this materially alters the analysis. It assumes that since cardholders and merchants are "comfortable" using credit cards, they will by and large continue to do so regardless of the price they face. Therefore the Joint Study's assessment is that if cardholders were to face the full cost of using cards, although credit card usage would contract, this would only be a marginal effect.

There is a significant tension between a view that the network will not be significantly affected by eliminating the no-surcharge rule, and a claim that the removal of the rule will lead to cardholders responding to merchant surcharges by switching to "low cost" payment instruments. Either price signals will work, in which case credit card network membership and usage will decline in favour of debit cards as credit cardholders are charged more, or price signals will not work, in which case removal of the no-surcharge rule would have little or no effect.

The Joint Study does not seem to be unduly troubled by the inconsistency involved in simultaneously claiming both that price signals will work and that they will not. Having said that, the Study's pre-dominant view appears to be that cardholders will respond to relative prices, so that having merchants pass on their merchant service costs to cardholders will reduce card membership and usage. What then would be the effects of this decline?

A recent paper, Wright (NECG, 2000) reproduced here as Attachment 2, analyses this question. It develops a model of a credit card payment service which is based on many of the properties assumed by the Joint Study, and looks at the effects of eliminating the no-surcharge rule. More specifically, the model assumes that in the absence of the no-surcharge rule, merchants will indeed surcharge – the assumption that the Joint Study makes, as it claims that removing the rule will alter price signals and hence induce changes in behaviour.[35] Given this assumption, the model measures the welfare consequences for the Australian economy, considering various types of merchant pricing behaviour. The analysis takes into account the effect on other customers, as well as on credit card paying customers.

The conclusion is that, having regard to the type of merchant pricing behaviour that the Joint Study assumes, the costs of eliminating the rule could be in the order of $4.5 billion to the Australian economy. This result is derived even though it is assumed that surcharges are applied in a cost-plus fashion, that cardholders use their cards to capture rebates, that cash customers subsidise cardholders, and that the interchange fee is set at around its current level. Varying these assumptions in the direction of greater realism can lead to significantly higher estimates of the welfare losses which would result if the policy the Joint Study recommends were implemented.

NECG's analysis demonstrates the importance of aligning marginal **social** costs and benefits. Allowing surcharging would permit the full costs of providing credit card services to be passed back to cardholders. Because a credit card service provides joint value to cardholder and merchant, this generally leads to too little card membership and too little card usage. Essentially, cardholders will face a price of using cards that is above the marginal social benefit arising from their membership. Because membership will ultimately decline, merchants can also be worse off, as they are more likely to face

---

[35] If removing the no-surcharge rule has no effect on merchant pricing behaviour, then it can have no effect on the choice consumers make between payment systems, so that the Joint Study's claims are simply misplaced. The assumption that given the opportunity, merchants will indeed surcharge is therefore fundamental to the Study's conclusions. The analysis reported here consequently starts from the assumption most favourable to the Joint Study's case – namely that absent the rule, surcharging will indeed occur.

[36] In practice they may have to make a smaller purchase than they would otherwise prefer, or delay purchase until they have available funds.

34

customers who will not be able to complete a purchase for lack of ready credit.[36] Each individual merchant would rather not pay to encourage a consumer to hold a card, since each merchant takes the number of cardholders as something it has no control over. Merchants therefore face a free-rider problem. Given this problem, merchants can collectively benefit from imposing the no-surcharge rule, which ensures all merchants contribute to promoting card membership.

This result, that the Australian economy would be worse off by $4.5 billion, reflects an attempt to measure the extent of these losses in a way that captures the characteristics of the Australian market put forward by the Joint Study. Because it is based on premises that are unrealistically favourable to the Joint Study, it is likely to provide a conservative measure of the true losses. Nevertheless, it demonstrates there is no prima facie case for assuming the no-surcharge rule is harmful, and in fact suggests that it is welfare improving.

For example, the analysis does not take into account the fact that merchants have fixed costs of operations. Because credit cards increase the number of merchant sales, allowing surcharging could then increase the average cost per-sale. This could result in higher merchant pricing, even for cash customers. Similarly, the analysis does not take into account the higher average costs that the credit card network will likely face if the number of credit-card transactions decreases under surcharging. Both of these factors would raise the estimate of the welfare costs of surcharging.

The $4.5 billion cost of abolishing the no-surcharge rule is derived assuming merchants will not be in a position to act monopolistically when they surcharge. Dr Wright's work also deals with a case in which merchants can behave monopolistically at the point of sale, and shows the welfare costs of removing the no-surcharge rule to be many times higher than the $4.5 billion figure. This outcome seems likely, at least for those transactions in which the cardholder has very few alternatives available. In these cases, merchants may well take advantage of their ability to surcharge to behave opportunistically at the point of sale.[37]

The opposing extreme assumption to monopolistic behaviour is to assume merchants are perfectly competitive. Even in this case, the analysis suggests that surcharging will not lead to a better outcome than under the no-surcharge rule. Facing the no-surcharge rule, perfectly competitive merchants will always either specialise in credit-card customers, passing on the full cost of credit in their posted prices, or not accept cards and serve the non-cardholding customers. On the assumption that the remaining customers are cheaper to service, any merchant that accepts both credit-card and non-credit card customers can be undercut by another merchant that just caters to the "low-cost" customers. Thus even in the extreme case of perfect competition, there is no reason to prohibit the no-surcharge rule.

No-surcharge rules are not a new feature of credit card systems: rather, they have been in place from the earliest days of the current networks. Regardless of the situation today, there can surely be no doubt that when these systems began, they lacked any market power. It is therefore reasonable to believe that the surcharge rules are not the result of market power – rather, they are more likely to reflect efficiency considerations and form part of the broader set of arrangements needed for the card systems' optimal development. The modelling carried out by NECG confirms this view.

The Joint Study takes a different view. However, it does not put forward any rigorous analysis, much less a complete model, which would support its approach. Nor, as will be shown, does it pay much regard to the empirical evidence that could inform an assessment of the no-surcharge rule's impact.

---

[37]   In this case, even if the systems had and used market power, the no-surcharge rule would operate to prevent double marginalisation.

5.3.5   Domestic and International evidence on consequences of surcharging

The concern identified by the Joint Study in relation to the no-surcharge rule is that it prevents cost-based pricing signals from being given to the cardholder by a merchant. The presumption is that, were surcharging permitted, cardholders would face accurate signals regarding the resource cost of their decision to purchase with a credit card.

However, except in the textbook case of perfect competition, merchants do not generally set prices equal to the marginal cost of obtaining and selling products. Since the vast majority of merchants operate in imperfectly competitive markets,[38] it is quite likely that the removal of the no-surcharge rule would lead to price changes for credit card customers that did not always reflect costs.

Two examples demonstrate this point. First, taxis in Australia which participate in the Cabcharge system apply a uniform 10% surcharge to credit card payments. This rate is the same for every Cabcharge-participating taxi in Australia, for every type of credit card, from Cabcharge-branded cards to American Express, and MasterCard. The only credit card to which this surcharge does not apply is VISA, which has declined to participate in the Cabcharge system consistent with its no-surcharge rule.

Quite obviously, the costs to the taxis of accepting credit card payment will differ across credit card brands, and most likely will also differ depending on the size, transaction volume, and banking relationships of the particular taxi company. In other words, the uniform 10% Cabcharge surcharge is not cost reflective. It might be thought that this surcharge is intended to recoup the merchant service fee levied by the credit card acquirer on each taxi company. Merchant service fees do vary from merchant to merchant, but 10% is more than double the merchant fee typically charged by an acquirer to a merchant at the high end of the risk range.

Since the 10% Cabcharge surcharge is clearly not cost reflective, it does not transmit a useful price signal to cardholders which might allow them to optimally adjust their cardholding behaviour. Rather, it provides an example of rent-taking by the intermediary in a credit card transaction.

Second, regard can be had to jurisdictions overseas where legislation prohibits credit card systems from imposing a no-surcharge rule. Two such jurisdictions are the United Kingdom and Sweden. There are some subtle but important differences in the way in which the prohibitions on no-surcharge rules operate in these two countries. In the UK the prohibition is complete, in the sense that credit card systems may not impose a no-surcharge rule, and acquiring banks may not enter into contracts with their merchants under which the merchant is prevented from surcharging credit card customers.

In Sweden, on the other hand, no-surcharge rules are prohibited only at the system level. Swedish law permits acquiring banks to enter into contracts with their merchants under which the merchant is prevented from surcharging. Such merchant-to-acquiring bank agreements enforcing a no-surcharge rule are now commonplace in Sweden.

In Britain, despite the complete prohibition of "no-surcharge" and related rules (such as the "no discount for cash purchases" rule) the practice of surcharging is not widespread. The reasons for the reluctance of British merchants to surcharge credit card customers, despite the freedom to do so, are likely to differ as between merchants. Some of the likely explanations are:

---

[38]   *Indeed, Australia has one of the most highly concentrated grocery retailing sectors in the OECD.*

i. Credit cards may not actually impose greater costs on some merchants than payment by cash, debit cards, or alternative payment means. For some types of merchant (such as supermarkets, for example) handling cash is costly and risky.

ii. Merchants may not wish to lose the opportunity to make sales to the cashflow-constrained customers who would use credit in their store if there were no surcharge.

iii. Merchants may be fearful of creating resentment among their credit card carrying customers, since this could do lasting damage to their own goodwill.

Whatever the reasons, and they are likely to vary across merchants, the fact that surcharging is not prevalent even in jurisdictions where it is permitted indicates that cost reflective surcharges for some credit card purchases may in fact be zero.

Although the initial British experience does not suggest that surcharging will immediately become widespread in the absence of the no-surcharge rule, the instances in which surcharging does occur in Britain are very revealing. For the most part, surcharging seems to occur during transactions in which there is significant potential for hold-up. These include taxis, travel agents (after deciding on itineraries, dates etc) and telephone bookings (for which few alternative means of payment are available). Moreover, these instances more closely resemble the opportunism reflected in the Cabcharge arrangements, with surcharges of around 10%, rather than the mere passing on of the (much lower) merchant service fee.

Consequently, experience to date casts doubt on the Joint Study's expectation that cardholders would receive cost-reflective pricing signals if surcharging were permitted. On the other hand, surcharging will damage the brand reputation of credit card systems which permit it – all the more so, as surcharging seems so closely allied to "hold up" behaviour. If cost-reflective pricing signals do not result from surcharging, then even judged by the Joint Study's own criteria, allowing surcharging will be harmful.

5.4 Summary: the "no-surcharge" rule

The Joint Study argues that on balance the "no surcharge" rule diminishes social welfare. The above analysis suggests that this apparent conclusion is not supported by economic analysis or by the available evidence.

First, the no-surcharge rule does not have the effects which the Joint Study suggests. In particular, there is no reason to assume that customers cross-subsidise each other, so the no-surcharge rule does not have the effect of enforcing a transfer between customers of different types.[39]

Second, a cross-subsidy could arise if merchants *were* permitted to surcharge, since this would very likely lead to free-riding behaviour. This kind of behaviour is particularly likely in respect of captive customers, who will probably be charged significantly more than the merchant service fee. International evidence supports this prediction, as does Australia's own experience with the taxi industry.

Third, though most merchants accept one or more credit cards, the force that motivates this practice is free and open competition rather than anything more sinister. One of the main virtues of competition is that it compels firms to offer services to customers that they would not otherwise provide.

Finally, the most general economic model of this rule available illustrates that the no surcharge rule can be welfare enhancing, even under the Joint Study's own assumptions about merchant pricing and cardholder behaviour.

---

[39] *Transfers between customers do not in and of themselves imply any change in aggregate welfare. Nevertheless, as with its discussion of bank margins, the Joint Study appears to view such transfers as inherently bad.*

T.1

6. Are VISA membership rules too restrictive?

The Joint Study claims that:

i.   interchange fees and merchant service fees are above cost;

ii.   the interchange fee should therefore be reduced; and that

iii. excessively restrictive membership rules have perpetuated too high fees, and that these rules consequently ought to be relaxed.

This response has already addressed the first and second of these views in some detail. It has been demonstrated that the Joint Study has not established the veracity of either of these views, and that the policy prescriptions it derives from these views are more likely to do harm than good. The third proposition is studied in this section.

The starting point for consideration of membership issues is the obvious fact that the open credit card networks are membership-based associations. They are a form of joint venture – but unlike many other joint ventures, the essence of the venture, in the case of the open systems, is the creation of a product which can only exist through the membership or joint venture form.

That product is the open credit card network itself – that is, a card for which the issuing and acquiring functions are carried out by a range of distinct but inter-dependent entities, with some essential coordination and promotion functions being carried out by the system itself. The transactions costs involved in attempting to organise such a system through a sequence of bilateral bargains would doom such attempts to failure; a central agent, that pools these costs and manages the process, is needed for open systems to emerge and develop.[40]

However, given such a central agent, the open nature of the systems allows them to secure important advantages when compared to their closed counterparts.[41]

i.   Open systems can achieve economies of scope between the services the network provides and the other services provided by their members.

ii.   By aggregating activities over a wide membership base, and providing central services to these members, they can more readily obtain economies of scale.

iii. last but certainly not least, they can use competition between members to enlarge the market for card services generally, including through greater innovation, and hence promote membership of specific networks.

These benefits are more readily obtained in networks that are not only open but also that are member-managed: that is, where it is members who exercise governance rights with respect to the system. This member-managed nature is important because it encourages members to make investments (for example, in promoting the VISA brand) that are specific to a particular network. Were the networks not member-managed, members would be reluctant to make investments of this kind: for once these investments had been made, the network manager might seek to expropriate the income gains they give rise to, preventing the members from receiving a reasonable return on their investment.

---

[40]   *Open credit card systems are similar in this respect to a number of other membership-based joint ventures; these include news associations, collecting societies (that administer copyright licenses on behalf of their members) and Inter-Flora.*

[41]   *That is, when compared to firms such as AMEX or Diners' Club.*

This member-managed status means that some restrictions on membership are needed, as member-ship decisions will alter the costs and difficulties involved in managing the system, as well as the returns from sunk investments. The fact that, as in cooperative ownership structures more generally, it is the members who assume residual income risk – that is, are liable for losses incurred by the entity – makes the need for controls over membership all the greater. Because credit card systems involve very large financial exposures, both as between system members and third parties and as between the members themselves, the prudential issues associated with membership decisions also loom large.

It is against this backdrop that the systems' current membership rules have developed. The Joint Study proposes that these rules should be relaxed: that is, that the systems should be compelled to admit members that could not meet the current standards. In considering this proposal, it is useful to start by considering the criteria that generally must be met before third parties can be granted access to facilities developed by others; these criteria are set down in the *Payment Systems (Regulation) Act 1998* ("**the Act**") and in Part IIIA of the *Trade Practices Act 1974* ("**the TPA**"). Having examined whether these criteria would be met in this case (they are not), this section then turns to a specific consideration of the current membership rules, and asks whether these rules are inconsistent with economic efficiency: it emerges that the Joint Study has failed to make out any case that they are. The recommendations the Study makes are therefore poorly based.

6.1 Mandatory Third Party Access

In Australia, a high hurdle generally needs to be met before private, commercial organisations can be forced by the exercise of government powers to make their resources available to third parties.

The specific criteria that apply to payment systems are set out in the Act. That legislation establishes powers whereby the RBA may determine a right of access to a payment system, where access in relation to a payment system is defined to mean "the entitlement or eligibility of a person to become a participant in the system, as a user of the system, on a commercial basis on terms that are fair and reasonable".[42] The RBA may only impose access conditions on a payments system if it considers that doing so would be in the public interest.[43] In determining whether such action would be in the public interest, the RBA must have regard to the desirability of payment systems being financially safe for use by participants, efficient and competitive.[44] Moreover, the RBA's actions in this respect should not materially cause or contribute to increased risk to the financial system.[45]

These criteria, at a minimum, imply that third party access should not be imposed unless there is a compelling case that:

i. the current membership criteria are inefficient and/or unduly restrictive of competition; and that

ii. alternative criteria exist which would better deliver objectives of economic efficiency and com-petitiveness, without causing increased risk to the financial system.

In implementing these criteria, it is important to bear in mind the fact that access regimes, such as those established under the Act and the National Competition Policy, are an exception to the general

---

[42] *Payment Systems (Regulation) Act 1998 s. 7.*

[43] *Payment Systems (Regulation) Act 1998 s. 11(1).*

[44] *Payment Systems (Regulation) Act 1998 s. 8(a).*

[45] *Payment Systems (Regulation) Act 1998 s. 8(b).*

rule whereby the property rights of firms, which include the right to use and dispose of those rights as they please, are protected in the interests of maintaining sufficient incentives to invest and innovate. These exceptions need to be narrowly circumscribed because mandating third party access can impose significant costs.

i.   It can reduce the incentives of established firms to expand, for fear that they will pass a threshold beyond which they cannot refuse third parties the right to use the assets they have built up.

ii. It can reduce the incentives for entrants to develop genuine alternatives to the facilities and services operated by incumbents, by making it easier and less risky for them to "cheap ride" on investments the incumbents have already made.[46]

iii. It can divert resources from competition in the market to entirely wasteful competition in the courts and in the regulatory process, as access seekers and access providers jockey for legal advantage.

As a result, "declaration" of access to essential facilities is subject to carefully set out processes of review and rights of appeal under our current competition law regime. These are specified on an economy-wide basis in Part IIIA of the TPA, and more particularly, in the provisions relevant to declaration.[47] While it is clear that the provisions of the Act differ in important respects from those of Part IIIA of the TPA,[48] the general concerns that underpin the rather restrictive nature of the declaration criteria in that Part IIIA are surely of economy-wide validity.

More specifically, the Act places considerable emphasis on efficiency and competition as goals that should inform access decisions. Yet too easy access would harm these goals: it would dissuade potential access seekers from developing alternatives to existing systems and discourage the owners of the property being opened to third party use from further investment in assets they would no longer fully control.

The fact that the systems are organised as membership-based joint ventures in no way alters the need for caution in mandating third party access. Thus, neither the Act nor Part IIIA of the TPA define one set of tests that must be met by unitary firms and another, weaker, set of tests that should apply to joint ventures. This, it can be inferred, is for the good reason that business decisions as to the choice of organisational form ought not to be distorted by legal requirements unrelated to policy objectives: and there is no policy rationale for forcing joint ventures to be more accommodating of third party access than their unitary counterparts.[49]

Under these circumstances, differential treatment of open networks such as VISA would clearly send the signal to business planners that in the long term the property rights created by joint ventures are less protected than other forms of property rights. This would most likely lead to increased incentives on the margin for members of joint ventures to merge in order to protect the value of their property

---

[46]   This is often summarised by saying that access requirements can promote "intra-brand" as against "inter-brand" competition.

[47]   Declaration is the process provided for under Part IIIA of the TPA. Declaration "opens the door" for access by third parties to the services provided by a facility owner.

[48]   Specifically, it is clear that the Payment Systems (Regulation) "the Act" 1998 is intended to allow for third party access even in instances that do not amount to natural monopolies. See Explanatory Memorandum, Payment Systems (Regulation) Act, 1998, at 3.4. Whether this is a valid objective of policy is debatable; nonetheless, it is the current state of the legislation.

[49]   A similar consideration underpins the specific exemption provided to joint ventures from the deeming provisions of s. 45A(1) of the TPA by s. 45A(2).

rights. These incentives to merge created by heightened perceptions of regulatory risk would be at the expense of allocative, productive and dynamic efficiency. Quite contrary to intention, the result would be a less competitive market overall, as the rivalry in supply that open network joint ventures permit would be lost.

Even if open network joint ventures survived, differential treatment of open and closed credit card systems would inevitably distort competitive outcomes. If systems such as VISA are required to accommodate new members, but Amex remains free to dispose of its assets as it chooses, competition between them will undoubtedly be affected. For example, VISA would be naturally reluctant to undertake high-risk marketing activities, if new parties could elect to join if these succeeded – thus benefiting from the risks existing members had borne – while staying out if they failed.[50] No such dissuasion of investment would apply to the closed systems, which would remain effectively insulated from "free riders".

6.2 The Current Membership Rules

It follows from these considerations that the membership rules of the credit card systems ought not to be tampered with lightly. It is with this in mind that the arguments put by the Joint Study about membership rules need to be examined.

The Joint Study, though it criticises the current rules, does not specifically establish that relaxing the existing membership requirements would increase welfare in any way. To make out such a case, the Joint Study would have had to define an alternative set of membership rules and demonstrate that these were likely to increase the gross value derived by users of the credit card network without imposing commensurately greater costs.

The Joint Study does not apply this basic welfare test. Indeed, the study makes absolutely no attempt to define a new set of rules. Rather, it simply assumes that a "better" set of rules than those currently in place must exist.

This assumption is difficult to accept at face value. Thus, the Study itself acknowledges the complexity of the service provided by open credit card networks; the implications of this for the design of system rules must be taken into account. In particular, it should be obvious that alterations to one aspect of the system are very likely to affect other parts. Any modifications therefore need to be mindful of the rationale underpinning the current arrangements. The Joint Study, however, shows only a very partial understanding of this rationale and of its implications.

This rationale consequently merits more detailed exposition. In setting it out, it is useful to proceed in three steps: first, by identifying some broad principles of sound design that membership rules should possess; then in a second step examining the extent to which the current rules with respect to issuers conform with those principles; and finally, by considering the rules with respect to acquirers.

---

[50] *In theory, an access charge exists which will exactly compensate existing members for this risk. However, merely writing out the expression for this access charge should suffice to make it clear that its practical implementation is inevitably complex, costly and controversial. The "regulatory risk" associated with the determination of access charges will add to the uncertainties bearing on the open system's investment decisions.*

6.2.1 Design Criteria for Membership Rules

Designing a set of rules that efficiently supports complex systems, such as the credit card networks, is a formidable challenge. Inevitably, there will be tradeoffs between the goals being pursued. One way to evaluate possible rules in this "second-best" environment is to postulate a set of principles that characterise desirable rules. Views will differ on the specification of these principles and on the weights that should be attached to each, but the process of defining them has the considerable merit of focussing attention on the issues that should be important.

The following is a minimal set of such principles. Membership or access rules for members (issuers and acquirers) should:

i.   prevent free riding by members of the credit card association;

ii.   promote the setting of efficient interchange fees;

iii.   be capable of being applied internationally;

iv.   be objective, not providing unnecessary discretion to existing members;

v.   introduce little settlement risk to the credit card system; and

vi.   minimise the transaction costs related to the making and enforcement of membership decisions.

Each of these principles is discussed separately below.

Prevent free riding by members of the credit card association

The joint nature of the credit card service has been discussed at length above. In order to efficiently provide this service, the network members must to some extent subjugate their own interests to those of the network as a whole. Because of this, a fundamental criterion for an efficient access rule is that, as far as possible, it should prevent members from free-riding on the network. In other words, the access rules should promote the interests of the network over those of the members (including potential members) where these are in conflict.

Promote the setting of efficient interchange fees

The previous section of this paper has already gone into detail about the complexities involved in the setting of an efficient interchange fee. One of the implications of that analysis is that there is no single "first best" interchange fee. Rather, any such fee can involve a trade-off between insufficient membership and excessive usage. The outcome of multilateral negotiations over the interchange fee depends on the interests of the individual members and the distribution of voting rights. Thus, membership rules can have an impact on the level of the interchange fee, and this effect must be considered in the design of these rules.

International application

An international payment system such as VISA, which is potentially open to organisations wishing to be issuers or acquirers in any country, must rest on membership rules that foster the international growth of the network as well as its growth in any given jurisdiction. These rules must be tolerant to differences in respect of the types and qualities of national institutional frameworks. Inevitably, some value will be put on international uniformity, as this allows economies of scale to be achieved in rule design, enhances the legitimacy of rules and perhaps most importantly, reduces the costs members in any one jurisdiction must incur in monitoring the rules applicable to membership in other jurisdictions (bearing in mind that members in each country are likely to have financial exposures relative to members in others).

Low levels of discretion

Objective membership rules which rest on criteria involving little discretion – for example, by using criteria that are assessed by authorities other than the association members – are preferable to rules which give existing members wide discretion. The main merit of a rule that reduces the need for discretionary decision-making is that it does not allow for interpretation by present members in ways that exclude potential new members for anti-competitive purposes. As a result, it helps insure the system against self-interested behaviour by existing members, when that behaviour would be contrary to the system's long-term aims.[51]

Introduce little settlement risk

Payment systems necessarily rely on the creditworthiness of participating organisations. This is most obviously the case for issuers who must be sufficiently sound to settle credit card transactions promptly and await settlement 20 – 50 days later from their cardholders. Less obviously, acquirers must also be financially sound, as they must be capable of settling merchant transactions promptly, bearing the risk of merchant fraud and other chargebacks, and reliably handling the large financial flows which must pass through their hands. Finally, networks such as VISA must have self-insurance type mechanisms in place that guarantee settlement of all legitimate transactions, even in situations where a particular issuer or acquirer fails. Ultimately, all members benefit from the system's financial solidity; it is consequently appropriate that the burden of providing for the system's self-insurance should lie on the membership as a whole. Reflecting this, all VISA members, both issuers and acquirers, are required to be able to contribute to a financial rescue package. Contributions are generally proportional to the value of the business each member transacts. Clearly, even acquirers must be capable of bearing this strain when the need for financial rescue arises.[52]

Minimise transaction costs

Transaction costs are incurred whenever an organisation applies for membership of an open credit card network. The new organisation's application must be weighed against the membership criteria—a process which could potentially involve extensive creditworthiness and financial soundness checks. If unsuccessful, an applicant may elect to litigate, as some have done in other jurisdictions. The litigation itself, regardless of its ultimate outcome, would involve significant costs and risks. Membership rules which minimise these costs, through streamlining the applicant evaluation process and reducing the risk of litigation (especially, but not solely, successful litigation), will improve the efficiency of the payment system overall.

---

[51] *As a general matter, systems such as VISA do not have an interest in diminishing competition in the issuing and acquiring activities – rather, greater competition will generally enhance the value of the network. However, any given set of members may have such an interest. There is consequently an agency problem that non-discretionary membership rules help resolve.*

[52] *Imposing the risk associated with this contingency solely on issuers would be plainly inefficient, as it would reduce the income pool over which it is spread, unnecessarily increase the covariance between the income to the risk-source and that of the insurer, and not reflect the distribution of the benefits from the insurance being provided.*

T.1

6.2.2 Current Rules for Issuers

The principles set out above provide guideposts by which to evaluate VISA's membership rules. In essence, the *prima facie* case for the efficiency of VISA's membership rules is stronger, the greater the extent of conformity between a rule and the principles that have been discussed. It is consequently reasonable to consider the degree to which the current rules for membership as an issuer indeed meet this test.

Ideally, the rules for access to membership as an issuer would be based on the financial and business capabilities required to discharge this role satisfactorily. The role of the issuer in the credit card network is crucial to network viability, as the Joint Study recognises.

However, while it acknowledges the need for financially strong issuers, the Joint Study queries the need for issuers to be deposit-taking institutions. It nevertheless concedes that the institutional framework for authorised deposit-takers, including continuous prudential supervision by the RBA, has been "a long established and effective screening device". In other words, the credit card associations have decided to draw upon the accrued "institutional capital" and organisational memory of the Australian financial system by basing admission rules for issuers on rules evolved and developed by the prudential regulation system.

In doing so, the credit card associations satisfy three of the criteria outlined above. Requiring issuers to be deposit taking institutions:

i.   is an objective rule, removing all discretion from existing members;

ii.   minimises settlement risk; and

iii.   minimises the transaction costs related to the making and enforcement of membership decisions.

In addition, by utilising a pre-existing framework that is essentially a public good, this approach realises a potentially valuable economy of scope. In effect, any other method would incur significant costs in designing, drafting, implementing and enforcing a separate set of rules. The economies of scope arising from VISA's use of the RBA's oversight of deposit-taking institutions would consequently be lost if this membership rule were dropped.

The Joint Study suggests that "a broader set of criteria … would not be difficult to devise"; but it offers no further insights as to how this could be achieved – much less actually propose concrete criteria. The fact of the matter is that any such criteria would need to consider exactly the same matters as have already been embodied in the prudential rules operated by the RBA. In addition, they would need to support the "low levels of discretion" objective in order to avoid the possibility of anti-competitive application of the criteria. Thus, when the costs of establishing and implementing suitable alternative criteria are weighed against the convenience of using current prudential regulations, the existing arrangements hardly seem inefficient.

It is clear that the present rules for access to issuing are sufficient to ensure the integrity of the network. The qualities which issuers must possess are those which are already targeted by the criteria for becoming an authorised deposit-taking institution. Thus, it is broadly sufficient that an institution qualify as a deposit taking institution for it to qualify as an issuer.

However, it might be argued that there is a gap here between what is sufficient to ensure the effective functioning of the system and what is minimally necessary for that effective functioning to occur. Though it may be *sufficient*, it may not be *necessary* for an institution to have all of the qualities of an authorised deposit taking institution. If the benefits of admitting additional issuers were greater than the costs of designing and implementing a weaker set of rules, welfare could be enhanced by incurring these additional costs.

44

The Joint Study does not discuss this trade-off, much less establish that such a welfare increase is available. However, in considering whether it might be, it is worth noting that the requirement that issuers be deposit-taking institutions is less of a constraint than it may seem to be, and has certainly become a weaker constraint in recent years. In effect, the result of the Wallis Inquiry[53] has been to make it less difficult to obtain regulatory approval to become a deposit taking institution, and hence to become an issuer. Thus, the net welfare effect of weakening the entry rules for issuers has surely fallen, making it even less likely that the views of the Joint Study would have any welfare advantages.

Finally, the benefits to an international payment system of having a membership rule which requires prudential supervision by the central bank in each country as a starting point should not be lightly dismissed. A departure from this rule in any one country would set a problematic precedent for others. Many of these other countries may not have the institutional endowments necessary to adequately safeguard a credit card system were central bank supervision not relied upon.

6.2.3 Current Rules for Acquirers

The Joint Study is even more critical of the membership rules with respect to acquirers than of those applying to issuers. Yet here too, the Study's claims seem to be based on a poor understanding of the rationale for the system rules. More specifically, the Study ignores the efficiency basis for the "balanced portfolio" rule; and then asserts that the prudential thresholds with respect to acquirers can and should be lower than those applying to issuers. Each of these elements in the Study's approach will be considered in turn.

The "balanced portfolio" rule

The "balanced portfolio" rule states that the dollar value of acquired transactions by a member must be no more than twice the dollar value of transactions for which the same member is the issuer. Violations of this rule attract a financial penalty of 0.03% of the volume of those transactions which exceed the balanced portfolio restriction. Thus, the practical effect of the rule is to encourage all acquirers to be issuers without equally encouraging all issuers to be acquirers. Even more importantly, it means that all acquirers must at least be capable of being issuers, and hence must satisfy the membership criteria required for issuing status.

Within the VISA system, the balanced portfolio rule has been established on a regional basis – that is, it is a rule that applies within VISA's Asia-Pacific region, which encompasses Australia. Regional organisation, and a degree of devolution of rule-making, allows the system to adapt to differing circumstances, albeit at the level of groups of countries. The balanced portfolio rule was initially seen as an efficient adaptation of the system rules to the requirements of emerging markets.

The reasons for this can be best understood in the context of the entry of the VISA network into a new country.

Consider, for example, the situation in Indonesia some years ago, prior to the introduction of VISA service in that country. A pure acquirer (that is, a member that engaged in no issuing) would not be entirely dependent on the activities of local issuers for the viability of its operations. Rather, it could take the stock of foreign card-holders as given, and target merchants (such as hotels in the main cities) which could gain by accepting those foreign card-holders' cards. The interchange fees paid by

---

[53]   *Financial System Inquiry Final Report, (1997) AGPS, Canberra.*

T.1

such an acquirer would be remitted to its home country, and would therefore be unavailable to assist in the promotion of the VISA network inside Indonesia. More importantly, relative to building an issuing business inside Indonesia, this kind of acquiring is virtually a risk-free activity. It involves a very low level of sunk costs, relative to those required to expand the issuing activity domestically.

As a result, there would be an important asymmetry between issuers and pure acquirers. Given the sizeable base of foreign card-holders, the pure acquirers could likely build viable businesses without relying on domestic issuing; in contrast, the issuers would have to expend substantial sunk resources to build card-holding, and could not hope that these investments would be profitable in the absence of a widespread merchant base. This asymmetry would, under most plausible assumptions, influence the negotiations between foreign acquirers and domestic issuers over the domestic interchange fee: as the acquirers faced greater "outside options", they would be able to force domestic interchange fees inefficiently low.[54] The balanced portfolio rule curtails this inefficiency, essentially by forcing the acquirers to bear some of the costs and risks that issuing involves.

While initially designed for emerging markets, the considerations underpinning the rule do not lose their force as one moves to more developed economies or to better-established card systems.

Here too, it is likely that at any point in time, it is issuers that bear the greatest burden of sunk investments. This reflects the fact that many of the costs involved in issuing are fixed and specific to the issuing activity, but are only recouped through subsequent income from card use.[55] As a result, in interchange negotiations that involved pure acquirers, issuers would always be vulnerable to hold-up; knowing this, they would price-protect, by demanding higher interchange fees than would be needed in the absence of the hold-up threat. These higher fees would not stimulate network growth, as they would merely be compensation for added risk.

The balanced portfolio rule eliminates or at least reduces this excess cost by placing the risk of hold-up on those best able to manage it: that is, acquirers. In effect, it requires acquirers to expose themselves to the financial consequences of too-low interchange fees.[56] It thereby limits acquirers' outside options, and hence places the negotiating dynamics on a more balanced plane.

Overall, it is a well established result that negotiating processes are affected by the interests and voting rights of the negotiating members.[57] Using standard bargaining models, it is possible to construct scenarios in which, in the absence of appropriate system rules, highly inefficient interchange fees are likely to emerge.[58] Encouraging a degree of member homogeneity, notably in terms of exposure to the risks associated with a failure to agree on the main system parameters (such as the interchange fee), can be an important factor making for better system outcomes.

---

[54]   *"Outside options" refers to the outcomes each party to a negotiating process can expect in the event that negotiations break down. It is a standard result in economic models of bargaining that the greater the outside options accruing to a negotiating party, the greater the share of the benefits it can appropriate from a negotiated agreement.*

[55]   *The fact of recouping fixed outlays through variable payments is by no means a sign of inefficiency. Rather, it is a common phenomenon in activities where suppliers have difficulty in committing in advance to the quality and hence consumer value of the service being provided. A scheme in which the full fixed outlays were recouped at the moment of issuing would be difficult or impossible to reconcile with efficient interchange – that is, with securing a contribution to issuing costs from the acquiring side.*

[56]   *Because the sunk costs associated with the system lie heavily on the issuing side, symmetric insurance against too low interchange fees is not required.*

[57]   *This is a central insight of the Nash bargaining solution, which was one of the contributions for which the Nobel Prize in Economic Science was awarded to John Nash.*

[58]   *For example, if interchange were required to be negotiated bilaterally, rather than multilaterally, highly asymmetric patterns of bargaining power would emerge. Depending on the (untested) legal interpretation of other rules in the VISA system, this would result in interchange fees that were extremely high or low (see Small and Wright (NECG 2000) reproduced here at Attachment 1 for more on this). It is worth noting that this is exactly what would be required by law if the ACCC's price-fixing allegations in respect of credit card associations were to succeed.*

46

There are consequently compelling reasons for requiring acquirers to also be issuers.[59] This in turn implies that the prudential standards for acquirers cannot be meaningfully different from those that issuers must meet.

Acquirer Rules and Settlement Risk

The Joint Study ignores the transactions costs considerations that underpin the balanced portfolio rule.[60] It consequently assumes that membership rules can be designed solely in the light of prudential requirements for issuers and acquirers taken separately.

By effectively requiring that acquirers be deposit-taking institutions, the balanced portfolio rule undoubtedly minimises settlement risk. The view advanced in the Joint Study, however, is that the rule is unnecessarily severe, given (what it perceives as) the relatively low exposure of acquirers to fraud and default risk. This assertion disregards the transactions costs considerations[61] that underpin the balanced portfolio rule; but even on its on own grounds, it is highly questionable.

Thus, while it seems clear that the individual exposure of acquirers to transaction related risk is lower than that of issuers, this is an operating cost differential which is most appropriately accommodated through the interchange fee. There nonetheless remains an important issue about the financial soundness of members. This issue arises from the joint and several responsibility of association members in the event that any one of the members defaults on its obligations. In the VISA system, this responsibility rests on acquirers as on issuers, and hence entails the exposure of the system to the financial soundness of both acquirers and issuers. This reflects the fact that both issuers and acquirers benefit from the financial soundness of the system and hence should jointly bear the burden securing that soundness entails.

Clearly, it is at least conceivable that this insurance aspect could be dealt with by explicit requirements for the purchase of insurance coverage, rather than through the membership rules. However, it is very unlikely that third-party coverage would suffice, as substantial reliance on third-party cover would give rise to obvious problems of moral hazard. Particularly when the extent of the commitment involved is difficult to predict, it is more common for the parties at risk to be required to self-insure, or at least too demonstrate a high degree of financial soundness relative to the risk at issue. Monitoring the ability of acquirers to underwrite the system's financial stability is therefore an essential feature of any efficient set of membership criteria.

Effective compliance with this requirement of financial soundness – not only by issuers but also by acquirers – is all the more important given the international nature of the VISA system. The failure of either issuers or acquirers to discharge their system obligations satisfactorily would have an adverse effect not only on VISA members in Australia but also on VISA worldwide.

---

[59]   *It may be claimed that in some jurisdictions this requirement is not imposed. That is indeed the case. However, the purpose of this analysis is not to explain the membership rules applied in other jurisdictions; those need to be seen in the context of their own history and setting. Rather, the issue is to understand the relevant rules that apply in this jurisdiction. Whether other jurisdictions would gain by applying these rules themselves is a matter for the systems in those jurisdictions and their regulators, not for the RBA/ACCC.*

[60]   *An improved understanding of transactions costs and of their role in shaping the structure of economic activity has been one of the most important areas of advance in economics in recent years. Indeed, work in this area has been recognised by the award of several Nobel Prizes for Economic Science, most notably to Ronald Coase. A considerable amount of work has been done specifically on the effects of transactions costs on forms of corporate governance, including on the design of membership rules in joint ventures (and more generally in organisations, such as clubs, that provide collective goods),. The Joint Study betrays no awareness of this work, and focuses solely on prudential considerations in its discussion of system membership rules.*

[61]   *These costs extend beyond the initial admission decision. Prudential supervision is an ongoing task and the costs associated with this task are saved under the current arrangements. Changing these arrangements would forego the social economy of scope which is currently being enjoyed.*

T.1

The current rules secure this financial soundness. They do so by relying on the supervision of the central bank to ensure that acquirers will be capable of meeting the financial demands that may be imposed upon them. As with the rules regarding issuers, there is no reason to believe that this goal could be met more cost-effectively by placing on the members the responsibility of vetting the financial soundness of potential members. This is all the more the case as the higher degree of discretion entailed by such an individualised vetting procedure could be used to purposes inconsistent both with the interests of the system and of the wider public. Certainly the Joint Study does not identify alternative rules that could more efficiently do the job.

Moreover, given the international attention being directed at the current Australian investigations, there is a real possibility that a compulsory relaxation in membership rules here may flow through to other jurisdictions with less stable networks. To the extent that this induced member failures in other countries, the damage would not be limited to the VISA brand – it would also impact on Australian VISA members and end-users.

Finally, it is clear that the existing access rule prevent allegations of lack of financial soundness from being used as an anticompetitive device by members against their competitors. This would be an ongoing risk for any access rule for issuers or acquirers that was not externally administered.

In short, the Joint Study ignores the transactions costs considerations that underpin the balanced portfolio rule. It consequently overlooks the arguments for requiring acquirers to be capable of being issuers – and hence for acquirers to meet the prudential standards that issuers must meet. However, even putting this aside, there are sound reasons for the systems to ensure the financial strength of acquirers – reasons that arise both from the financial risks inherent in the acquiring function and from the need for members to underwrite the exposure of the system as a whole. The current rules allow this need to be met in a manner that is both effective and minimises discretionary choice by existing members – hence minimising the scope for anti-competitive exclusion.

6.3 Conclusions

As a general matter, a high standard of proof needs to be satisfied before owners of existing assets can be forced to share those assets with others. The appropriate nature of this standard has been extensively considered by Parliament. The results of that consideration are set down on an economy-wide basis in Part IIIA of the TPA. It is clear that the open credit card systems would not meet the requirements those provisions establish for third party access to be mandated.

The criteria set down in the Act differ from those established under Part IIIA of the TPA. Nonetheless, that Act emphasises the goals of competition, efficiency and financial stability.

"Competition" cannot be confused with the number of competitors. Even more plainly, forcing an increase in the membership of one credit card system would do nothing to promote competition if it was at the expense of the development of genuine alternatives to that system. Before it could be said with any confidence that third party access to credit card networks would promote competition, a careful analysis needs to be made of competitive conditions in the supply of payment systems. Such a study would identify whether competition was deficient in any respect, and if so, how it would be affected by mandatory third party access.

The Joint Study provides no such assessment. Indeed, it is a surprising feature of the Joint Study that though co-authored by the ACCC, it makes no attempt whatsoever to define the markets in which the services at issue are being provided, or to apply the standard techniques of competition analysis to those markets. There is consequently no way of knowing what market the Study is gauging the degree of competition in, or why it regards that degree of competition as inadequate.

This does not prevent the Joint Study from seeking potentially significant changes in the membership rules of the current systems. However, the claims the Joint Study makes do not seem to be based on a sound analysis of the current rules. No less importantly, the Study does not point to any alternative set of rules that could meet social goals more efficiently. Rather, it contents itself with assuming that such an alternative exists. It does not seem unduly harsh to suggest that this is a poor basis for policy formulation.

The risks such an approach creates are especially great in respects of systems such as VISA. These systems' reliance on the joint venture structure differs from that which characterises many other areas of economic activity. In those other areas, the joint venture form is used to supply goods and services that could well be supplied by unitary organisations. In the case of open credit card systems, however, the joint venture form is intimately related to the character of the good being supplied: namely, a credit card that can be supplied by any number of competing issuing and acquiring institutions. This provides consumers with the undoubted benefits of competition; but it involves numerous complexities that can only be effectively addressed by system rules developed and refined in the course of commercial experience.

These rules deserve a degree of deference – at the very least, they merit careful analysis in economic terms. Moreover, proposals to impose change upon them need to be mindful of the costs of distorting supply as between these systems on the one hand and the closed systems on the other.

It is, in this respect, striking that the Joint Study does not even discuss the implications of its membership proposals for the closed credit card systems. These systems do not admit members; however, increasingly, they license aspects of their activities to third parties. The clear implication of the Study is that while the closed systems would be free to license whomsoever they chose, the open systems would face membership rules imposed through the coercive powers of the Commonwealth. No rationale is provided for this distinction, which would inevitably induce evasion and avoidance on a substantial scale.

Overall, the Joint Study's discussion of membership rules is long on criticism and short on alternatives and on analysis. It is merely assumed that regulators could do better, without bothering to say how. Successful businesses such as VISA can surely legitimately demand, and expect, more than the Joint Study to date has provided.

T.1

7. CONCLUSION

The Joint Study has discussed many aspects of the 4-party credit card systems that operate in Australia and around the world. In doing so, it has characterised these systems as containing several attributes that are not socially desirable. In particular, the Joint Study has suggested that 4-party card systems:

i.   create less social welfare than debit cards;

ii. have unnecessarily high interchange fees; and

iii. have facilitated a subsidy to non-revolving users of credit cards from consumers using other payment methods.

The analysis presented in this paper shows that none of these claims is defensible. This analysis rests on some simple, and widely accepted propositions, namely that:

i.   social welfare is the difference between the value and cost of production;

ii. interchange fees help promote network membership and usage; and

iii. credit cards allow merchants to make more sales and larger sales.

Since welfare depends on valuations as well as costs, it is not possible to conclude from any analysis of costs alone whether one service is better than another. The Joint Study accepts that interchange fees serve a useful balancing role in card networks. Given this, one way to assess whether the fee is too high is whether the network is unbalanced, and since the VISA network is continuing to grow in cardholder and merchant participation this appears not to be the case for VISA. Moreover, since merchants almost always face significant fixed operating costs of doing business, an increase in sales allows them to reduce their per-unit cost, so even on this ground alone, it is not possible to conclude that credit card holders are receiving any subsidy from other customers.

The Joint Study also suggests credit card networks are mature and that, as a result, the traditional justifications for setting the interchange fee to encourage network growth are weakened. The evidence presented in this paper establishes that credit card networks in Australia are not mature; but even if they were, this would (a) not affect the relevance of the theory of interchange fees based on network externalities and (b) if anything, would imply that higher rather than lower interchange fees are optimal since (according to the Joint Study itself) merchant demand elasticity would fall as the systems became better established.

Two alternative approaches to the setting of interchange fees are put forward by the Joint Study, both being essentially cost-based. According to the established theory of interchange, an interchange fee level based solely on costs would be optimal only by chance. Optimal interchange fees depend on several factors including differences in costs, benefits, demand elasticities, and competition across the issuing and acquiring sides of the market. Given that the bulk of the costs of credit card networks fall on the issuing side, while many of the benefits flow to merchants, there is no case for the cost-based approaches to interchange put forward by the Joint Study. In fact, given the Joint Study's own characterisation of the market, analysis shows that there is a good case that the interchange fee paid by acquirers to issuers should be well above cost.

The exclusion of closed credit card networks from the scope of the Joint Study is a matter of serious concern. Understanding the differences between closed card and open card networks is central to understanding the purpose of the interchange fee. An explicit interchange fee is required for open credit card systems precisely because of their joint venture nature. Given that closed credit card systems typically have higher merchant service fees and more substantial cardholder loyalty pro-

grams, the claim that interchange fees are the cause of high merchant service fees or cardholder loyalty programs is clearly not justified. Moreover, regulating the interchange fee of open credit card networks, while failing to regulate the cardholder and merchant service fees of closed card networks, is likely to promote inefficient closed card systems at the expense of more competitive open credit card systems.

The Joint Study presents the view that two other aspects of 4-party card networks should be changed. The proposed changes are:

i.   the abandonment of the no-surcharge rule; and

ii. relaxation of the entry rules.

The likely effects of the first proposal have been shown in several ways:

i.   by analogy with shopping centre car-parks;

ii. by reference to formal modelling; and

iii. using local and international evidence.

In each case, the conclusion is the same. Permitting merchants to surcharge will lead to: sporadic holdup of captive customers; the possibility of a cross-subsidy from non-surcharging merchants to surcharging merchants; and a reduction in the value and size of 4-party card networks.

In respect of entry rules, the challenges involved in designing entry criteria for a network joint venture such as an open credit card system have been set out. Taking account of the need for:

i.   financial stability of the payment system;

ii. international consistency in rules;

iii. minimal discretion to existing members in the application of those rules; and

iv. minimal transaction costs of implementation and enforcement of the rules,

as well as the need to maximise intra-system competition, it has been demonstrated that the existing membership rules, which rely on the fact that members are deposit-taking institutions, represent a reasonable trade-off among difficult, and sometimes conflicting objectives.

Additionally, it has been explained why any changes to these membership rules would potentially impose substantial efficiency costs.

The conclusions drawn above rely on well-accepted principles of economic analysis and on economic modelling. Each of these is necessarily imperfect; but they are essential in clarifying complex issues – such as those involved in analysing open credit card systems. It is troubling that the Joint Study has made so little use of economic analysis – for example, so as to assess the welfare consequences of its recommendations. The fact that the Study provides virtually no empirical evidence (and when it does, seems to process it inaccurately) makes this concern all the greater.

The issues raised by the Joint Study are of great importance both to the institutions directly involved and to the Australian community more generally. As such, it is reasonable to expect that they will be examined with the utmost care. This is all the more the case when claims are made that successful, privately run ventures should be brought within the net of administrative regulation. As has been shown, the Joint Study is far from meeting the criteria that can reasonably be thought to apply before intervention of this kind is justified. It can be hoped that further work by the RBA and the ACCC on these issues will provide a firmer foundation than the Joint Study does for policy development.

T.1

8. APPENDIX: SPECIFIC FACTUAL ERRORS

The Joint Study report contains a number of assertions which are at odds with known facts. While these factual issues are important, they have been placed in an appendix owing to their relatively detailed nature.

8.1 Store credit

On page 14, the third bullet point is mostly incorrect insofar as it says that the store issues the card, and inconsistent with what is said at pages 47-48 regarding provision of credit.

Unless it has sold off its portfolio, David Jones issues its own cards and also does its own processing and provides its own credit - a major difficulty in selling off its store card portfolio would be finding a way of effecting a transfer of the receivables out of the retail/credit provider entity to a purchaser whilst simultaneously complying with the Consumer Credit Code requirements and not incurring substantial conveyance duty and financial institutions duty liability. So, the statement is correct in relation to David Jones, but that is a small part of the store card market.

The largest part of the store card market is constituted by the Coles Myer group (Myer, Grace Bros, K Mart, Target and Katies) store cards, which are issued by GE Capital Finance Australia (formerly called Australian Retail Financial Network), and which also does most of the processing and provides the credit. The entire portfolio was sold to GE Capital five years ago because the retailers did not want the costs of providing their own credit. Another point is that the store cards are used to generate loyalty (by special previews and offers, etc.) but, more to the point, store cards are offered to people who are not sufficiently creditworthy to get a bank credit card (as well as to creditworthy people if they apply for them).

8.2 Disclosure of direct ATM charges

Regarding paragraph 4.3, page 41, compliance with the Consumer Credit Code in regard to disclosure of a direct ATM charge would be very difficult, if not impossible as the Code currently stands. This is because the fee would be a "credit fee or charge" and would need to be disclosed in accordance with s.15G of the Code.

i.   Paragraph (a) of S.15G requires a statement of the fee or charge and when it would become payable. While it might be wordy, that could be implemented.

ii. Paragraph (b) of S.15G requires disclosure of the amount of the fee or charge if ascertainable. Given that different ATM operators may impose different fees under the Joint Study's proposed outcomes, there is a question as to whether or not the fee or charge is ascertainable because there is no guidance in the Consumer Credit Code about how much effort the credit provider has to put into ascertaining an amount. It is possible for the credit provider to find out who are all the ATM operators in Australia and then find out what their current ATM fee is and then to disclose all of them in the credit contract. That might be what is required for compliance. If the suggestion of having different fees for different ATMs depending on their costs were implemented, this would be considerably more complicated and probably more readily arguable that the amount of the fee was not ascertainable.

iii. S.15H of the Code allows the amount of a credit fee or charge to be changed so long as the credit contract permits it.

---

iv. S.61 then goes into the mechanics of making a change. It requires the credit provider (that is, the card issuer not the ATM operator) to give at least 20 days (in effect, one statement cycle) notice before there is a change in the amount of a credit fee or charge. The notice can be given by a newspaper advertisement but, even if it is a reduction in the fee or charge, needs to be notified on the card holder's next statement. Therefore, each issuer would have to have in place arrangements for receiving notification from each ATM operator before it made a change to the fee so that it could follow the procedures required by S.61.

v. The suggestion on page 42 that the fee could be posted on the ATM does not solve the Code disclosure problem as the Code requires disclosure by the credit provider/card issuer and disclosure by the ATM operator (even if the credit provider could be sure that it would always be done) would not be sufficient to comply with the strict requirements of the Code.

T.1

9. Attachment 1 - Decentralised Interchange Fees in Open Payment Networks An Economic Analysis

John Small and Julian Wright

NECG and CRNEC, University of Auckland

December 2000

Abstract

This paper considers the consequences of changing one feature of open payment system networks: the centralised setting of default interchange fees. The motivation for this work is the concern that has been expressed over possible competition law violations arising from an alleged agreement between competitors, namely the banks that perform the issuing and acquiring functions made possible by the network. Our analysis focuses on open payment card networks and demonstrates that the centralised setting of interchange fees is a necessary element of these networks. Eliminating only this feature would impose significant one-off and ongoing economic costs, expose acquiring firms to hold-up by small issuers, and lead to an escalation of interchange fees. The net effect would be to destabilise these networks and compromise their open nature. We conclude that open payment card networks would not exist without centralised interchange.

1. Introduction

Payment systems involve the exchange of assets. In barter transactions, the assets exchanged are goods and/or services, while in the exchange of goods for cash the "medium of exchange" property of money is used, enhancing the number of feasible transactions. Cheque, debit card and credit card transactions are similar in that the buyer effectively issues the seller with a transaction record that can later be used to claim value from the buyer's bank account. Each particular payment system differs in the exact technology it uses, the cost of this technology, the timing and terms of payments, the benefits that flow to end users, and the distribution of system costs through participant charges.

The aim of this paper is to consider the method by which charges are set for a particular type of payment card service. We distinguish two types of payment card service according to whether the underlying networks are open or closed. The American Express and Diners Club are proprietary payment card services and are based on a closed network. This means that a person wishing to hold an Amex card (for example) has no choice about which organisation to arrange this through. A single organisation issues Amex cards, acquires the value resulting from Amex card usage at merchant outlets, and collects the necessary payment from card holders. The same is true for Diners Club.

This approach contrasts with that used in open payment card networks, such as those operated by VISA and MasterCard. For each of these cards, potential cardholders have a large set of banks and other financial institutions from which to select a card issuing service. The participation of banks and other financial institutions allows open network credit card services to be readily integrated with other banking requirements, such as cheque and savings accounts and mortgages, providing customers with a full range of financial services at a single bank. Because of this, open network cards are quite distinctly different to closed networks. This situation is mirrored on the merchant's side of the transaction, where merchants can select from a range of acquiring banks, each of which can provide many other financial services that the merchant may require.

54

Open payment card networks are established through a set of contracts embodying the rules by which the network will operate. Amongst other things, these rules govern the terms and conditions on which cards will be issued, value will be acquired, and settlements will be effected. The aim of this paper is to consider one particular rule in some detail. This rule concerns the amount paid by acquiring banks for settlement with a cardholder's bank. The amount is known as the "interchange fee" and our interest in this paper lies primarily with the centralised method by which it is set. In practice, the interchange fee is set by a committee comprised at least partially of representatives of the card issuing and acquiring banks. Since some of these banks compete with each other in respect of the merchant acquiring function, it has been argued that the joint determination of the interchange fee has the effect of "controlling or maintaining" charges set for merchant services by competing banks. By setting the interchange fee cooperatively, it is argued, banks are effectively engaged in price-fixing.[62]

Our purpose is to provide an economic analysis of the effect of the centralised setting of interchange, and of the alternatives to this practice. In doing so, we identify a number of costs that would arise if centralised setting were prohibited. These costs, which include a severe destabilisation of the network, imply that the centralised setting of the interchange fee is necessary for the success of an open card network. Without this feature, open card networks would never have become established, and if this feature were now removed these networks could not operate as they currently do.

The costs we identify arise from the following problems:

i. *Transaction cost problem*: Bilateral negotiation in respect of interchange is the only decentralised alternative to the current arrangements. The transaction costs of bilateral setting of interchange fees would significantly impede the ability to provide an ubiquitous payment network.[63]

ii. *Divesture problem*: If it is per-se illegal for competitors to agree on interchange fees, no bank could be permitted to offer both issuing and acquiring services, since any bilateral agreement between two banks, at least one of which offers both issuing and acquiring services, is, using the logic of the ACCC's arguments, also an agreement between competitors. The necessary splitting up of the banking sector into sole issuers or sole acquirers would impose large one-off and ongoing efficiency losses.

iii. *Hold-up problem*: If interchange fees are set bilaterally, issuing institutions will have substantially more bargaining power than those that acquire. This creates the ability for issuing banks to "hold-up" acquiring banks by extracting highly skewed interchange fees during bargaining. Even though this jeopardises the viability of the whole network, banks acting in their own private interests have a strong incentive to engage in such hold-up.[64]

---

[62] Alan Frankel (1998 "Monopoly and competition in the supply and exchange of money", Antitrust Law Journal, 66, 313-361, 1998.) claims that interchange fees are an anticompetitive abuse of market power.

[63] Note that no additional choices are made available through the authorisation process embodied under Australian competition law.

[64] This problem arises because only one rule (the one complained of) is dropped, while all others remain. Far from being a weakness in our analysis, this highlights our main point - centralised interchange is a *necessary* input into the production of the service that is currently being supplied.

T.1

iv. *Escalation problem*: Given the hold-up power of issuing banks, there will be an escalation of fees that will all but destroy the payment network.[65] Although every issuing bank individually manages to secure a very large interchange fee with acquiring banks, the issuing banks will not end up being better off, and in fact are likely to be worse off as the network becomes imbalanced and implodes. Nevertheless, this does not stop each individual issuing bank further negotiating ever higher interchange fees with acquiring banks, as it takes advantage of its hold-up power to try to secure higher interchange fees than its rivals. Acting individually to maximize its own profit without regard for the other members of the network, each bank ends up destroying the very network that it relies on.

In the VISA payment system the centrally set interchange fee plays a critical balancing role and ensures that the value of the network is maximized for both cardholders and merchants. Requiring that the fee be set bilaterally will destroy this value. This is of course not to say that all payment systems require a centralised interchange fee – debit card networks for instance do not.[66] While these different payment systems clearly compete with one another, each system has its own set of rules, costs, and provides a particular set of services to consumers. The rules governing each network form a coherent and interdependent foundation which supports the payments system itself. The unilateral isolation and removal of one of these rules is akin to removing one of the cornerstones of the system. Without further attention, there is a serious risk of collapse. Our analysis suggests that when the rule to be removed is the centralised setting of interchange in an open payment card network, this risk is extremely high.

There is very little existing analysis of what would happen in the absence of a centralized interchange fee. Baxter (1983, p.577)[67] provides one hypothesis, predicting the network would greatly diminish or collapse without a centrally set interchange fee. The reasoning underlying his analysis is that when a single issuer increases its interchange fee, the decrease in demand by merchants only partially impacts on this issuer, since merchants cannot (or will not) price discriminate based on the issuing card, and so any decrease in demand by merchants is spread across all card issuers. Facing such inelastic demand, each issuing bank will set very high fees. He finds the centralised setting of the interchange fee is necessary to avoid such exploitation of the system. Our modeling of the escalation of interchange fees reveals a similar logic in a model where interchange fees are set through bilateral bargaining rather than being determined by network participants, competition between banks is imperfect, and customers' and merchants' participation in the system is determined within the model.

The rest of the paper proceeds to explain how each of these results arises in more detail, as well as detailing their full consequences. Section 2 starts by briefly reviewing the key features of the VISA payment system, and explains why a centrally set interchange fee is needed. The motivation behind the other VISA rules – the honour all cards rule and the no-surcharge rule is also explained. Section 3 discusses the transaction costs arising from the bilateral setting of interchange fees, and the various implications of the splitting up of banks into their acquiring and issuing functions. Section 4 describes exactly how the hold-up problem arises under the honour all cards rule. The formal model of the hold-up problem is presented in Appendix A. Section 5 explains the dynamics of the escalation problem that plagues the bilateral setting of interchange fees; this is based on formal modeling which is presented in Appendix B. Finally, Section 6 provides some brief conclusions.

---

[65] *The symmetric problem exists for acquiring banks, if they instead possess the hold-up power.*

[66] *The most obvious reason for this difference is that debit card networks offer only very limited free financing to cardholders. In an open credit card network, this uncharged finance provided to cardholders can be seen as both an investment in the ongoing vitality of the network and an integral part of the service provided. Without the need to fund such investment, there is no comparable role for an interchange fee in debit card networks.*

[67] *Baxter, W.F. (1983) "Bank Interchange of Transactional Paper: Legal and Economic Perspectives,"* Journal of Law and Economics, *26, 541-88.*

2. Visa's payment system

The VISA network in Australia is comprised of 26 banks, all of which issue Visa cards to consumers and nine of which also acquire VISA debt from merchants. The acquiring market is more concentrated than the issuing market, as measured by the Herfindahl Hirschmann index (HHI).[68] These banks provide many other products and services besides VISA, with the result that accurate measurement of the cost of the VISA service for either issuing or acquiring banks is difficult.[69]

VISA cardholders generally pay an annual membership fee, which varies across issuers and by customer characteristics, in return for which they receive a card that is co-branded by VISA and the issuing bank. Purchases can be made on this card without any payment being required until the end of the billing period, so that through careful use a cardholder could enjoy a non-trivial period of interest free credit. Having paid the membership fee, cardholders clearly have a strong incentive to use their cards. The value that cardholders receive from the VISA network depends on:

i.   the number and distribution of merchants that accept VISA;

ii.   the level of the membership fee;[70] and

iii. the interest cost avoided.[71]

Merchants who accept VISA pay low (often zero) standing charges but remit a "merchant discount" to their acquiring bank. This discount is set by negotiation between the merchant and the bank as some small percentage of the sale price. The value that merchants receive from the VISA network depends on:

i.   the number and distribution of VISA cardholders; and

ii.   the level of the merchant discount.

2.1 Network Effects

A "network effect" arises from the fact that merchants and cardholders both prefer large numbers of VISA members on the "other side" of the transaction. Each time a new cardholder signs up to VISA, all of the existing merchant members of VISA receive more value from the system. This is because the probability that the merchant's VISA membership will allow an additional sale to be made, has increased. Similarly, each additional merchant that begins to accept VISA creates value for all the existing cardholders because they now have more choice about where to use their VISA card.

---

[68] *As at December 1999, the issuing market had a Herfindahl-Hirschmann Index (HHI) of 1565 while the figure for the acquiring market was 1973.*

[69] *All of the familiar problems associated with cost allocation in multi-product firms arise. Fundamentally, these concern the allocation of costs that are joint and common to a wide range of products and services. This necessarily limits the accuracy of the model that VISA uses to determine the best interchange fee, and highlights the importance of other indicators of network health (such as the relative rates of growth of cardholder and merchant numbers) in determining the fee. Secondly, since banks can only imperfectly measure their own cost of VISA service, any cartel attempting to use the interchange fee to control or maintain the merchant service fees is likely to be highly unstable.*

[70] *The membership fee is most appropriately seen as a filtering device that excludes high risk cardholders. Even under quite intense competition for cardholders, we would not expect the membership fee to fall significantly for new users with limited credit histories.*

[71] *Note that this item is zero if the consumer's most preferred alternative to using VISA is to pay by cash, cheque or debit card.*

The largely variable nature of the merchants fees significantly increase the value merchants derive from the network, and reduce the risk associated with VISA membership, since they only pay when sales are generated. This also means that membership of multiple networks with this fee structure is feasible for merchants, which is why merchants typically accept several different cards. This fee structure promotes network growth on the merchant side, thereby increasing the value that cardholders receive.

Network growth on the cardholders side is promoted by making card usage attractive through the provision of an interest free period and other benefits. These benefits are not costless, however, and issuing banks need to recoup those usage costs that are not covered by the membership fee. This recoupment is primarily achieved through the interchange fee which is remitted to issuers by acquirers.

Clearly, issuing banks have other sources of funds in addition to the interchange revenue. The two which are relevant to the VISA system are the annual fee and the interest rate margin earned on credit card balances not settled by the due date. As is the case in all non-linear tariff schemes, however, there is no requirement for any single tariff element to recover all of the costs caused by the activity on which it is levied. For example, there could be sound business reasons for setting the annual fee above (or below) the annual fixed costs of servicing a cardholder. Thus, while it is clear that all card usage causes the issuer to incur some financing costs, and it is also clear that the interchange fee is the only component of issuers' revenue that is generated by card usage, these two revenue streams need not be perfectly balanced in general.

The pricing and other rules established at the network layer effectively define the VISA product and separate it from other credit cards, such as Mastercard, Diners Club and Amex, and from other payment cards, such as debit cards. All plastic cards are easy to carry and all are used in much the same way, so there is nothing about the physical manifestation of a payment card that differentiates it from any other payment card. The payment card product is defined entirely by the uses to which it can be put, and the terms and conditions of that usage. All of these characteristics must be established before the product can be marketed, since without these the product is not defined. While any single company can create its own payment card, the creation of an open credit card network requires a balance to be struck that ensures the participation of a far more diverse range of interests.

2.2 The Interchange Fee

There is no universally accepted interpretation of the interchange fee. VISA views it as a tool for balancing and growing the network, while MasterCard apparently see interchange as a fee for services provided by issuers to acquirers.[72] Others, notably Frankel (1998) consider interchange fees to be an anticompetitive abuse of joint market power. Rather than enter into a debate over the label, we concentrate on analysing the effect of interchange fees.

The current system has the effect that VISA members set a default interchange fee, although any individual members are free to negotiate other interchange fees if they choose. By the setting of a default rate, VISA members ensure an open network whereby any institution which meets VISA's standard rules can obtain the centrally set rate for interchange. This approach has the effect of promoting competition between issuers and acquirers, which has the additional effect of expanding the market for the underlying payment system.

---

[72] *Cruickshank, D., 2000, Competition in UK Banking, HMSO, Norwich*

The level of a centrally set interchange fee places a cap on the benefits that issuers can supply to cardholders, assuming that the system covers its own costs. The higher the interchange fee, the more heavily issuers can invest in attracting and retaining cardholders. This investment in system growth benefits all participants including merchants, who benefit from the network effects discussed above. High interchange fees reduce the advantages of membership to merchants, however, because these fees are recovered from the merchant discounts. The network participants therefore need to set the interchange fee at a level which balances the benefits enjoyed by cardholders against the need to attract and retain merchants.

The optimal level of the interchange fees is also important from the members' perspective for another reason. An essential feature of open credit card networks is that members can join on either side of the transaction.[73] If issuing was significantly more profitable than acquiring for any reason, members would desert the acquiring function and the network would collapse. High interchange fees could contribute to such an incentive, with the result that the VISA members must have regard to the position of both issuers and acquirers when setting this fee. Indeed, an attractive criterion against which to evaluate possible interchange fees is whether the relative profitability of issuing and acquiring is equalised.

2.2.1 Who should set the interchange fee?

There are compelling reasons for setting the interchange fee at the network level. The VISA service is jointly produced by its member banks, none of which could provide the VISA service without the co-operation of others. Because the VISA service is jointly provided, it is natural to centralise the allocation of the costs of the service.

Most of the VISA service costs arise on the issuing side, but many of the benefits of VISA flow to merchants. Centralised determination of interchange allows the acquiring side of the network to be allocated some of the joint costs. These costs are only loosely related to the cost structure of any given merchant or acquiring bank since these do not include the cost of promoting the issuing side of the service. Thus, setting interchange at the network level is the only strategy that is consistent with a thriving and well balanced network.

Though some[74] argue that there should be a zero interchange fee, the level (and hence existence) of this fee is not the subject of this paper. We note in passing, however, that without an interchange fee, issuers could not cover their cost of funds except through higher card holder fees. With higher cardholder fees, customers will not want to hold cards as much, and fewer cardholders will reduce the benefits to merchants of accepting VISA. The net result would be a smaller network.

---

[73]   It is not uncommon to place some restrictions on specialisation in order to promote network growth. In Australia, banks can fully specialise in issuing, but all acquirers are required to issue as well. This rule prevents the network being destabilised by excessive concentration on the acquiring function. The existence of this rule does not necessarily imply that acquiring is the more profitable activity, however. It may simply reflect a natural preference for serving relatively larger customers (merchants).

[74]   Frankel (1998) for example.

2.3 Other Network Rules

The VISA network includes two additional rules that are important for our analysis. The first is known as "honour all cards", and prevents merchants (and their banks) from refusing to accept VISA cards issued by any particular bank, or refusing to accept a particular type of VISA card. This rule contributes to the integrity and ubiquity of the VISA network. It has the procompetitive effect of ensuring that no issuing bank can be foreclosed from the network by the unilateral or concerted actions of acquirers. Even very small issuing banks can attract cardholders on terms that are similar to their larger rivals in respect of the acceptability of their cards. Without the honour all cards rule, the value of membership to cardholders would be dramatically reduced, since they could not be sure of paying by VISA. The confidence cardholders have in this aspect of the VISA brand is particularly important when they are traveling outside their home region.[75]

Another important rule is the "no surcharge" rule. This is designed to prevent merchants from adding anything onto their retail price to cover the cost of the merchant discount. If merchants were to discriminate in this way, the costs of operating the system could simply be passed back to cardholders with the result that it could not be used to balance the benefits to cardholders with those of merchants in order to maximise the value of the network. Although merchants, in aggregate, value this role of interchange fees, each would prefer not to be the merchant paying for it. Each merchant therefore has a private incentive to free-ride on the VISA network. The no-surcharge rule, combined with a centrally set interchange fee, prevents individual merchants from engaging in opportunistic behaviour which would reduce the value of the network.

2.4 The Incidence of Network Costs

The combined effect of centralised setting of the interchange fee, honour all cards and no surcharge, is to guarantee benefits to cardholders, and thereby drive network growth. As noted above, however, the cost of these cardholder benefits must be paid for by someone, and the incidence of these costs is poorly understood. Frankel (1998) argues that they fall on final consumers who do not pay with payment cards, providing a cross-subsidy to those who do. Therefore, the argument goes, final prices would be lower in the absence of payment cards, or under some other set of rules. This argument assumes that all merchants who accept payment cards are either perfectly competitive, or have margins that are smaller than the merchant discount. Neither of these assumptions is reasonable, nor consistent with the observed widespread acceptance of credit cards. Moreover, even if retailing were extremely competitive, there is no reason to believe that a cross-subsidy would emerge in equilibrium. In such an environment, if merchants were to set prices at cost plus the merchant discount, new entrants would be able to serve all of their non-credit card customers at lower prices. The market would then reach a separating equilibrium[76] in which merchants specialised completely in serving either credit card customers at high prices, or non-payment card customers at low prices. This would unwind any cross subsidy.

---

[75]   *A recent Joint Study into payment cards by the Australian Competition Consumer Commission (ACCC) and the Reserve Bank of Australia (RBA) acknowledged the important role of the honour all cards rule in payment card networks.*

[76]   *In this case, the term "separating equilibrium" refers to a stable situation in which merchants specialise completely in serving only one type of customer. Some merchants will accept payment cards and charge higher prices, while others will not accept payment cards and set lower prices. Thus, even under the extreme competitive assumptions that are implicit in Frankel's hypothesis, the cross-subsidy he alleges will not occur.*

A more reasonable interpretation is that retailing is imperfectly competitive and merchants choose to provide the additional service of payment through a card system without increasing their prices. The acceptance by merchants of payment cards is conceptually identical to the provision of many other services for which explicit prices are not charged, such as gift wrapping, free delivery, and lay-by facilities.

3. Bilateral negotiations and the ACCC allegation

The allegation at issue is that members collusively set a price (the interchange fee) for a particular service, and that this has the effect of "controlling or maintaining" the terms on which these members, or some subset of them, compete in a market. Putting aside for the moment the substantive issue of whether the interchange fee actually does "control or maintain" the merchant discount,[77] this paper examines the first leg of the ACCC's claim, which is that there is a pre-existing service which would be competitive were it not for the centralised setting of interchange fees.

To consider this, we need first to define the service over which competition occurs. There is no apparent sense in which the banks compete with each other in the supply of interchange services. Rather, interchange is an input into the provision by members of open credit card network services, which they then compete to provide to cardholders and merchants. Thus, we consider whether genuine competition in the supply of these services could exist without the centralised setting of interchange fees.

There is only one alternative to centralised setting of the interchange fee which is a set of bilateral interchange fees negotiated between banks.[78] By itself, this is not sufficient to avoid the "agreement between competitors" allegation, however. To see this, suppose that bank A is solely an issuer, while bank B is both an issuer and an acquirer. If A and B negotiate an interchange fee bilaterally, they are agreeing on a price which, according to the logic of the ACCC claim, controls or maintains the terms that A offers to its cardholders in competition with B. The same issue arises if A is solely an acquirer, since in this case A and B would be competing to provide services to merchants.

There is no material difference between these situations and the one which is the subject of the ACCC's allegations. In both cases, firms that are said to be in competition are negotiating over a price that is alleged to have some effect on[79] the terms of that competition.[80] Consequently, bilateral negotiation between banks is not sufficient to avoid the conduct which is the subject of the current price-fixing allegations. In order to avoid that allegation, we need to restrict the activities of each bank to either issuing or acquiring, but not both. This necessarily means splitting up the banking sector into sole issuers and sole acquirers.

---

[77] In order to establish this "controlling or maintaining" relationship, it would be necessary to present evidence that banks were either unable to reduce merchant discounts (which seems implausible given the number of other services typically provided by banks to merchants) or, as a matter of fact, do not vary these discounts across merchants.

[78] We observe that even if the members set the interchange fee at zero, the fact that this was a decision by the members would persist. Thus, under these circumstances, the ACCC allegation would remain relevant (and invalid).

[79] Note that to "control or maintain", under the Trade Practices Act 1994 implies far more than merely having an effect. From an economic perspective, a price is only "controlled or maintained" by some arrangement if significant variations in the price are not possible while that arrangement remains in force. Given the fact that merchant discounts are not set directly, the multi-product nature of the relationship between banks and merchants, and the prevalence of non-linear pricing, we expect that it would be exceptionally difficult to establish empirically that the interchange fee does actually "control or maintain" the merchant discounts.

[80] As an aside, the same could be said of any pair of telecommunications companies that negotiate over the terms of access to a network and then subsequently compete in a telecommunications market. Yet such negotiation is explicitly sanctioned by the Telecommunications Act 1997.

The breakup of issuing and acquiring functions would lead to significant economic costs. These costs can be thought of as comprising real resources that would be available for production and consumption if they were not used in the divestiture process and its ongoing consequences. Some of the economic costs of divestiture depend on the pattern of the resulting sales. Since all of the acquiring banks in Australia are also issuers of VISA, these nine banks would be most affected. Some of these could conceivably sell their acquiring business to other existing acquirers, provided the buyers of these businesses themselves divested their issuing activities. This, however, would increase concentration in both issuing and acquiring. The scenario that is perhaps least objectionable from a public policy standpoint would be if all acquirers sold their issuing business to existing issuers. This scenario is used below for illustrative purposes - it almost certainly leads to an understatement of the true costs of divestment. For example, we take no account of the ongoing effects of divestment on competition in the broader banking market.

3.1 Once-only costs of divestment

Customer change-over

One-off customer switching costs would arise for all cardholders and all merchants whose existing bank divests the service they happen to use. For instance, if John is a cardholder of bank X, and bank X drops its issuing function, then John will either have to set up a new relationship with another issuing bank, or perhaps automatically be switched to the bank which takes over bank X's business (with the possibility of switching anyway if he is unhappy with the new relationship). Either way, this process is costly to both the banks and the cardholders affected. Similar effects would arise in respect of merchants required to switch banks as a result of divestiture. A lower bound on the number of network users affected is set by the total number of merchants accepting VISA. If all acquiring functions were sold to new entrants, all merchants would be required to shift to those entrant firms. Any other divestment scenario must affect more end-users of the VISA network.

Redeployment of productive resources

There are a number of obvious one-off costs arising from divestures. These include the costs of any staff redundancies, retraining and redeployment; the cost of scrapping and repurchasing, or relocating and reinstalling specialist equipment; and alterations to software, accounts, and forms.

3.2 On-going costs of divestment

No on-us transactions

Resource costs would arise from no longer being able to enjoy any on-us (or intra-bank) transactions, which are generally cheaper than centrally cleared transactions since they do not involve matching an issuing and acquiring bank. An estimate of the cost of this matching service is the fee VISA charges member banks for providing its clearance service. This is currently set at 0.3% of the value of transactions. In Australia about 25% of all transactions (by value) are on-us transactions. Thus, a cost of about 0.075% of the value of all transactions would be incurred.

No economies of scope

The loss of economies of scope from banks no longer being able to offer both issuing and acquiring services may be a significant resource cost. Economies of scope could arise from fixed costs that are common to both issuing and acquiring functions, including: management resources; credit assessment and risk management functions; and computer hardware and software.

Greater market concentration

Additional costs are likely to arise out of the greater market concentration that would inevitably follow from the breakup of issuing and acquiring functions. When all acquiring banks are required to divest either their issuing or acquiring functions, assuming that the cardholders they divest are captured by existing issuing banks, and that the merchants they divest are captured by existing acquiring banks, then the number of banks competing on each side of the VISA transaction will suddenly and dramatically fall.[81] For example, if all acquirers sold their issuing functions to existing specialist issuers, who subsequently retained the same market shares relative to one another, the HHI for card issuing would rise from 1565 to 2011. While greater concentration is not necessarily bad, in this case it makes the possibility of anti-competitive pricing in payment services more likely, not less, especially since it would likely coincide with the demise of the current open and competitively neutral[82] VISA network.

Two suppliers rather than one

The requirement to deal with two separate banks rather than one may well impose some costs on consumers. This is likely to affect many cardholders and merchants on an ongoing basis. For cardholders, compulsory switching of card issuer will probably also mean using different banks for card services and other banking services. Similarly, any merchant who also requires card services, must now deal with separate banks. This generates ongoing costs, as the majority of cardholders and merchants will prefer a bundled service.

Less preferred supplier

Other ongoing costs arise from the customer switching process. These arise from a number of sources. Firstly, because customers originally chose their most preferred bank, forcing them to switch to another (presumably less desired) bank can lead to ongoing welfare losses. Moreover, because in many cases people choose to have their bank account at the same bank that they get their VISA card from, when their bank no longer can provide an issuing service, they will be forced to either switch their whole bank account to an issuing bank, or to have their VISA card issued by a different bank from the bank they use for their other banking services. This involves either one-off switching costs as they switch all their banking services over (and again ongoing costs from not being able to use their preferred bank provider) or the ongoing costs of having to deal with two different banks (as discussed above). These costs also arise for merchants whose bank drops their acquiring service; in this case the costs are likely to be even more significant.

---

[81] *We note that the inability to issue (acquire) cards will foreclose a substantial market segment to those banks that specialise in acquiring (issuing). The flow-on effects of this in the broader banking market are likely to be significant.*

[82] *The VISA network is competitively neutral in the sense that any institution that agrees to the rules may join and compete with pre-existing members. As will become clear, bilateral bargaining will unambiguously raise the barriers to entry, with the result that the existing participants will face less competition than is currently the case.*

T.1

3.3 Costs of bilateral bargaining

Other costs arise from the fact bilateral bargaining is more costly than a centrally agreed upon rule.

An important advantage of a centrally set interchange fee (together with the other rules of the VISA network) is that it creates a standardized and open market with minimal transaction fees. If each bank that is a member of the VISA system has to negotiate separately with every other bank to determine interchange fees (with no default rate being set by Visa), then the cost of the necessary bargaining, and inevitable bargaining failures will be significant. Failed negotiations reduce the ubiquity of the VISA network, which creates a negative externality that individual negotiating banks do not take into account. For smaller firms, these bargaining costs represent fixed costs, which are likely to act as a barrier to entry.

In Australia, there are 26 issuing banks and nine acquiring banks. Assuming the same numbers prevail after divestment (so the increase in concentration is minimised), the number of interchange fees that will need to be negotiated will be 26 x 9 or 234. Even this may be a gross understatement, however, since it assumes that interchange with the rest of the world was still set by the default rate adopted by VISA International. At the international level, bilateral bargaining would require many more separate negotiations.

When the costs of many separate negotiations are sufficiently high, it is natural for a private organization to develop a way to coordinate these negotiations. This is especially true when the terms being negotiated fundamentally relate to ensuring maximal value is created for the whole network, rather than any specific member of the network. As will be shown below, the fact that the private interests of members do not coincide with the broader objectives of network growth is the main reason that centralised interchange is required, and that bilateral bargaining is extremely dangerous.

4. Bilateral bargaining and the hold-up problem

In this section we consider what would transpire if bilateral bargaining were mandated, while maintaining the rules of the open card network operated by VISA. The outcome of bilateral bargaining is as much about what happens if two parties cannot agree, as it is about what happens if they agree. Under the honour all cards rule that VISA has in place, if two banks cannot agree on terms for interchange, the acquiring bank will not be able to offer merchant services. To see why, suppose that bank X issues cards and bank Y offers an acquiring service to merchants. If bank X and bank Y cannot reach an agreement, then there can be no interchange service between bank X and bank Y, and bank X's cardholders cannot use their cards at the merchants serviced by bank Y. However, if bank X's cardholders cannot use their cards at these merchants, then under honour all cards, bank Y's merchants will not be able to accept *any* cards, in which case bank Y will no longer be able to acquire.[83]

Because any issuing bank, no matter how small, has the ability to hold-up an acquiring bank – destroying its business if it does not agree with the issuing bank's terms, this leaves issuing banks in a very strong bargaining position. Although each issuing bank still wants to reach an agreement, so as to be able to offer an issuing service, the amount it stands to lose is far less than the acquiring bank

---

[83] *Under a different, and less obvious interpretation of the honour all cards rule, the issuing bank would not be able to issue cards unless it had agreement with all acquiring banks. Bank X would have to stop issuing cards which cannot be accepted everywhere. In this case, acquiring banks have all the bargaining power, and all the analysis that follows in the paper can be replicated for the acquiring banks rather than the issuing banks. For this reason it is sufficient to restrict attention to the more obvious interpretation of the VISA rules.*

with which it is negotiating. Suppose, there are a total of N equally-sized issuing banks and the same number of equally-sized acquiring banks, and that issuing bank X is negotiating with acquiring bank Y over the terms of interchange. If bank Y does not agree with the issuing bank's terms, it stands to lose all its acquiring business, which is proportional to N, the number of issuing banks it deals with. However, if bank X cannot reach an agreement it stands to lose only the business with bank Y. In other words, in negotiation, the acquiring bank stands to lose N times as much business as the issuing bank. This leaves the acquiring bank in a very poor bargaining position. It will be willing to accept a very high interchange fee with bank X, in order to maintain its business.

In Appendix A we provide a formal model of this hold-up problem. It shows that, using the most accepted economic model of how agreements are determined under bilateral bargaining (the Nash bargaining solution), acquiring firms are indeed left in a very poor bargaining position.[84] The Nash bargaining solution, which was introduced by Nobel prize winner John Nash in 1953,[85] says (in general terms) that bilateral bargaining will lead to profits being split in relation to the difference between what each firm can achieve through an agreement and what they will get if they disagree. In the case at hand, this implies issuing bank X will not only be able to capture half the revenue (less its cost) generated from business between itself and bank Y, but also half the profit that bank Y earns through all its other acquiring business. This outcome is further skewed once we consider the fact not all banks are equally sized. Very small issuing banks, will still be able to capture half the profit that a large acquiring bank earns, which they do through very high interchange fees.

The hold-up problem which characterises bilateral negotiation over interchange fees arises because of the setting we assumed banks operate under. The setting is exactly the one which the current allegations require us to consider: the current VISA network but without centrally set interchange fees. We have demonstrated a serious hold-up problem arises under these conditions. The asymmetric bargaining power that arises from issuing banks having hold-up power is used in the next section to show how this leads to an escalation of interchange fees – to the point that the Visa network implodes.

5. The escalation of interchange fees

In the previous section we demonstrated that under bilateral bargaining, an issuing bank will have a much stronger bargaining position when negotiating with an acquiring bank, which they exploit by securing a high interchange fee. In this section we explore the consequences of this asymmetric bargaining position, when all banks are setting their interchange fees bilaterally, and when the full dynamic effects of bank competition and demand feedbacks are allowed to play out.

Suppose, for the sake of argument, we start from a point where interchange fees had previously been set centrally so as to maximize the size of the card network. From such a point, under bilateral bargaining, the holdup problem will mean that each issuing bank can secure a higher interchange fee compared to the previously centrally set fee. When securing a higher interchange fee, an issuing bank faces the following two effects:

---

[84] *Using a simpler model of bilateral bargaining, in which the issuing bank makes a "take-it-or-leave-it" offer, leads to even more skewed outcomes, and even higher interchange fees.*

[85] *Nash, J., 1953, "Two person co-operative games" Econometrica, 21, 128-140.*

i. higher market share due to the ability to lower its fixed charge to card holders (or increase benefits to cardholders) while maintaining the same revenue per customer; and

ii. reduced overall size of the market as a result of decreased demand from merchants assuming higher merchant discounts

The first effect will dominate the second effect in terms of increasing the bank's profitability. To see why, note firstly that any reduction in demand from merchants caused by a single issuing bank raising its interchange fee reduces demand for all issuing banks in equal proportion to their size.[86] Thus, an issuing bank that increases its interchange fee gets all the benefits, but shares the costs, of such an action. More importantly, the reduced size of the network does not necessarily affect the bank's profitability. With a smaller market, assuming contestability and fixed costs to cover, the issuing firms will still end up making the same amount of profit, as will the acquiring banks. Any surplus (economic) profits are always competed away. Even without perfect contestability profits will not fall in proportion to market size since, with given fixed costs, a smaller market will typically sustain either fewer firms or less aggressive competition, or both.[87] On the other hand, by increasing its market share, an issuing firm can directly increase its profits. Thus, the primary way for an issuing bank to secure higher profits is to set a higher interchange fee than its rivals.

This dynamic leads to an escalation of interchange fees. Each issuing bank, taking as given the interchange fees set by its rivals, will only be able to secure higher profits if it can obtain higher interchange fees with acquiring banks than its rivals enjoy. In this case, regardless of the level of the interchange fees of its rivals, each issuing bank will negotiate a higher interchange fee. Given its bargaining position, this is a certainty. However, once all issuing banks have negotiated higher interchange fees, none are better off. At this point all banks are roughly in the same position they started with when interchange fees were centrally set. The only differences are that all interchange fees are now higher, issuing banks are competing either more or less aggressively for cardholders (depending on whether the effect of higher interchange fees dominates the effect of reduced market size or not), and acquiring banks are competing less aggressively to sign up merchants due to the high interchange fees they face. At this point, with profits largely unaffected, each issuing bank still faces the same superior bargaining position as before – it can extract profits from the acquiring bank through bilateral negotiations. However, in order to do this, given the now higher level of interchange fees, the issuing bank has to negotiate an even higher interchange fee than its rivals.

This escalation problem will lead to extremely high interchange fees, and a dramatically reduced card network. The number of cardholders and merchants will drop to very low levels, and the network will contract. This dynamic is likely to be accentuated by the fact the network becomes less attractive to cardholders and merchants as there are less other cardholders and merchants in the network. Generally, escalation will not stop until the card network is dramatically reduced in size, and if the network effects dominate, the network could collapse altogether.

---

[86]   This would not necessarily be the case if merchants were allowed to price discriminate across different card types – this is ruled out under the current VISA rules.

[87]   The structural competitiveness of an industry depends on the relationship between the unit costs of production for a typical firm and the level of market demand. With sufficiently high fixed costs, an industry is a natural monopoly. Similarly, if the size of a market being supplied by many firms shrinks, some may find themselves unable to cover their fixed costs and will therefore exit the market.

In Appendix B we provide a full model of interchange fee setting in the context of (imperfectly) competing issuing and acquiring banks. The model also determines how many customers will hold cards, how many merchants will accept cards, and thus the size of the network. We start by assuming interchange fees are initially set to maximize the number of transactions using the card network. We then find that at each round of bilateral bargaining the interchange fee will increase. This escalation occurs no matter how small the asymmetry in bargaining power, although the more bargaining power issuing banks have, the faster escalation arises. Figures 1 and 2 in Appendix B illustrate the outcome of this modeling. Figure 1 describes the average interchange fee at the end of each bargaining round, for two different degrees of asymmetry in the banks' bargaining positions. Clearly interchange fees quickly escalate over time, although the speed of escalation is greater when issuing banks have relatively more bargaining power. To illustrate what is happening to the network size during this escalation process, Figure 2 shows the corresponding number of total card transactions at each bargaining round, again for two different degrees of asymmetry in the banks' bargaining positions. The network quite rapidly contracts, with the rate of contraction again depending on the degree of asymmetry.

Appendix A

In this appendix we provide a formal model of bilateral bargaining over the interchange fee. The emphasis in this section is to model the bargaining process in detail, leaving the details of how the resulting interchange fee feeds back into profits and affects competition to be fully analysed in Appendix B.

We suppose initially every issuing bank i initially has n_i cardholders and there are N such issuing banks, while each acquiring bank j has m_j merchants and there are M such acquiring banks. Thus, we suppose the total number of card transactions is

$$\sum_i \sum_j z n_i m_j$$

where z is a scaling parameter. We suppose each transaction generates profit to the acquiring bank of $(r_{i,j} - a_{i,j})$, where $r_{i,j}$ is revenue per transaction earned by the acquiring bank and $a_{i,j}$ is the existing interchange fee between issuing bank i and acquiring bank j. Likewise, we suppose each transaction generates profit to the issuing bank of $(a_{i,j} - c_i)$, where c_i captures the fact that if it wasn't for interchange revenue, issuing banks incur losses rather than profits on each transaction. We consider the case of two banks, one being an issuer (denoted k) and the other an acquirer (denoted l), which are negotiating an interchange fee. If they both agree on an interchange fee of $a*$, than the issuing bank will have profits of

$$\Pi_k = \sum_{j \neq l} z n_k m_j \left(a_{k,j} - c_{k,j} \right) + z n_k m_l \left(a* - c_{k,l} \right)$$

and the acquiring bank will have profit of

$$\Pi_l = \sum_{i \neq k} z n_i m_l \left(r_{i,l} - a_{i,l} \right) + z n_k m_l \left(r_{k,l} - a* \right)$$

Under honour all cards, if they cannot reach an agreement, then the acquiring bank can no longer continue offering its service, and their respective profits are

$$\Pi_k^0 = \sum_{j \neq l} z n_k m_j \left(a_{k,j} - c_{k,j} \right)$$

and

$$\Pi_l^0 = 0$$

We use the Nash Bargaining Solution (NBS) to determine the outcome of negotiating given these payoffs[88]. The NBS is the value of a that maximises:

$$NBS = [\Pi_k - \Pi_k^0][\Pi_l - \Pi_l^0]$$

Denoting the value of a which maximizes NBS, $a*$, it is straightforward to show that

$$a* = \frac{\sum_{i \neq k} z n_i m_l (r_{i,l} - a_{i,l})}{2 z n_k m_l} + \frac{r_{k,l} + c_{k,l}}{2}$$

so that interchange revenue paid by acquiring bank l to issuing bank k is

$$\frac{\sum_{i \neq k} z n_i m_l (r_{i,l} - a_{i,l})}{2} + z n_k m_l \left[\frac{r_{k,l} + c_{k,l}}{2} \right]$$

Clearly bank k secures interchange revenue that not only pays for half of its costs of providing business with bank l and half of bank l's revenue from these transactions, but also half of bank l's profit arising from all its other acquiring business. The issuing bank's profit, resulting from bargaining, are then

$$\Pi_k^* = \sum_{j \neq l} z n_k m_j (a_{k,j} - c_{k,j}) + \frac{1}{2} \sum_{i \neq k} z n_i m_l (r_{i,l} - a_{i,l}) + z n_k m_l \left[\frac{r_{k,l} - c_{k,l}}{2} \right]$$

while the acquiring bank's profit is

$$\Pi_l^* = \frac{1}{2} \sum_{i \neq k} z n_i m_l (r_{i,l} - a_{i,l}) + z(n_k) m_l \left[\frac{r_{k,l} - c_{k,l}}{2} \right]$$

---

[88] *The Nash Bargaining Solution is the most commonly used method in economics to determine how bilateral negotiations will be concluded. The NBS assumes that the parties share of surplus will be determined by how much they can get in the event they disagree. This solution concept leads to weaker results than would be obtained by assuming that the party with the most bargaining power could make a "take-it-or-leave-it" offer. In this sense, our results are if anything an understatement of the extent to which bilateral bargaining over interchange would lead to undesirable outcomes for open credit card networks.*

From the above profit results, it is clear that the issuing bank captures half of the profit that the acquiring bank derives from its trading with other issuing banks, plus half of the revenue (less cost of issuing) that the acquiring bank earns from business with the issuing bank's cardholders. If an acquirer does not accept the terms put forward by the issuer, the acquirer stands to loose all its existing acquiring business. When negotiating with a small issuing firm, this leaves it in a very poor bargaining position. Note that the issuing bank k necessarily earns more profit than the acquiring bank l, regardless of their market shares or profitability prior to the bilateral bargain. Moreover, the interchange fee agreed through bilateral bargaining is increasing in the number of merchants the acquiring bank has, and is decreasing in the number of cardholders the issuing bank has. Thus a small issuing firm, extracts half of the acquiring firms profit by setting a very high interchange fee.

Implicit in the analysis above is that the other variables, such as the number of customers each bank secures remains constant, as does their revenue and cost per transaction. In practice, any increase in the interchange fee could have a variety of complicated effects. For the purposes of demonstrating the nature of the hold-up problem that arises from bilateral bargaining, these effects are very much second order. However, to show that taking them into account, might if anything accentuate the results already obtained, assume that the gross revenue acquirers earn per transaction increases in the interchange fee, but that as a result of a decrease in demand by merchants, the number of merchants it has decreases. As a result of a higher interchange fee we also assume that the issuing bank can offer more attractive terms to cardholders and thus increases its number of cardholders. Assuming, to a first approximation, that the total number of transactions between the cardholders and merchants remains the same[89], we can then derive the interchange fee which maximizes *NBS* above:

$$a^* = \frac{\sum_{i \neq k} z n_i m_l (r_{i,l} - a_{i,l})}{(2 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}) z n_k m_l} + \frac{r_{k,l} + c_{k,l} - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}}{2 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}}$$

Since both derivatives with respect to the interchange fee a^* are positive, the interchange fee is generally considerably higher once these two effects are taken into account. This has the effect of further increasing the asymmetry in bargaining power between issuing and acquiring banks. Bank k's profit from issuing is now:

$$\Pi_k^* = \sum_{j \neq l} z n_k m_j (a_{k,j} - c_{k,j}) + \frac{1}{2 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}} \sum_{i \neq k} z n_i m_l (r_{i,l} - a_{i,l})$$

$$+ z n_k m_l \left[\frac{r_{k,l} + c_{k,l} - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}}{2 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}} - c_{k,l} \right]$$

---

89   *To be precise, we assume that* $z n_k n_l$ *remains constant over the range of* a* *considered.*

Bank l's profit from acquiring is now:

$$\Pi_l^* = \frac{1 - \dfrac{\partial c}{\partial a^*} - \dfrac{\partial r}{\partial a^*}}{2 - \dfrac{\partial c}{\partial a^*} - \dfrac{\partial r}{\partial a^*}} \sum_{i \neq k} z n_i m_l (r_{i,l} - a_{i,l})$$

$$+ z n_k m_l \left[r_{k,l} - \frac{r_{k,l} + c_{k,l} - \dfrac{\partial c}{\partial a^*} - \dfrac{\partial r}{\partial a^*}}{2 - \dfrac{\partial c}{\partial a^*} - \dfrac{\partial r}{\partial a^*}} \right]$$

For all reasonable parameter values, the issuing bank's profit is even higher and the acquiring bank's profit is even lower, compared to before.

Appendix B

We develop a model of a card payment system that allows for

i.   imperfect competition between issuing banks;

ii.   imperfect competition between acquiring banks;

iii.   heterogeneous consumers, not all of whom will choose to hold cards;

iv.   heterogeneous merchants, not all of whom will choose to accept cards;

v.   benefits to cardholders that increase with the number of merchants who accept cards;

vi.   benefits to merchants that increase with the number of cardholders;

vii.   the merchant fee, issuers fixed fee (net of any rebate), market shares, and customer and merchant participation are all determined within the model; and

viii.   asymmetry across banks; banks can differ in customer loyalty and costs.

To make the model tractable we assume there are only two issuing banks, and two acquiring banks. In the model we will still be able to allow for different intensities of competition between issuing banks and between acquiring banks by adjusting the degree of product differentiation in each case; the service that each bank offers is differentiated from its competitor.

In the model there are a fixed number of potential cardholders (consumers) and potential suppliers who accept cards (merchants), all of whom vary in two different dimensions. Consumers vary in their preference for the bank their card is issued from. Merchants vary in their preference for the bank that provides them with their acquiring service. We use a Hotelling product differentiation model to capture these preferences. Consumers and merchants are each endowed with a value of x which is drawn from a uniform distribution on the interval [0,1].[90]

---

90 *See Laffont et al. (1998), "Network Competition,"* Rand Journal of Economics*, 29, pp.1-37 and Carter and Wright (1999), "Interconnection in Network Industries,"* Review of Industrial Organisation*, 14, pp.1-25 for models of network competition which use the equivalent form of Hotelling product differentiation. The latter paper explains why customer loyalty can be modeled in the form used below.*

A consumer with value x who has decided to hold a card, receives extra benefits

$$\theta_1 = \frac{0.25 - x}{2\sigma^I} + \frac{\beta^I}{4\sigma^I} \quad \text{and} \quad \theta_2 = \frac{x - 0.75}{2\sigma^I} - \frac{\beta^I}{4\sigma^I}$$

from getting their card issued from bank 1 and bank 2 respectively. The super-script I denotes issuing rather than acquiring (which is denoted with a super-script A). Consumers with low (high) values of x prefer the service offered by bank 1 (bank 2). The higher the value of the parameter σ^I, the more price differentials matter between the banks and the smaller are the perceived differences between the banks, represented by θ_1 and θ_2. Thus a high σ^I indicates more intense competition between banks. The parameter β^I measures the degree of asymmetry between the two banks. If one bank has higher brand loyalty on its cards than another bank, we define this to be firm 1, which corresponds to assuming β^I is positive. Such brand loyalty can capture the superior services (or complementary services) offered by one bank than another. If everything else is equal and one bank has higher brand loyalty towards its cards, it will have more cardholders. When the two banks are perceived to be identical, then $\beta^I = 0$.

Exactly the same model of product differentiation is used to model merchants' preferences towards the two competing acquiring banks. The above expressions for x only differ in that the super-script I is replaced everywhere with a super-script A. Thus, we allow a potentially different degree of product differentiation for the service offered by acquiring banks, compared to the service offered by issuing banks. Likewise, brand loyalty can differ for acquiring banks, compared to that assumed for issuing banks.

Consumers and merchants also vary in the value they place on usage of the card payment service. We assume a consumer gets a usage benefit from having a card which is $v^c(m_1 + m_2)\rho$, where v^c is uniformly distributed over v_{min}^c and v_{max}^c and measures the benefit a consumer gets for each transaction from using cards, m_1 and m_2 are the number of merchants who are serviced by acquiring bank's 1 and 2 respectively, and ρ is the fixed probability a transaction will occur between any particular consumer and any particular merchant. Likewise merchants get a usage benefit of $v^m(n_1 + n_2)\rho$ where v^m is uniformly distributed over v_{min}^m and v_{max}^m and measures the benefit a merchant gets for each transaction from cards being used, and where n_1 and n_2 are the number of cardholders who are serviced by issuing bank 1 and 2 respectively. We also assume cardholders and merchants get a fixed benefit from having the option of using the card service, which we denote v_0^c and v_0^m respectively (both of which are independent of usage and could be zero).

To normalise the scale of the card network we assume there are potentially N cardholders and potentially M merchants who accept cards. We assume cardholders pay the issuing bank they get their card from fixed amounts per-year, which we denote r_1^I and r_2^I for issuing bank 1 and bank 2 respectively.[91] We assume merchants pay their acquiring bank an amount per transaction, which we denote r_1^A and r_2^A for acquiring bank 1 and bank 2 respectively.

---

[91] *These fixed charges incorporate any rebates that customers get. Thus if higher interchange fees lead to higher rebates to customers, then this will show up as lower fixed charges in our analysis. In fact, fixed charges can become negative - in other words, on net, cardholders are being paid to join the network.*

In summary, the net utility that a consumer (with preference for issuing bank 1 of x and benefits from a card transaction of v^c) gets is [92]

$$w_1^c = v^c(m_1 + m_2)\rho + v_0^c + \frac{0.25 - x}{2\sigma^I} + \frac{\beta^I}{4\sigma^I} - r_1^I$$

while if the same consumer gets their card from issuing bank 2, they receive utility of

$$w_2^c = v^c(m_1 + m_2)\rho + v_0^c + \frac{x - 0.75}{2\sigma^I} - \frac{\beta^I}{4\sigma^I} - r_2^I.$$

Similarly, the net benefit that a merchant (with preference for acquiring bank 1 of x and benefits from a card transaction of v^m) gets is

$$w_1^m = v^m(n_1 + n_2)\rho + v_0^m + \frac{0.25 - x}{2\sigma^A} + \frac{\beta^A}{4\sigma^A} - r_1^A(n_1 + n_2)\rho,$$

while if the same merchant gets their acquiring service from issuing bank 2, they receive utility of

$$w_2^m = v^m(n_1 + n_2)\rho + v_0^m + \frac{x - 0.75}{2\sigma^A} - \frac{\beta^A}{4\sigma^A} - r_2^A(n_1 + n_2)\rho.$$

Given prices $(r_1^I, r_2^I, r_1^A,$ and $r_2^A)$ we can calculate the number of cardholders belonging to each issuing bank (n_1 and n_2) and number of merchants belonging to each acquiring bank (m_1 and m_2). To do this, we note that consumers will only get a VISA card if they expect to get a positive benefit from the card (that is either w_1^c or w_2^c is positive); and which bank they choose depends on which of these net benefits is higher. The merchants will only accept cards if either w_1^m or w_2^m is positive, and choose the bank which gives them the maximal net benefit w^m. It remains to describe how the prices $(r_1^I, r_2^I, r_1^A,$ and $r_2^A)$ are set.

We assume each bank sets its price to maximise its profits, taking the other banks prices as given, and given the consumers' and merchants' behave as describe above. Thus we look for the standard Nash equilibrium in the banks' prices. The profits of issuing bank i are

$$\Pi_i^I = n_i[r_i^I - f_i^I + (a_{i,1} - c_i^I)(m_1\rho) + (a_{i,2} - c_i^I)(m_2\rho)] - F_i^I$$

while an acquiring bank j is assumed to earn profits of

$$\Pi_j^A = m_j[r_j^I(n_1 + n_2)\rho - f_j^I - (a_{1,j} + c_j^A)(n_1\rho) - (a_{2,j} + c_j^A)(n_2\rho)] - F_j^A$$

where f is used to denote costs which are fixed for each cardholder or merchant, F is used to denote costs which are fixed regardless of the number of cardholders, merchants or transactions, and a denotes an interchange fee per-transaction.[93]

---

92 Note each transaction is assumed to have a constant size. The fact the size of the transaction does not vary with the prices set by banks is consistent with the fact that the cardholder does not pay a fee that depends on the size of the transaction, and that when purchasing using a card, the cardholder pays the same price as they would pay if they used cash.

93 Note we have allowed different interchange fees between each issuing bank and each acquiring bank, so that the bilateral setting of interchange fees can be allowed for.

Profits for issuing bank i arise from the collection of fixed fees for cards (less fixed costs of servicing each cardholder) and interchange revenue (less the costs of servicing each transaction). These profits, which are measured per-cardholder, are multiplied by the number of cardholders for issuing bank i, and then any fixed costs (which do not vary with cardholders or with the number of transactions) are subtracted. For an acquiring bank, profits per-merchant arise from the merchant discount, which is multiplied by the number of transactions received by each merchant, less fixed costs of servicing each merchant and less the interchange cost and other costs per-transaction (which are also multiplied by the number of transactions received by each merchant). These profits per-merchant are multiplied by the number of merchants belonging to acquiring bank j, and then any fixed costs are subtracted.

Thus for a given set of interchange fees, we can determine using this model the prices set by each of the issuing and acquiring banks, and simultaneously, the number of cardholders that each issuing bank has, and the number of merchants that each acquiring bank has. We call this solution the equilibrium. However, solving for this equilibrium is far from straightforward since:[94]

i.   the number of cardholders in equilibrium depends on the number of merchants;

ii.   the number of merchants in equilibrium depends on the number of cardholders;

iii. both of these depend on comparing a number of inequality constraints which vary with each customer and each merchant; and

iv.   each of the four banks sets their price to maximise their profit taking into account all of the above considerations, plus the prices of all the other banks.

To add to these complications, the interchange fees themselves have to be determined. According to the Nash bargaining solution, which we defined in Appendix A, the bilaterally determined interchange fees which come out of the Nash bargaining depend not only on the above considerations, but also on the disagreement points, which also depend on all of the above complications. To simplify, we follow the approach of Schmalensee (1999) and represent the solution to bilateral bargaining between issuing bank i and acquiring bank j as the interchange fee $a_{i,j}$ which maximises $\alpha \Pi_i^I + (1-\alpha)\Pi_j^A$, for some α greater than one-half.[95] We choose α greater than one-half to represent the fact that issuing banks have more bargaining power than acquiring banks, because of their hold-up power (as explained in Appendix A). By varying α we can capture various degrees of hold-up. By only requiring that α is greater than one-half, we show that we do not require the extreme hold-up that actually is likely to happen in practice (and predicted by Appendix A), to still get an escalation of charges under bilateral bargaining.

---

94   *We have solved this by writing a program in an econometric package which is commercially available called TSP. Our program, along with assistance in understanding it, using it and interpreting results is available upon request.*

95   *Schmalensee (1999) also assumed that the profit of the issuing banks gets more weight when determining the interchange fee which maximises the weighted sum of issuing and acquiring banks' profit. In his model, there is just one interchange fee, which either represents the rate set between a monopoly issuing bank and a monopoly acquiring bank, or between a group of issuing banks and a group of acquiring banks.*

74

Using this model, we find that for all reasonable parameter values, the interchange fee escalates (increases at each round of bargaining), even though we start from a centralised interchange fee which we pick to maximise the total number of card transactions $\rho (n_1+n_2)(m_1+m_2)$.

We illustrate the results of the model with the following plausible parameter values.[96] To normalise the potential size of the network, we assume N=10 million, M=1 million, and ρ=0.0001, so there are potentially 10 million cardholders, 1 million merchants and that the total number of possible VISA transactions in a year (the assumed length of a period), is 1 billion. We assume that cardholders get gross benefits per-transaction from using cards (ignoring benefits arising from rebates which are reflected in the fees r^I) that vary between v^c_{min} = -0.50 and v^c_{max} =2.5, so that some consumers can actually get (gross) disutility from using cards for transactions. The gross benefit v are all measured in dollars per-transaction. For merchants these benefits range from v^m_{min} =0 to v^m_{max} =5.00, so for the merchant that values cards the most, the merchant gets $5 worth of benefits per-transaction of being able to accept cards. At the other extreme, some merchants put no intrinsic value on accepting card transactions. We also assume $v^c_0 = 0$ and $v^m_0 = 10$, which says cardholders do not have any value for cards if they never use them, while merchants might put a small value on being able to accept cards even if they are never used (e.g. a merchant may value being able to say they accept cards, even if no one ever uses this facility).

We assume σ^I =0.01 and σ^A =0.01, so that both issuing and acquiring functions are equally competitive. We also assume no bank has particular brand loyalty over another, so that all banks are symmetric (β^I =0 and β^A =0). Finally, we posit the following values for the costs parameters (all values are in dollars): c^I_1 =0.50, c^I_2 =0.50, c^A_1 =0.25, c^A_2 =0.25, $f^I_1 = f^I_2 = f^A_1 = f^A_2$ =25. These number simply that the cost per-transaction of providing an issuing service is higher than the cost per-transaction of providing an acquiring service. They also imply the fixed cost per-cardholder is the same as the fixed cost per-merchant, both of which equal $25. Because the overall fixed costs F do not affect the results of this analysis at all, we leave them unspecified.

We consider two values for α to illustrate the affect of different degrees of bargaining power on the part of the issuing bank. In both cases our choice of α is conservative, relative to the likely asymmetry in the bargaining position of issuing banks arising from hold-up. To show our results hold for even very small differences in bargaining power, we use α=0.55, while to examine more realistic differences in bargaining power we take α=0.75.

---

[96]   We have analysed the model with a wide range of parameter values; in every case escalation of interchange fees results. The particular parameter values were chosen to illustrate the model, rather than to provide an exact match with actual numbers, which are in some cases are not available to even Visa, and where they may be known would in any case be confidential.

We start by finding the common and centrally set interchange fee that maximises the total volume of card transactions. We then consider rounds of bargaining, where in each round of bargaining, each issuing bank reaches a separate bilateral agreement with each acquiring bank taking as given the interchange fees set in other agreements as those existing in the last round. This captures the idea that all issuing banks reach an agreement with all acquiring banks through simultaneous but separate agreements.[97] At the end of each round of bargaining we record the average interchange fee and the total number of card transactions (per-year) for the network.[98]

Figure 1 presents the interchange fee after each round of bargaining, starting from the level which maximises the size of the network for two cases. The top curve illustrates the path of interchange fees when there is greater hold-up (α=0.75), while the bottom curve illustrates the path of interchange fees for less asymmetry in the issuing banks' bargaining position (α=0.55).

Figure 1: Escalation of Interchange Fees

---
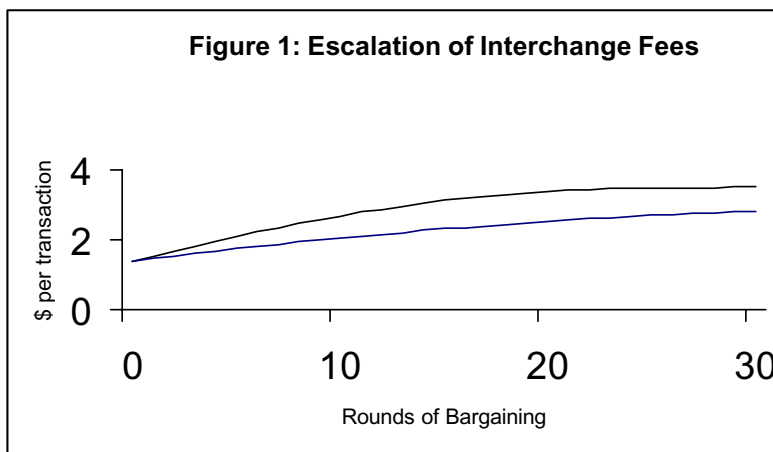
97   We have also modeled the case that agreements between banks are reached sequentially taking the agreements of the other banks as given. This leads to the same qualitative results of escalation and an implosion of the network size, but means the order of negotiation between banks affects the exact path of interchange fees.

98   Given our assumptions of symmetry, the interchange fee turns out to be the same for all banks at each round.

While escalation is clearly slower when there is less asymmetry in the banks' bargaining position, nevertheless interchange fees do still rise steadily. The result of the increase in interchange fees is initially a reduction in merchants signed up and an increase in cardholders, as cardholders receive greater rebates from merchants through high interchange fees. However, ultimately as interchange fees become very high, the level of merchants falls to a level where cardholders actually get less value from the network, and their participation can also fall.

As a result of reduced demand for the card payment service at very high interchange rates, the size of the card network will implode. Over the thirty rounds of renegotiation (or bargaining) which we consider, the network falls in size by about 75% when there is a considerable degree of asymmetry in the banks' bargaining positions (as predicted in Appendix A) and by about 33% when hold-up is quite small.

These results are illustrated in figure 2, where the bottom line indicates the case of stronger hold-up.

Figure 2: Volume of VISA Transactions
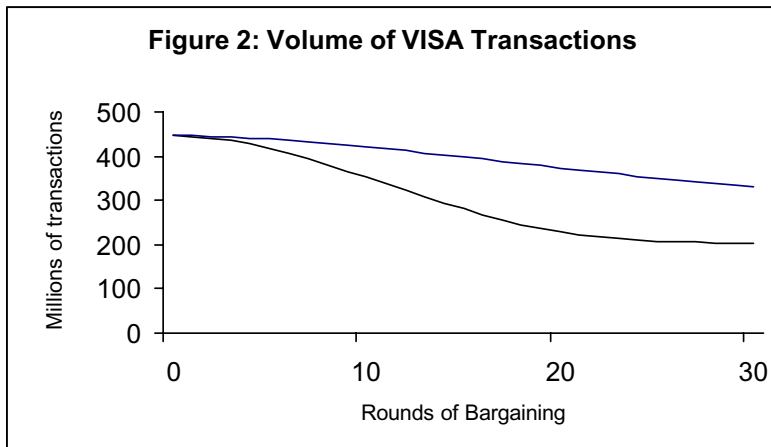
10. Attachment 2 - An economic analysis of a card payment network

Julian Wright

NECG and University of Auckland

December 2000

Abstract

This paper presents a model of a four-party card payment system to address the social optimality of the rules that govern such systems. The model extends the framework introduced by Rochet and Tirole (1999) by modeling the credit functionality of cards, by examining the optimal usage as well as membership of the card network, and by considering different modes of merchant competition. We show that the no-surcharge rule, which credit-card associations have adopted to prevent merchants surcharging customers for the use of their cards, is generally welfare improving. We also characterise the socially optimal interchange fee, and explain why regulating interchange fees on the basis of cost is likely to be inefficient.

1. Introduction

This paper provides a model of a four-party card payment system and uses it to address the social welfare effects of the rules that govern the payment system. A four-party card system (such as the ones offered by Mastercard and VISA) involves the issuance of cards (by issuing banks) to cardholders. These cardholders use their cards to make purchases with merchants who have been signed up by acquiring banks. Such systems differ from three-party card systems (such as the ones offered by American Express and Diners Club) in that they allow, and in fact encourage, competing banks and other institutions to issue cards and acquire merchants. Because for a typical transaction, cardholders pay a different institution from that which receives payment from the merchant, the card association plays the role of clearing payments between issuers and acquirers. In addition, the card association sets an interchange fee, which is a payment made by acquiring banks to issuing banks on each transaction. Apart from the setting of the interchange fee, two other rules set by card associations have attracted attention by policymakers - these being, the no-surcharge rule, which says that merchants cannot set a surcharge on goods purchased using cards (as opposed to other forms of payment) and the honour-all-cards rule, which says that merchants must accept all legitimate cards within a card system, regardless of the particular institution that issues them.

78

In the context of such a four-party card system we study the welfare implications of lifting the no-surcharge rule. We also examine implications of the level of the interchange fee between issuers of the cards and banks that sign up merchants to accept cards, examining the optimal setting of such an interchange fee.

The model analysed draws heavily on Rochet and Tirole (1999), but extends their analysis in a number of dimensions. In a seminal paper, Rochet and Tirole analyse the impact of the no-surcharge rule (which they refer to as the no-cash-discount rule) in a four-party card system. They use a Hotelling model of merchant behaviour, where the demand for card usage is derived from individual consumer preferences. Because firms set prices optimally, the impact of the no-surcharge rule on firms pricing and consumer welfare can be examined for the first time. They find that if the no-surcharge rule is lifted, the interchange fee no longer impacts on the level of credit card services, and aggregate card membership is reduced regardless of whether the interchange fee is set by the credit card association to maximise the banks' profit or by a social planner.

A potential limitation of Rochet and Tirole's model, and other models of card networks, is the fact that cardholders, by assumption, make one transaction each. In this paper we allow for the fact that in some situations cards are needed by consumers more than at other times. This leads to a clear distinction between cardholders' membership and usage decisions. We do this for several reasons.

Allowing for the fact that cards are needed more in particular circumstances can capture the credit functionality of the cards we study.[1] As well as providing regular payment services, credit cards allow consumers to make purchases in certain circumstances that would otherwise be lost or have to be delayed. For instance, this captures the fact credit cards allow consumers to purchase goods even if temporarily they do not have the cash, or funds at their disposal, to make the purchase. Credit allows consumers to make purchases based on the value offered through the purchase, rather than based on the constraints of their funds available at any particular point in time.

Compared to the existing literature, our model also captures the fact that cardholder rebates arise on transactions made, whereas cardholder fees generally arise for membership. This allows us to address one of the key concerns policymakers have with respect to allegedly high interchange fees when there are cardholder rebates and no merchant surcharges for card transactions. Policymakers appear to be concerned not so much with the number of cardholders induced by these practices, but with the possibility that there is excessive usage of credit cards by cardholders. By modeling both the credit functionality and payment functionality of credit cards we can address this concern. In particular, we characterise the optimal level of membership and usage in our model.

---

1   *For this reason, we often call the payment technology that we analyse a credit card.*

By making cardholders' membership and usage decision distinct we also clarify an important reason why absent the no-surcharge rule Coasian style bargains, in which consumers and merchants negotiate prices that internalise externalities and enhance efficiency, will not take place. In our model, in order that consumers are able to make payments with a credit card they first need to hold a card. Because the benefits of holding a card depend on the value obtained ex-post and because each individual merchant can safely ignore the effect of its decisions on consumers decision to hold a card (since it is just one of millions of merchants that consumers have the chance of dealing with), individual merchants have no incentive to negotiate agreements ex-post which lead to the efficient membership by cardholders ex-ante. Thus we explain why the no-surcharge rule and an appropriately set interchange fee can play an important efficiency enhancing role.

Using the model, we first characterise the solution to the central planner's problem. The solution depends on the likelihood of a credit card being needed to complete a transaction. When this is unlikely, the central planner's first-best solution involves allowing membership to the network up to the point where the social value generated from the marginal cardholder just covers the incremental technology costs of providing card services, and that cardholders should always use their card. When the likelihood of a credit card being needed to complete a transaction is higher than some critical point, the central planner's first-best solution involves more card members, but with some (low valuation) cardholders using their cards less.

We compare this hypothetical first-best solution, with what could be achieved with and without the no-surcharge rule in the context of three different models in which merchants set prices. In the first model we assume merchants each sell a distinct good over which they have monopoly power. In the second model we assume Bertrand competition, with merchants' selling a homogenous good. The third and final model considers an intermediate case in which firms compete according to the standard Hotelling model - that is, where there is some product differentiation (this is the model of merchant pricing that Rochet and Tirole consider).

The models highlight a basic problem that arises from surcharging - merchants pass the cost of card usage back to cardholders. Given a payment network provides value to both types of users, but only does so when there are sufficient numbers of both types of users, it is very unlikely to be socially desirable to have cardholders bear the full burden of paying for the network. Preventing surcharging, and setting an appropriate interchange fee, ensures that cardholders gain enough surplus from the network, so that the number and usage of cardholders maximises (as much as possible) the value of the network to all concerned. Even taking into account the utility of cash paying customers, we show social welfare is generally higher under the no-surcharge rule.

Our first model, where merchants have monopoly power, shows the role that the no-surcharge rule plays in preventing merchants expropriating the additional surplus that arises from the use of credit cards. If merchants are not constrained from such pricing, customers will not be prepared to pay anything to join the payment network. Moreover, an interchange fee would not help this unfortunate

situation, since merchants would simply raise the price that customers face when using credit, in response to any increase in the interchange fee. Thus when merchants have monopoly power, no one will hold or use credit cards.

In aggregate, merchants value a constraint that prevents them setting a surcharge, even though each merchant, taking the number of cardholders as given, would like to expropriate all the cardholders' surplus. The social inefficiency of dropping the no-surcharge rule arises because of a combination of an externality and a free-rider problem. Potential cardholders do not take into account the value merchants get from transactions that are made possible by customers holding cards (the externality). Merchants, on the other hand, are not willing to voluntarily subsidise customers to join the credit card network. Each individual merchant takes as given the number of cardholders, realizing that its own contribution to a system wide subsidy will have a negligible effect on the decision of any potential customer.[2]

We find that imposing the no-surcharge rule most likely increases, and never decreases social surplus, when merchants have monopoly power. We also characterize the optimal level of the interchange fee under the no-surcharge rule, which either achieves the central planner's first best solution, or if it does not, leads to the optimal trade-off between the size and usage of the payment network. When having access to the payment network is seldom required to make a transaction, there is no conflict between the optimal membership and usage of the network. The marginal cardholder has to have a sufficiently high valuation from using the payment network to justify covering the fixed cost of servicing the customer, so the marginal cardholder would want to make use of the service all the time. In this case, the central planner's first-best solution can be achieved by an appropriate choice of the interchange fee.

If, in reality, situations where credit cards are required to make a transaction are somewhat more frequent, then a trade-off arises. The optimal trade-off occurs where the social benefit created by the marginal cardholder making transactions with the credit card just equals the incremental cost of such transactions. At this point, compared to the first best solution, there can be too few cardholders, but each may use their cards too often (if there are situations where cash or other payment mechanisms are more efficient).

In our second model of merchants' pricing behaviour we take the other extreme - that of perfectly competitive merchants. In this case, under the no-surcharge rule, some merchants will accept card payments and charge more (assuming card payments are more expensive for merchants to handle than cash) while others will only accept cash and charge less. Any firm that accepts both card and cash payments is vulnerable to a competitor that undercuts its price and just accepts the low-cost cash customers. This outcome has the same welfare consequences as the surcharging case, in which merchants simply pass on all costs of payments directly to consumers.

---

2   *Equivalently, each merchant knows that a customer will use the card across a wide range of merchants, so that an individual merchant would never be able to capture the benefits of its own subsidy, designed to get cardholders to join the network.*

T.1

An intermediate case occurs where merchants set prices above marginal cost, but below the monopoly price level. To analyse this case we follow Rochet and Tirole and adopt the Hotelling model of competition, in which merchants compete over price, but where there is product differentiation. This implies merchants do not act to expropriate surplus from consumers, but rather set prices that reflect the underlying marginal costs they face. They only markup prices to the extent their product is differentiated from their rival. In this case, we show that relative to the first-best solution, dropping the no-surcharge rule generally leads to both under-subscription to the network and under-usage by each member, suggesting total social surplus is lower than achievable under the no-surcharge rule. Under surcharging, consumers get the private benefits of using credit cards, but face the full social costs of doing so. The benefits that merchants receive are not fully internalised by potential cardholders. The interchange fee can be used as an instrument to price out this externality. However, like in the monopoly model, the ability to use the interchange fees to maximize the value of the network is destroyed under surcharging.

We show that provided the probability consumers need cards to make transactions is not too high, there will be an equilibrium where all merchants participate in the network. Given this, we explain how to evaluate the welfare effects of dropping the no-surcharge rule. To get specific results we adopt some plausible parameter values and calculate the welfare losses from dropping the no-surcharge rule, comparing it to the welfare losses from dropping the rule for the other models of merchant pricing. The results demonstrate that the no-surcharge rule can enhance social welfare, even in a model where cash customers subsidise credit card customers and merchants set prices in a 'cost-plus' fashion.

Given that merchants are not allowed to (or choose not to) surcharge, the models can also be used to characterise the socially optimal interchange fee. We find this fee generally depends on all the parameters of the model. The complexity of an analysis of the socially optimal interchange fee reflects the network externalities inherent in a payment system. Card membership and usage involves costs and provides benefits to all four parties of a transaction. The interchange fee provides a way to balance these to maximize network benefits. Despite the complexity of the analysis, there are a few things that the model clearly implies.

Firstly, like Baxter (1983) and Schmalensee (1998), the model shows why it would only be by chance that the level of the socially optimal fee would be zero. The optimal interchange fee depends on the relative costs and benefits that cardholders, cash paying customers and merchants face, the extent of competition in the issuing and acquiring sides of the market, and the interaction of these factors with the network externalities arising between the different users. There is no particular reason why it should be set at zero.

Secondly, the modeling in this paper shows that setting the interchange fee based purely on the technological cost of issuing is flawed. Not only is this approach completely arbitrary, it is actually likely to be particularly inefficient. If interchange fees are set based purely on private cost there will

T.1

be too little usage of cards and too few cardholders, as cardholders ignore the benefits they generate for merchants. These benefits include those that arise when people who choose to hold credit cards are able to use them to make purchases that would otherwise be missed. In order for merchants to capture these benefits, consumers have to be encouraged to obtain and carry credit cards in the first place. This is the role of an interchange fee above cost.

Moreover, cost-determined interchange does not provide incentives for the efficient usage or provision of cards. To see this consider a situation in which issuance costs decrease, creating an opportunity for society to benefit from an increase in issuance activity (more cardholders and more card usage) relative to other activities. If the interchange fee is regulated at the cost of issuing, a decrease in costs will decrease the interchange fee and there will be no change in the gross income banks earn on each transaction, and so no particular reason for cardholder fees or rebates to change. As a result, there will be no new cardholders and no increase in card usage. To the extent the decrease in interchange fees reduces average merchant service fees, it will if anything shift the balance of network growth towards an increase in the number of merchants rather than an increase in the number of, and in usage by, cardholders.

It is sometimes observed that merchants will not in practice surcharge for small cost differences, even if they are free to do so - a concept sometimes referred to as price coherence.[3] If this is the case, it has several implications for our paper. Firstly, if the phenomenon arises because the transaction costs of differential pricing outweigh the difference between the net cost of accepting credit and the net cost of accepting cash to merchants, then the analysis of the paper still holds. In particular, if the cost of credit to merchants were to increase significantly, then whether card associations allow surcharging or not will have implications for social welfare. In this situation merchants will surcharge if allowed, and the analysis in this paper addresses the welfare implications of this surcharging. Alternatively, if price coherence is expected to remain regardless of movements in costs, then by definition welfare will not be affected by the no-surcharge rule.[4] However, even in this case, there are likely to be particular markets, where merchants have substantial market power over the terms on which customers pay (for example, in over-the-phone ticketing, in taxis, or in transactions connected with foreign travel). These markets are likely to be associated with very high surcharges, so that prices will be well above cost. If most merchants do not surcharge even if allowed, this is hardly justification to force card associations to allow surcharging. Where the result of allowing surcharging is monopolistic surcharging in a few selected markets, but otherwise unchanged pricing, the welfare implications do not require economic modelling. There is simply no good justification for a policy that would have the sole effect of allowing monopolistic surcharging in a few selected markets.

---

3   *For instance, this is the term used by Frankel (1998).*

4   *In this case one can still use our model to analyse how to set the interchange fee optimally.*

The rest of the paper proceeds as follows. Section 2 sets up the basic model, but does not specify how merchants set retail prices. Section 3 provides the first-best solution that would be adopted by a hypothetical central planner who was able to observe where and when customers needed to use the card to complete their payments. In Section 4 three different models of merchant pricing are used to examine the welfare implications of the no-surcharge rule, as well as the optimal level of the interchange fee and the optimal size and usage of the network. The rule for setting VISA's interchange fee in Australia is characterised in terms of our model in Section 5. Using this interchange fee, the model is calibrated to Australian data in Section 6 and the welfare impact of dropping the no-surcharge rule for each of the models of merchant competition is evaluated. Section 6 also compares the welfare implications of different ways of setting the interchange fee. Section 7 concludes.

2. Model set-up

Although our analysis has strong similarities to the paper by Rochet and Tirole, it differs in a number of important respects:

- The card provides value both because it is useful in general for making transactions, but also because it makes *some* transactions possible, which otherwise may not be possible; and

- because of this, a distinction between network membership and usage arises; so that

- the rebates that cardholders earn are modeled as occurring per transaction rather than on the fixed fee cardholders pay; and

- given fixed costs of card membership, the number of transactions matters; as does

- the value created by each transaction.

- Moreover, rather than assuming perfect competition on the acquiring side and imperfect competition on the issuing side of the market, as Rochet and Tirole do, we allow for some arbitrary level of markup above cost on both sides of the market.

Where we need to be specific, we will refer to the payment system analysed here as a credit card payment system and, like Rochet and Tirole, take the alternative payment technology to be cash. We start by making some explicit assumptions and defining some terminology.

(A1) All consumers want to buy a fixed number (M) of "goods" over the given time period.[5] The gross benefit of each purchase to a consumer, ignoring any benefit to using a particular form of payment, is v per-transaction. The gross cost of the item sold by each merchant is d.

---

5   *Where relevant, the time period is taken to be one-year long.*

(A2) Using a card for any transaction adds value of b_c to cardholders and b_m to merchants. These parameters measure the additional value created by using a credit card to conduct a transaction versus the relevant alternative, cash. We assume consumers have types b_c which are continuously distributed according to the density function $h(b_c)$ and distribution function $H(b_c)$.[6] In contrast, we assume all merchants are identical, so that b_m is the same for all merchants.[7] Consumers know their own b_c, and both parties know the distribution $h(b_c)$ and the value of b_m.

(A3) Merchants and cardholders also derive benefits from using credit cards because this allows them to carry out some transactions which otherwise would be missed. Specifically, suppose that for each good that a consumer wishes to purchase, there are two possible states of the world - state 1 and state 2. State 1 occurs with probability q and state 2 with probability 1-q. In state 1 consumers will only be able to make a purchase if they have a credit card and merchants accept credit cards. Consumers know the state that applies to each of their purchases, and also whether each merchant accepts cards or not. Merchants do not know the state particular cardholders are in when they set their prices.[8]

(A4) A transaction that is done using credit cards costs the issuing bank c_c and the acquiring bank c_m. These are technological costs as opposed to costs that acquiring banks might face through the interchange fee (denoted a), or costs that issuing banks may face from providing loyalty benefits or interest-free terms to cardholders. We also allow there to be a fixed costs per-cardholder of F_c.[9] Like Rochet and Tirole we assume that the technological costs of cash are zero, although in practice there are clearly many costs that merchants face to accepting cash. One view of the transactional benefits of accepting credit cards (b_m) is that it measures the costs that merchants save by not having to handle cash for a transaction.

(A5) We assume that as a result of bank competition, prices emerge for cardholders and merchants which take the following form. Cardholders are charged a fixed amount (annual fee) r_c, which does not have to equal the per-customer fixed cost F_c, and then receive a rebate of $f_c(a) = a - c_c - \pi_c$ per credit card transaction, where π_c allows for a markup over marginal cost, earned by issuing banks.[10]

---

6   We could equivalently assume that there are a finite number of cardholders, and each has their benefit b_c independently drawn from the distribution $H(b_c)$.

7   Rochet and Tirole also assume that cardholders receive a benefit per-transaction that varies across different cardholders while assuming merchants all receive the same benefit per-transaction.

8   Thus, state 1 does not refer to the type of product being sold or the merchant. Rather it should be thought of as the financial state of the consumers, and/or the urgency with which the purchase is required. More generally one could interpret q as the percentage discount in value that consumers face if they are not able to use credit in making their purchasing decisions. Some purchases will have to be delayed, some made smaller, while others may be lost altogether.

9   In Rochet and Tirole's model there is no distinction between fixed costs per-cardholder and marginal costs of transactions since cardholders all make only one transaction.

10   Once the fixed and common costs of the services provided are allowed for, such pricing does not necessarily imply any supernormal profits are earned by issuing institutions.

This rebate per-transaction captures the interest-free period cardholders' receive and any loyalty programs that provide rewards per-transaction, which is assumed to be funded through the interchange fee.[11] Acquiring banks charge merchants a per-transaction fee $f_m(a) = a + c_m + \pi_m$, where π_m denotes the markup over marginal costs earned by acquiring banks.[12]

(A6) We assume

$$\pi_c + \pi_m + (r_c - F_c)/M \geq 0,$$

so that the average margin to the banking sector per card-holder transaction must be non-negative (and where there are fixed or common costs to cover the inequality must be strictly positive).

Clearly, in reality issuing and acquiring banks will face different costs and different profitabilities with respect to different cardholders and different merchants. Moreover, each issuing and acquiring bank will have their own distribution of these variables across cardholders and merchants. One could think of our measures of costs, profitabilities, rebates and merchant discounts above as capturing the mean of their respective distributions.

Given these assumptions, a cardholder of type b_c will receive utility of $b_c + v - p_{card} + f_c(a)$ if they use their card for the transaction and $v - p_{cash}$ if they use cash, where p_{card} is the price charged by the merchant if they use their card and p_{cash} is the price charged by the merchant if they use cash. Merchants, likewise, will receive $b_m + p_{card} - d - f_m(a)$ if a card is used, and $p_{cash} - d$ if cash is used. In the case customers do not have a credit card and state 1 occurs, there will be no sale, and the benefit received by both consumer and merchant is zero. These payoffs are summarized in the following table:

Payoff table for a transaction

	Use Card		Use Cash	
	State 1: prob q	State 2: prob (1-q)	State 1: prob q	State 2: prob (1-q)
Customer	$b_c + v - p_{card} + f_c(a)$	$b_c + v - p_{card} + f_c(a)$	0	$v - p_{cash}$
Merchant	$b_m + p_{card} - d - f_m(a)$	$b_m + p_{card} - d - f_m(a)$	0	$p_{cash} - d$

---

11 *Another way of interpreting the equation is that the margin earned on transactions must be the difference between interchange revenue and the costs of providing transactions and covering loyalty and interest-free benefits, when measured on a per-transaction basis.*

12 *The assumption that customers face fixed fees and merchants face usage fees is an approximation to actual pricing.*

The timing of decisions is then as follows.

(1) Payment system rules are set. In particular, a rule is set whereby merchants are either allowed to set a surcharge for card payment, or not. Also the centralized interchange fee "a" is set.

(2) Issuing and acquiring banks set their prices for issuing and acquiring respectively - these are determined exogenously given the interchange fee.

(3) Consumers and merchants decide whether to join the payment network.

(4) Merchants set prices for goods (p_{card} and p_{cash}).

(5) States of nature are observed.

(6) Consumers decide where to purchase and what payment method to use.

The ordering of these decisions is quite natural. Payment system rules are set first. Given these, banks set their prices for issuing and acquiring. Consumers and merchants observe these and decide whether to join the network. Each individual merchant, realising that its own pricing will have a negligible effect on the decision of any potential customer to join the card network, ignores the effect of its pricing on consumers' membership decision in working out its optimal prices. This is modeled by having merchants set their prices after membership decisions are made. However, merchants set their prices prior to the states of nature being observed, since merchants will not generally know whether a particular consumer will need to use their credit card or not for a particular transaction. On the other hand consumers will generally know whether they will need to use credit or not to make a transaction, and which merchants accept credit or not. Thus, their purchase decision is modeled as occurring in the last stage taking all theses factors as given.

The final step in closing the model is to detail how many merchants there are, how these merchants set their prices, and how then consumers choose where to purchase. To do this, we will consider three different models of merchant pricing. However, before doing so, we first calculate the central planner's first-best solution.

3. Central planner's solution

The first-best solution involves calculating the optimal size of the card network (which types of consumers should hold credit cards), and when cards should be used. Because cards provide more social value in state 1 than state 2, there are two separate cases a social planner would consider. If the probability of state 1 occurring (q) is less than some critical value, then given the fixed cost of membership, a social planner would ensure every cardholder used their cards at all times. For higher values of q, the central planner will have only some cardholders using their cards in state 2.

Case (a) $q < F_c/(M(v-d))$

In this case, even the marginal cardholder generates social surplus by using their card in state 2. A social planner, who could control the use of cards across the different states of the world, would have consumers own cards provided the social benefits arising across both states $[q(b_c+b_m+v-d)+(1-q)(b_c+b_m)]M$ exceed the technological costs $(c_c+c_m)M+F_c$. Thus, the first-best solution, has the marginal cardholder having benefit $b_c=b_s$, where b_s is defined as:

$$b_s = c_c+c_m-b_m+F_c/M - q(v-d) \tag{1a}$$

All cardholders with benefits greater than b_s should join the network. Moreover, it would be optimal for all cardholders that do belong to the network to use their cards in both states since $b_s+b_m>c_c+c_m$ from (1a).

Thus the marginal cardholder to use credit in state 2 is

$$b_u = b_S = c_c+c_m-b_m+F_c/M - q(v-d) \tag{2a}$$

The first best level of usage of the network is achieved when all cardholders use their cards all the time. The total social surplus generated from this first-best outcome is

$$TS = M\int_{b_s}^{\infty}(b_c + b_m + v - d - c_c - c_m)h(b_c)\,db_c + M\int_{-\infty}^{b_s}(v-d)(1-q)h(b_c)\,db_c - \int_{b_s}^{\infty}F_c h(b_c)\,db_c \tag{3a}$$

Case (b) $q \geq F_c/(M(v-d))$

Under this condition there is a high option value of holding a card. It will be optimal for some lower valuation customers to join the network although they should only use them in state 1. Thus, a social planner who could control the use of cards across the different states of the world, would have consumers own cards provided the benefits arising from usage in state 1 $q(b_c+b_m+v-d)M$ exceeded the technological costs $q(c_c+c_m)M+F_c$. Thus, the first-best solution, has the marginal cardholder having type $b_c=b_s$, where

$$b_s = c_c+c_m-b_m+F_c/(qM)-(v-d) \tag{1b}$$

88

Given those consumers with $b_c > b_s$ have cards, the social planner would restrict usage in state 2 to only those cardholders who generate benefits from using cards in state 2 that exceed the additional cost of doing so ($c_c + c_m$). Thus, the cardholder who is on the margin between using credit and cash in state 2 has benefits defined by $b_u + b_m = c_c + c_m$ or equivalently

$$b_u = c_c + c_m - b_m \qquad\qquad (2b)$$

In this case, if only cardholders with transactional benefits b_c greater than b_u are allowed to use cards in state 2, the first best level of usage of the network can be achieved. The total social surplus implied by this outcome is then equal to

$$TS = M \int_{b_u}^{\infty} (b_c + b_m + v - d - c_c - c_m) h(b_c)\, db_c + M \int_{b_s}^{b_u} q(b_c + b_m + v - d - c_c - c_m) h(b_c)\, db_c$$
$$+ M \int_{-\infty}^{b_u} (v - d)(1 - q) h(b_c)\, db_c - \int_{b_s}^{\infty} F_c h(b_c)\, db_c \qquad\qquad (3b)$$

4. A welfare analysis of eliminating the no-surcharge rule

To analysis the welfare implications of eliminating the no-surcharge rule, we consider three different models of how merchants price. These are (i) monopoly pricing; (ii) perfect competition; and (iii) Hotelling competition.

4.1 Merchants with monopoly power

In this section, we assume merchants are monopolists.[13] Thus, we assume each merchant produces one of the M different goods that the consumers buy. Equivalently, we could assume that there are M/p merchants, where **p** is the exogenous probability that a consumer will want to buy the good supplied by any particular merchant. We consider both the case surcharges are allowed, and they are not.

---

13   *We do not actually need to assume that merchants are monopolists over the goods they sell. Rather, if merchants compete over the cash price of the good they sell, but have market power over their charge for card payment then we can derive similar results to those below.*

89

(A) Surcharges allowed

Suppose the marginal cardholder is of type b_1; that is, the value of b_c for the consumer that just wants to join the card network. Under surcharging, merchants will set a price for using credit that at least extracts all the payment benefits that the marginal cardholder receives. That is, p_{cash} will be greater than or equal to $v+b_1+f_c(a)$. Knowing this, the cardholder of type b_1 would never join the network in the first place, since joining entails a fixed cost but provides no benefits. This means the valuation of the marginal cardholder is in fact greater than b_1, which contradicts the definition of b_1 as the valuation of the marginal cardholder. No matter how we define the marginal cardholder, only cardholders with somewhat higher valuations will want to join. By logical deduction, no consumer will choose to hold cards in the first place. Consequently, there will be a single price of goods sold ($p_{cash}=v$). Although it would be in the collective interests of merchants to pay cardholders to join the network, since each takes the supply of cardholders as given, each merchant will behave opportunistically. The result will be that there are no cardholders.

Total social surplus is then easily calculated as the value created when all sales are cash. This is simply

$$TS = (1-q)(v-d)M \tag{4}$$

Under-subscription of the card network compared to the first-best level is equal to the fraction of cardholders in the first-best solution. This is

$$(1-H(b_s))>0 \tag{5}$$

This provides a measure of how many less cardholders there are relative to the first-best level. Since everyone uses cash with surcharging, there is also too little use of credit in state 2 (as well as of course in state 1). Relative to the first-best solution, the under-usage of credit in state 2 is

$$M[q(1-H(b_s)) + (1-q)(1-H(b_u))]>0 \tag{6}$$

(B) No surcharges allowed

In this case, each merchant must set a common price for cash and credit. We denote this price p^*. As we will show, for some parameter values there may be no cardholders, or no merchants that accept cards. In these cases the welfare under the no-surcharge rule is no different to the case where surcharges are allowed. However, for most reasonable parameter values, there will be some cardholders, and merchants will accept cards. In these cases, we show social welfare is higher compared to the case where surcharges are allowed. Thus, unambiguously, welfare is never lower, and usually higher under the no-surcharge rule. To show this result we first prove the following lemma.

90

Lemma 1

In any equilibrium where cards are used for some transactions, the following conditions will be satisfied

$$b_c{}^* + f_c(a) > 0 \tag{7}$$

$$p^* = v \tag{8}$$

$$b_m \geq f_m(a) - q(v-d) \tag{9}$$

where $b_c{}^*$ is the valuation of the marginal cardholder.

Proof

In any equilibrium where cards are used, all merchants must accept cards. This follows because all merchants are identical, and there is no competition between them. Moreover, the marginal cardholder must be willing to join the network in the first place. If $b_c{}^* + f_c(a) \leq 0$ and $p^* > v + b_c{}^* + f_c(a)$, the cardholders will never want to use their cards, contradicting that the marginal cardholder will join the network. However, if $b_c{}^* + f_c(a) \leq 0$, a monopolist will never set $p^* < v + b_c{}^* + f_c(a)$ since this would imply lower margins than could be obtained from cash customers without generating any extra sales (since the number of cardholders is given). Thus, if $b_c{}^* + f_c(a) \leq 0$ then it must be that $p^* = v + b_c{}^* + f_c(a)$, which leaves no surplus for the marginal cardholder to cover their costs of network membership r_c. Thus any equilibrium must have $b_c{}^* + f_c(a) > 0$.

Given that $b_c{}^* + f_c(a) > 0$, cardholders will always use their cards rather than cash. If merchants set $p^* > v$, there will be no cash customers since such customers obtain no surplus. Facing only card customers, merchants will set p^* to extract the surplus at least from the marginal cardholder. However, given this, the marginal cardholder will not join the network in the first place, thus contradicting the definition of a marginal cardholder. Thus, any equilibrium must have $p^* \leq v$. However, merchants will not want to set $p^* < v$, since in so doing they will earn less from cash customers, and not attract any additional card transactions. The reason for the latter result is that given $b_c{}^* + f_c(a) > 0$, all cardholders use their cards instead of cash, and so there are no more card transactions to attract by setting a lower price. Thus any equilibrium must have $p^* = v$.

If merchants accept cards, we have shown they will want to set $p^* = v$. Their profits are

$$(1-q)(v-d)H(b_c{}^*) + (v-d+b_m-f_m(a))(1-H(b_c{}^*))$$

91

The alternative to this is that they do not accept cards at all. In this case they achieve profits of

$$(1-q)(v-d)$$

They will accept cards provided

$$(1-q)(v-d)H(b_c^*) + (v-d+b_m-f_m(a))(1-H(b_c^*)) \geq (1-q)(v-d)$$

which is true if and only if

$$b_m \geq f_m(a)-q(v-d). \hspace{3cm} QED$$

Given the conditions above, the marginal cardholder compares her utility from joining with not joining, so that

$$[b_c^*+f_c(a)]M-r_c = 0, \hspace{3cm} (10)$$

where the utility from each transaction is multiplied by the expected number of merchants M for which transactions are made. Solving (10) for b_c^* implies

$$b_c^* = r_c/M - a + c_c + \pi_c$$

Using these results we can now show the following proposition.

Proposition 1

In any equilibrium where cards are used at all, social surplus is higher under the no-surcharge rule.

Proof

Total surplus under the no-surcharge rule is

$$TS = M\int_{b_c^*}^{\infty} (b_c+b_m+v-d-c_c-c_m)h(b_c)\,db_c + M\int_{-\infty}^{b_c^*} (v-d)(1-q)h(b_c)\,db_c - \int_{b_c^*}^{\infty} F_c h(b_c)\,db_c \hspace{1cm} (11)$$

Comparing this to the total surplus under surcharging, which is $(1-q)(v-d)M$, we find that total surplus under the no-surcharge rule is higher if and only if

$$M\int_{b_c^*}^{\infty} (b_c + b_m + q(v-d) -c_c - c_m - \frac{F_c}{M})h(b_c)\,db_c$$

is positive. Using the conditions that $b_c^*= r_c/M - f_c(a)$ and that $b_m \geq f_m(a)-q(v-d)$, it follows that

---

92

T.1

$$b_c^* + b_m \geq f_m(a) - q(v-d) + r_c/M - f_c(a) = c_c + c_m + \pi_c + \pi_m + r_c/M - q(v-d)$$

and so a sufficient condition for social welfare being higher under the no-surcharge rule is assumption A6; that $\pi_c + \pi_m + (r_c - F_c)/M \geq 0$.

<div align="right">QED</div>

Clearly, if the conditions for there to be an equilibrium in which there are some cardholders do not hold, then social welfare will be the same under surcharging or no-surcharging – as there will be no cardholders in either case. However, this is generally not optimal. Under the no-surcharge rule, the interchange fee can be adjusted to ensure there is an equilibrium. In this case, it is instructive to compare the outcome under the no-surcharge rule with the first-best outcome that a omniscient central planner would choose. We consider both of the cases from section 3.

Case (i) $q < F_c/(M(v-d))$

In this case the central planner would have cardholders using their cards in both states, which is also what happens according to the no-surcharge rule above. The first-best outcome would seem to be achievable by picking the interchange fee so that the marginal cardholder is the same type as would be chosen by the central planner (that is, $b_c^* = b_s$). This occurs when

$$a = (r_c - F_c)/M + \pi_c + q(v-d) - c_m + b_m$$

However, substituting this interchange fee into the condition that merchants accept cards ($b_m \geq f_m(a) - q(v-d)$) implies $\pi_c + \pi_m + (r_c - F_c)/M < 0$ which violates assumption A6. Given assumption A6, merchants will not want to accept cards at the above interchange fee. The socially optimal interchange fee under the no-surcharge rule is then the highest one for which merchants accept cards; that is, $a^* = b_m - c_m - \pi_m + q(v-d)$. At a^* there will be less cardholders and less usage than the first-best outcome. For any other interchange fee for which merchants accept cards this will also be true. However, unlike the first-best world in which we derived the central planner's solution, a card association cannot in practice force merchants to accept cards.

Case (ii) $q \geq F_c/(M(v-d))$

In this case, the first-best solution calls for some low valuation cardholders ($b_c < b_u$) not to use their cards in state 2. However, under the no-surcharge rule, $b_c^* = r_c/M - a + c_c + \pi_c$, and all cardholders use their cards in both states. The result that merchants accept cards implies

$$a \leq q(v-d) + b_m - c_m - \pi_m$$

$$\leq q(v-d) + b_m - c_m + \pi_c + (r_c - F_c)/M \text{ from the assumption that } \pi_c + \pi_m + (r_c - F_c)/M \geq 0$$

$$< b_m - c_m + \pi_c + r_c/M + (v-d) - F_c/(qM) \text{ from the fact } q \geq F_c/(M(v-d)) \text{ and } q > 0$$

so that

$$b_c^* > r_c/M - (b_m-c_m+\pi_c+r_c/M+(v-d)-F_c/(qM)) + c_c + \pi_c = c_c+c_m-b_m+F_c/(qM)-(v-d)=b_s$$

This says that the proportion of consumers who hold cards is less than the first-best level; that is, there is under-subscription. Comparing the level of b_c^* with the first-best level of usage in state 2, it can be shown that when $q=F_c/(M(v-d))$ there is under-usage of cards in state 2 (and of course under-usage of cards in state 1). As q increases, b_c^* falls and for sufficiently high q there can be over-usage of cards in state 2 compared to the first-best outcome. If this over-usage is sufficiently great, it may even offset the under-usage in state 1. In general, the no-surcharge rule and positive interchange fees, allow a card network to compensate cardholders so that the extent of under-subscription and under-usage compared to the first-best outcome is attenuated.

In conclusion, imposing the no-surcharge rule most likely increases, and never decreases social surplus (regardless of the interchange rate set), in a world of monopolists. The no-surcharge rule prevents merchants expropriating the additional surplus that arises from the use of credit cards. If merchants are not constrained from such pricing, customers will not be prepared to pay anything to join the card network. Moreover, an interchange fee would not help this unfortunate situation, since merchants would simply raise the price that customers using cards face, in response to any increase in the interchange fee. In aggregate, merchants value a constraint that prevents them setting a surcharge, even though each merchant, taking the card network as given, would like to price discriminate.

4.2 Bertrand competition

In this section we take the opposite extreme to the previous model by assuming all merchants offer an identical product and compete over price according to Bertrand competition. Thus we assume each merchant produces one of the M goods that the consumers buy, but in contrast to before, that each of the goods is identical.[14]

(A) Surcharges allowed

In this case prices are driven to (net) costs, so that

$$p_{cash}=d$$

$$p_{card}=d+f_m(a)-b_m$$

are the equilibrium prices provided $f_m(a)-b_m \geq 0$

---

14   *As before, there can be more or less merchants than there are goods sold.*

At these prices merchants are indifferent about accepting cards or not, but in equilibrium at least one merchant will accept cards. To see this is an equilibrium, first note that the merchants earn no margin on each transaction. Consider the possibility no merchants accept cards. Then a merchant could offer to accept cards, setting a price for card usage above $d+f_m(a)-b_m$, and they would attract some cardholders to their store (for instance, all cardholders in state 1) and would earn a positive margin from doing so. Moreover, if any merchant offered a lower price than those above, they would make a loss. If any merchant increased their price above the levels above, they would receive no customers. Thus, merchants cannot do better by changing their prices.

If the interchange fee is set low, so that $f_m(a)-b_m < 0$, then merchants in competition would like to set the price of card transactions below that of cash. The prices above would still represent the equilibrium prices if there were not other constraints. However, given the legal tender restrictions on cash, merchants may not be able to set a higher price for cash than credit. In this case the best they can do is to accept cards and set a common price (p*) at which they earn zero profits. This price will be less than d, given that $f_m(a)-b_m < 0$. Despite this, a merchant cannot profit by excluding cardholders, since to cover their costs they will have to raise their price to d, but they would not attract any cash customers if they did so. In fact, they make a loss on cash customers and a profit on cardholders. They could increase profits by refusing cash and charging slightly less than p*, but since they are not allowed to do this by law, no firm can do better than setting a common price for cash and cards.

Returning to the equilibrium with $f_m(a)-b_m \geq 0$, total surplus depends on whether the marginal cardholder uses their card in both states or not. This depends on whether $q < r_c/(M(v-d))$ or not.

(i) $q < r_c/(M(v-d))$

For the marginal cardholder the benefits of joining the network are the same as not joining; that is,

$$[f_c(a)+b_c*+(v-d-f_m(a)+b_m)]M-r_c = (1-q)(v-d)M$$

which implies

$$b_c* = r_c/M + f_m(a)-b_m-q(v-d)-f_c(a)$$

Comparing this outcome to the first-best solution (1a) and (2a), the extent of under-subscription and under-usage is proportional to $(r_c-F_c)/M + \pi_c + \pi_m$ which from assumption A6 is non-negative. Total surplus is

$$TS = M\int_{b_c*}^{\infty}(b_c+b_m+v-d-c_c-c_m)h(b_c)\,db_c + M\int_{-\infty}^{b_c*}(v-d)(1-q)h(b_c)\,db_c - \int_{b_c*}^{\infty}F_c h(b_c)\,db_c \quad (12)$$

(ii) $q \geq r_c/(M(v-d))$

In this case the marginal cardholder only uses their card in state 2. Thus the marginal cardholder compares $[q(f_c(a)+b_c*+(v-d-f_m(a)+b_m))+(1-q)(v-d)]M-r_c$ with $(1-q)(v-d)M$ and is defined by

$$b_c* = r_c/(qM)-(v-d-(f_m(a)-b_m-f_c(a))$$

The type of consumer that is indifferent between using their card in state 2 and not using it is defined as

$$b_I = f_m(a)-b_m-f_c(a) \geq b_c*$$

Comparing these outcomes to the first-best solution (1b) and (2b), under-subscription is proportional to $(r_c-F_c)/(qM) + \pi_c + \pi_m$, while under-usage in state 2 is proportional to $\pi_c + \pi_m$. Using these definitions, total surplus is then

$$TS = M\int_{b_I}^{\infty}(b_c + b_m + v - d -c_c - c_m)h(b_c)\,db_c + M\int_{b_c*}^{b_I}q(b_c + b_m + v - d -c_c - c_m)h(b_c)\,db_c$$

$$+ M\int_{-\infty}^{b_I}(v-d)(1-q)h(b_c)\,db_c - \int_{b_c*}^{\infty}F_c h(b_c)\,db_c \tag{13}$$

(B) No-surcharges allowed

In this case each merchant must set a single price p*. If $f_m(a)-b_m < 0$, then the equilibrium for sur-charging is identical to case (A) above, since merchants do not want to surcharge in equilibrium even if they are allowed. When $f_m(a)-b_m \geq 0$, the equilibrium prices are also identical to above. Some firms will set their price at p_{cash} defined above and will not accept cards, while others will set their common price at p_{card} defined above and accept cards. To see this is an equilibrium note that the merchants earn zero margins on all transactions. Moreover, those merchants that accept cards, only attract customers who will pay by card - since $p_{card}>p_{cash}$. If any individual merchant increased their price, they would attract no customers, while if any merchant decreased their price they would make a loss.

Consider the possibility no merchants accept cards. Then a merchant could offer to accept cards, setting a price for card usage above $d+f_m(a)-b_m$, and they would attract some cardholders to their store (for instance, all cardholders in state 1) and would earn a positive margin from doing so. Thus, at least one merchant will accept cards and will set $p_{card} = d + f_m(a) - b_m$.

The equilibrium is identical in pricing and outcomes to the case where surcharging is allowed. Under perfect competition, merchants will "separate", either accepting cards or not. Those merchants that accept cards will set a price to cover the cost of doing so. Given this, they will only attract cardholders. The remaining merchants will set a lower price, but only accept cash customers. Because consumers face the same pricing as under surcharging, because there is always one merchant that accepts cards, and because all merchants sell an identical good, subscription, usage, and welfare are identical in the two cases.

96

4.3 Hotelling competition

In this section we see how our model of the card payment system applies under the Hotelling model of product differentiation, as used by Rochet and Tirole. Under this model merchants compete on price, but are differentiated so that even if one merchant sets a lower price they will not necessarily capture all the market. We use the standard Hotelling set-up of a "linear city", in which two merchants compete, but suppose there are M such separate cities. Equivalently, we can think of there being many separate markets or industries, and in each two differentiated merchants compete for customers. Even though there are only two firms in each sector, the "transportation cost" parameter t allows us to capture varying degrees of competition between the firms. In each sector we suppose there are a continuum of consumers with locations x distributed uniformly between 0 and 1. Their location x is independent of their value of b_c. Although the distribution of consumers is over both the b_c and x dimensions, the total measure of consumers is still one. A consumer located at x obtains a benefit from purchasing from merchant 1 of $t(1-x)$ and from purchasing from merchant 2 of tx, in addition to the benefits discussed in section 2. Each consumer still makes up to M transactions, depending on the extent to which they carry out transactions in state 1.

(A) Surcharges allowed

Under the Hotelling model, when surcharges are allowed, firms will set the price for credit and for cash just based on a markup of t over net-costs. This implies $p_{card}=d+f_m(a)-b_m+t$ and $p_{cash}=d+t$. Where cards are not needed to carry out the transaction, these prices leave the merchant indifferent between accepting cards and not.[15] Regardless of the form of payment, the merchant earns a margin of t. However, in state 1 where cards are needed, the merchant will strictly want to accept cards since they earn a positive margin, as opposed to nothing if they accept cash. Note, however, the merchants' decisions are independent of the interchange fee. Similarly, a consumer's decision to hold a credit card, or to use it, does not depend on the interchange fee. Cardholder receive benefits of $b_c+v-(d+f_m(a)-b_m+t)+f_c(a)$ if they use cards in either state 1 or 2 and face a cost of r_c per-card, both of which are independent of the interchange fee a. Because of this, the interchange fee cannot be used to influence consumers' or merchants' decisions to hold or use cards. Given all the costs of card usage are passed back to cardholders, cardholders face the full social costs of a card transaction but receive only some of the social benefits. An appropriately set interchange fee could help to reallocate some of the costs back to the merchant, but because under surcharging they are passed straight back to the cardholders this cannot happen. This suggests surcharging will lead to too few cardholders and too little usage.

---

15 *Because merchants are always willing to accept cards, symmetry implies each will share half of the cardholding customers and half of the cash customers at the equilibrium prices.*

To see this we first look at customers' usage decisions, and then examine their subscription decision to the network. For those customers who have a card, they will use this in state 1 if

$$b_c + v - (d + f_m(a) + t) + f_c(a) > 0$$

and in state 2 if

$$b_c + v - (d + f_m(a) + t) + f_c(a) > v - (d + t)$$

These conditions can be simplified to say customers will use a credit card in state 1 if

$$b_c + v - d > c_m + c_c + \pi_m + \pi_c + t$$

and in state 2 if

$$b_c > c_m + c_c + \pi_m + \pi_c$$

These conditions give rise to two possibilities. Either

(i)   those holding credit cards will use them all the time; or

(ii)   there will be some cardholders (with low b_c) that will use cash in state 2.

Which case we are in depends on the parameters of the model. It can be shown that case (i) holds if and only if

$$q < r_c/(M(v-d-t)) \tag{14}$$

In this case the type of cardholder that is just indifferent between joining the network or not, is defined by

$$b_c^* = c_c + c_m + \pi_c + \pi_m - q(v-d-t) + r_c/M - b_m \tag{15}$$

and all cardholders that join the network, use their cards in both states of the world. In comparing this to the first-best solution, we need to consider both the case that $q < F_c/(M(v-d))$ and the case that $q \geq F_c/(M(v-d))$. In the former case, $b_s = c_c + c_m - b_m + F_c/M - q(v-d)$ and so under-subscription is proportional to

$$\begin{aligned}
b_c^* - b_s \; &= c_c + c_m + \pi_c + \pi_m - q(v-d-t) + r_c/M - b_m - (c_c + c_m - b_m + F_c/M - q(v-d)) \\
&= \pi_c + \pi_m + (r_c - F_c)/M + qt \\
&> qt
\end{aligned}$$

from (A6). There is too little network membership, relative to the first-best solution. Moreover, in both state 1 and state 2 there is under-usage proportional to the level of under-subscription.

In the alternative case that $q \geq F_c/(M(v-d))$, $b_s = c_c + c_m - b_m + F_c/(qM) - (v-d)$ under-subscription is proportional to

$$b_c^* - b_s = c_c + c_m + \pi_c + \pi_m - q(v-d-t) + r_c/M - b_m - (c_c + c_m - b_m + F_c/(qM) - (v-d))$$
$$= \pi_c + \pi_m + (r_c/M - q(v-d-t)) + ((v-d) - F_c/(qM))$$
$$> \pi_c + \pi_m$$

since $q < r_c/(M(v-d-t))$ and $q \geq F_c/(M(v-d))$. Because cardholders always use their cards in state 1, under-subscription implies the equivalent amount of under-usage in state 1. Moreover, even in state 2, where the first-best solution requires that not all cardholders use their cards, we find that the outcome under surcharging implies insufficient usage of cards. The under-usage in state 2 is proportional to

$$b_c^* - b_u = c_c + c_m + \pi_c + \pi_m - q(v-d-t) + r_c/M - b_m - (c_c + c_m - b_m)$$
$$= \pi_c + \pi_m + (r_c/M - q(v-d-t))$$
$$> \pi_c + \pi_m$$

Thus, allowing surcharges leads to inefficiently low usage and inefficiently low membership under markup pricing. Total surplus for this case $(q < r_c/(M(v-d-t)))$ is

$$TS = M \int_{b_c^*}^{\infty} (b_c + b_m + v - d - c_c - c_m) h(b_c) \, db_c + M \int_{-\infty}^{b_c^*} (v-d)(1-q) h(b_c) \, db_c - \int_{b_c^*}^{\infty} F_c h(b_c) \, db_c \qquad (16)$$

where b_c^* is defined in equation (15). Since interchange fees do not affect consumers' choice of card membership under surcharging they do not enter this total surplus function.

In the second case that $q \geq r_c/(M(v-d-t))$, we assume q will also be greater than $F_c/(v-d)$. This implies the marginal cardholder has benefits of using cards of

$$b_c^* = c_c + c_m + \pi_c + \pi_m - (v-d-t) + r_c/(qM) - b_m \qquad (17)$$

Compared to the central planner's choice of b_s, this is again higher; the amount of under-subscription is proportional to:

$$b_c{}^* - b_s = c_c + c_m + \pi_c + \pi_m - (v-d-t) + r_c/(qM) - b_m - (c_c + c_m - b_m + F_c/(qM) - (v-d))$$
$$= \pi_c + \pi_m + (r_c - F_c)/(qM) + t$$
$$> t$$

As is always the case, the under-subscription in state 1 also corresponds to under-usage in state 1. We define the benefit of the cardholder who is indifferent between credit and cash in state 2 as b_1, and note that by construction $b_1 > b_c{}^*$. The marginal cardholder who uses credit in state 2 with surcharging is

$$b_1 = c_c + c_m + \pi_c + \pi_m - b_m \tag{18}$$

Comparing this to the first-best solution, the under-usage of credit in state 2 is proportional to

$$b_1 - b_u = c_c + c_m + \pi_c + \pi_m - b_m - (c_c + c_m - b_m)$$
$$= \pi_c + \pi_m$$
$$> 0$$

Thus in all cases considered, there will be too few subscribers and too few credit card transactions compared to the first-best solution. Cardholders do not take into account the revenue to banks and merchants, which their subscription generates, and thus only those with sufficiently high private benefits from using cards will join. To get the more marginal cardholders to join requires a subsidy, but since merchants pass on any interchange fee to cardholders through the surcharge, no such subsidy is possible. The ability to use the interchange fees to maximize the value of the network is destroyed. Merchants would be better off with more cardholders, but each individual merchant takes the number of cardholders as a given, and so does not lower the surcharge on card usage to promote greater network membership. Total surplus for this case ($q \geq r_c/(M(v-d-t))$) is

$$TS = M \int_{b_1}^{\infty} (b_c + b_m + v - d - c_c - c_m) h(b_c) \, db_c + M \int_{b_c^*}^{b_1} q(b_c + b_m + v - d - c_c - c_m) h(b_c) \, db_c$$
$$+ M \int_{-\infty}^{b_1} (v - d)(1 - q) h(b_c) \, db_c - \int_{b_c^*}^{\infty} F_c h(b_c) \, db_c \tag{19}$$

where $b_c{}^*$ and b_1 are defined in equations (17) and (18) respectively.

(B) No-surcharge rule

In this case, merchants must set a common price for cash and credit, or not accept cards at all. We denote this common price p*. There are two separate cases to consider. If $b_c*+f_c(a) \geq 0$ then even the lowest value cardholder will always use their cards when facing a common price for cash or card-usage. Alternatively, if $b_c*+f_c(a)<0$ then facing a common price for cash and credit, the marginal cardholder will only use their card when it is required to make the purchase; that is, in state 1.

Case (i) $b_c*+f_c(a) \geq 0$

Consider each city or sector. Suppose merchant 1 sets a common price of p_1 and merchant 2 sets a common price of p_2 and assume initially that both firms accept cards. A cardholder of type b_c located at x get a payoff of $b_c+v-p_1+f_c(a)+t(1-x)$ if they purchase with their cards from merchant 1 and a payoff of $b_c+v-p_2+f_c(a)+tx$ if they purchase with their cards from merchant 2. Thus the share of cardholders who purchase from merchant 1 is

$$s_{cards} = 1/2 + (p_2-p_1)/(2t)$$

Similarly, for non-card holding customers, the benefit of purchasing from merchant 1 in state 2 is $(v-p_1)+t(1-x)$ and the benefit of purchasing from merchant 2 in state 2 is $(v-p_2)+tx$. The share of cash paying customers that merchant 1 captures is also

$$s_{cash} = 1/2 + (p_2-p_1)/(2t)$$

In this case the expected profit of merchant 1 is

$$\text{Profit}_1 = s_{cards}(1-H(b_c*))(b_m+p_1-d-f_m(a))+s_{cash}H(b_c*)(1-q)(p_1-d) \tag{20}$$

Taking b_c* as given, merchant 1 will set p_1 to maximise its expected profits. Differentiating equation (20) with respect to p_1 and setting the resulting derivative equal to zero gives the first order condition for merchant 1. Solving both merchant's first order conditions simultaneously gives equilibrium prices of p*, where

$$p*=p_1=p_2= d+t+(f_m(a)-b_m)\ (1-H(b_c*))\ /\ (1-qH(b_c*)), \tag{21}$$

Substituting (21) back into (20) and simplifying the resulting expression implies

$$\text{Profit}_1* = (t/2)(1- qH(b_c*)) \tag{22}$$

and likewise for merchant 2.[16] This is an equilibrium provided both merchants want to accept cards. The following proposition shows they will want to accept cards, provided a is not too large and q is not too large.

Proposition 2

Assuming the parameters q and a are not too high, merchants will accept cards. In this case prices will be determined by the markup formula described in equation (21).

Proof

Suppose firm 1 deviates and does not accept cards. Suppose as a result it sets a price for cash purchases of p_1 and firm 2 sets a new price of p_2. All cardholders will choose merchant 2 in state 1. In state 2 a cardholder of type b_c located at x gets benefits of $v-p_1+t(1-x)$ from merchant 1 and benefits of $b_c+v-p_2+f_c(a)+tx$ from merchant 2. Thus the share of cardholders of type b_c who purchase from merchant 1 in state 2 is

$$s(b_c) = 1/2 + [p_2-p_1 - (b_c+f_c(a)] / (2t)$$

For consumers who do not hold cards, they will not be able to purchase anything in state 1. In state 2, a cash customer of type b_c located at x gets benefits of $v-p_1+t(1-x)$ from purchasing from merchant 1 and benefits of $v-p_2+tx$ from purchasing from merchant 2, so that

$$s(b_c) = (1/2) + [p_2-p_1] / (2t)$$

Thus the expected profit merchant 1 will get if it stops accepting cards is

$$\text{Profit}_1 = (1-q)(p_1-d)[1/2 + (p_2-p_1-I(b_c^*))/(2t)] \tag{23}$$

where $I(b_c^*) = \int_{b_c^*}^{\infty}(b_c + f_c(a))h(b_c)db_c > 0$. Choosing p_1 to maximise expected profits in (23) implies

$$p_1^* = 0.5(t+d+p_2 - I(b_c^*)) \tag{24}$$

Firm 2 faces all cardholders in state 1, as well as the market share $1-s(b_c)$ of customers of each type b_c in state 2. Thus the expected profit merchant 2 gets is

---

16 *Interestingly, equilibrium merchant profit can increase as the interchange fee is increased. To the extent a higher interchange fee enables greater cardholder rebates and thus more card membership, this will increase merchant profits. With more card members, merchants are less likely to miss sales when consumers have insufficient funds. A higher merchant discount does not affect their profits in equilibrium, as these are passed through into prices.*

102

$$Profit_2 = q(p_2-d+b_m-f_m(a))(1-H(b_c^*)) + (1-q)[(p_2-d)H(b_c^*))(1/2 + (p_1-p_2)/(2t))$$

$$+(p_2-d+b_m-f_m(a))(I(b_c^*)/(2t) + (1-H(b_c^*))(1/2 + (p_1-p_2)/(2t)))] \tag{25}$$

Choosing p_2 to maximise expected profits in (25) implies

$$p_2^* = 0.5(t+d+p_1+I(b_c^*)+(1-H(b_c^*))(f_m(a)-b_m)) + tq(1-H(b_c^*))/(1-q) \tag{26}$$

Solving (24) and (26) simultaneously implies

$$p_1^* = [t + d + (1/3)(f_m(a)-b_m)(1-H(b_c^*)) - (1/3)I(b_c^*) + (2/3)(tq(1-H(b_c^*))/(1-q)] \tag{27}$$

and $\quad p_2^* = [t + d + (2/3)(f_m(a)-b_m)(1-H(b_c^*)) + (1/3)I(b_c^*) + (4/3)(tq(1-H(b_c^*))/(1-q)] \tag{28}$

Substituting (27) and (28) into (23) and simplifying implies

$$Profit_1 = (t/2)(1-q)[1-y(b_c^*)/(3t)]^2 \tag{29}$$

where $y(b_c^*) = (1-H(b_c^*)) [I(b_c^*) / (1-H(b_c^*)) - (f_m(a)-b_m) - 2tq/(1-q)]$. Provided profits in (29) are no more than profits in (22), merchant 1 will want to accept cards. That is, provided

$$(1-q)[1-y(b_c^*)/(3t)]^2 \; < \; (1- qH(b_c^*)).$$

A sufficient condition for this to be true is if $y(b_c^*)$ is non-negative.

Turning to consumers' subscription decision, a consumer gets benefits of $[b_c+v-p^*+ f_c(a)]M-r_c$ from belonging to the card payment network, and $(1-q)[v-p^*]M$ if they do not belong. Thus b_c^*, the level of b_c for which customers are indifferent between belonging and not, solves the equation

$$b_c^*=r_c/M - q(v-p^*) - f_c(a) \tag{30}$$

Because each individual merchant takes the number of cardholders as given when it considers deviating and not accepting cards, the determination of the marginal cardholder is determined under the posited equilibrium in which both firms accept cards. Given (21) and (30) it can be shown that provided q is not too high, then for sufficiently low a, $y(b_c^*)$ is non-negative. To get this result we make use of the fact that b_c^* is decreasing in a and so $I(b_c^*) / (1-H(b_c^*))$ is decreasing in a.[17] If q is too high, then the last term in $y(b_c^*)$ will be large and so $y(b_c^*)$ can be negative. In this case, there may be no equilibrium where both merchants accept cards.[18]

QED

---

[17] This is the same condition Rochet and Tirole use.

[18] We leave the more difficult analysis of what the equilibrium is with a higher q to future research.

Proposition 2 extends the first part of proposition 1 in Rochet and Tirole (1999) to the case that q can be positive, although not too large.[19]

Substituting (21) into (30), we get that the marginal cardholder's type can be found by solving the equation

$$b_c^* = r_c/M - q(v-d-t) - a(1-q) / (1-qH(b_c^*))$$
$$+ [q(1-H(b_c^*)) / (1-qH(b_c^*))](c_m+\pi_m-b_m) + c_c + \pi_c \qquad (31)$$

Social surplus is found by integrating the social surplus generated across different cardholders and across the two states. This is:

$$TS = M\int_{b_c^*}^{\infty}(b_c + b_m + v - d - c_c - c_m)h(b_c)\,db_c + M\int_{-\infty}^{b_c^*}(1-q)(v-d)h(b_c)\,db_c - \int_{b_c^*}^{\infty}F_c h(b_c)\,db_c \qquad (32)$$

Differentiating (32) with respect to b_c^* we can find the level of subscription which is socially optimal under the no-surcharge rule. This socially optimal level of b_c^* (denoted B*) satisfies the first-order condition

$$B^* + b_m + q(v-d) = c_c + c_m + F_c/M \qquad (33)$$

Thus, the socially optimal interchange fee is such that the social benefit created by the marginal cardholder making transactions with cards in both states just equals the cost of such a transaction. Equating (31) and (33) implies

$$r_c/M + qt - a(1-q) / (1-qH(B^*))$$
$$+ [q(1-H(B^*)) / (1-qH(B^*))](c_m+\pi_m-b_m) + \pi_c = c_m + F_c/M - b_m \qquad (34)$$

The level of the interchange fee *a* which solves (34) is the socially optimal interchange fee. That is,

$$a^* = b_m - c_m + [(1-qH(B^*))/(1-q)][(r_c-F_c)/M + qt] + [(1-qH(B^*))\pi_c + q(1-H(B^*))\pi_m]/(1-q) \qquad (35)$$

Provided $q \leq F_c/(M(v-d))$, the marginal cardholder implied by this interchange fee is the same as the first-best solution; that is, equations (1a) and (33) are the same. Because all cardholders use their cards in both states, this corresponds to the first-best level of membership and usage of the network.

---

19 In the benchmark version of the calibrated model (section 6) we show that not too large means q<0.18.

104

For $F_c/(M(v-d)) < q$, the benefit b_c received by the marginal cardholder is higher in the outcome above (B^*), relative to the first-best solution (b_s). To show this result compare equations (1b) and (33) for q between $F_c/(M(v-d))$ and 1. There is under-subscription (and therefore, under-usage in state 2). However, according to the first-best solution, if $q \geq F_c/(M(v-d))$, the marginal cardholder should not use their card in state 2. This contrasts with above, implying there will be over-usage in state 2. It follows that for any higher level of interchange fee, there will be less under-subscription (and under-usage in state 1), but more over-usage in state 2.

The no surcharge rule with appropriately set interchange fees implies either the first best level of card membership and usage - the case $q \leq F_c/(M(v-d))$; or where this cannot be achieved, a tradeoff between membership and usage. This is in strong contrast to the results from relaxing the no surcharge rule, which show that in this case there is generally too little network membership and too little network usage.

The socially optimal interchange fee is clearly a complicated function of almost all of the parameters of the model. This reflects the role the interchange fee plays in balancing the costs and benefits that cardholders, cash customers and merchants face so as to get the right level of subscription and usage from a social perspective. One result which is clearly implied by the model is that the optimal interchange fee should not be set at the "technological" costs incurred by the issuing bank. This follows from the fact c_c does not enter *directly* into the optimal interchange fee equation (35). The intuition for this is clear. A higher c_c increases the cost of offering card services, and should optimally lead to lower subscription and usage, other things equal. This is clear from equation (33). By holding the interchange fee constant, as c_c increases, the cardholder rebate will fall, as issuing banks have less interchange left to pass back to cardholders, and so subscription and usage will indeed fall.

Taking into account the reduction in subscription means that the fact that cardholders do not fully internalise the value their transactions create in state 1 causes a smaller welfare loss. Thus as c_c increases the socially optimal interchange fee actually decreases. Formally, this follows from the fact

$$\frac{da^*}{dc_c} = \frac{-q}{1-q}(\frac{r_c - F_c}{M} + \pi_c + \pi_m + qt)h(B^*) < 0$$

The result highlights the complexity of calculating the socially optimal interchange fee, even in this simplified setting. The results also show why the optimal interchange fee is not equal to the technological cost of issuing.

T.1

Apart from the complexity in working out exactly what this socially optimal interchange fee is there are some qualifications to note. If the interchange fee which maximises total surplus is too high, then proposition 2 suggests that merchants may not want to accept cards. This not only puts a natural constraint on the interchange fee member associations will want, but also on the level that a social planner can choose without causing merchants to drop out. If the interchange fee is lowered to respect this constraint, then it may be that not all cardholders will use their cards in state 2. In this case there will be less over-usage in state 2, compared to that described above. We look at this case now.

Case (ii) $b_c^* + f_c(a) < 0$

In this case some low value consumers (those with b_c close to b_c^*) will not want to use credit in state 2. Rather, facing a common price p^*, they will prefer to use cash in state 2, while still using their cards in state 1. In particular, those cardholders of types $b_c < b_1$ will use cash in state 2, where

$$b_1 = -f_c(a)$$

We characterise the equilibrium where both merchants accept cards in competition. In this case, regardless of the type of consumer b_c, the share of consumers on the unit interval which will choose firm 1 is simply

$$s = 1/2 + (p_2 - p_1)/(2t)$$

The expected profit of merchant 1 is

$$\text{Profit}_1 = s[(1-q)H(b_1)(p_1-d) + q(1-H(b_c^*))(b_m+p_1-d-f_m(a)) + (1-q)(1-H(b_1))(b_m+p_1-d-f_m(a))] \quad (36)$$

Taking b_c^* as given, merchant 1 will set p_1 to maximise its expected profits. Differentiating equation (36) with respect to p_1 and setting the resulting derivative equal to zero gives the first order condition for merchant 1. Solving both merchant's first-order conditions simultaneously gives equilibrium prices of p^*, where

$$p^* = p_1 = p_2 = d + t + (f_m(a) - b_m)[1 - H(b_1) + q(H(b_1) - H(b_c^*))] / [1 - qH(b_c^*)] \quad (37)$$

Substituting (37) back into (36) and after considerable simplification we find the expression for profits is the same as in (22); that is,

$$\text{Profit}_1^* = (t/2)(1 - qH(b_c^*)) \quad (38)$$

and likewise for merchant 2. This represents an equilibrium provided both merchants want to accept cards. Like proposition 2, the following proposition shows merchants will want to accept cards, provided q and a are not too high.

Proposition 3

Assuming the parameters q and a are not too high, merchants will accept cards. In this case prices will be determined by the markup formula described in equation (37).

Proof

Suppose firm 1 deviates and does not accept cards. Suppose as a result it sets a price for cash purchases of p_1 and firm 2 sets a new price of p_2. All cardholders will choose merchant 2 in state 1. In state 2 a cardholder of type $b_c \geq b_I$ located at x gets a payoff of $v-p_1+t(1-x)$ from merchant 1 and a payoff of $b_c+v-p_2+f_c(a)+tx$ from merchant 2. Thus the share of cardholders of this type who purchase from merchant 1 in state 2 is

$$s(b_c) = 1/2 + [p_2-p_1 - (b_c+f_c(a)]/(2t) \qquad \text{for } b_c \geq b_I \qquad (39)$$

For consumers with $b_c < b_I$ who hold cards, they will compare (in state 2) purchasing with cash from merchant 1 and with cash from merchant 2, as will those without cards. Thus, in state 2, a consumer of type $b_c < b_I$ located at x gets benefits of $v-p_1+t(1-x)$ from purchasing from merchant 1 and benefits of $v-p_2+tx$ from purchasing from merchant 2, so that

$$s(b_c) = (1/2) + [p_2-p_1] / (2t) \qquad \text{for } b_c < b_I \qquad (40)$$

The expected profit merchant 1 will get if it stops accepting cards is thus

$$\text{Profit}_1 = (1-q)(p_1-d)[1/2 + (p_2-p_1- I(b_I))/(2t)] \qquad (41)$$

where $I(b_I) = \int_{b_I}^{\infty} (b_c + f_c(a))h(b_c)db_c > 0$. This is analogous to the profit expression (23).

Choosing p_1 to maximise expected profits in (41) implies

$$p_1 = 0.5[t + d + p_2 - I(b_I)] \qquad (42)$$

Firm 2 faces all cardholders in state 1, as well as the market shares $1-s(b_c)$ for each type of customer b_c. Thus, the expected profit merchant 2 gets is

$$\text{Profit}_2 = q(p_2-d+b_m-f_m(a))(1-H(b_c^*)) + (1-q)[\ (p_2-d)H(b_1))(1/2 + (p_1-p_2)/(2t))$$
$$+(p_2-d+b_m-f_m(a))(\ I(b_1)/(2t) + (1-H(b_1))(1/2 + (p_1-p_2)/(2t)))\ ] \tag{43}$$

Choosing p_2 to maximise expected profits in (43) implies

$$p_2^* = 0.5(t+d+p_1+I(b_1)+(1-H(b_1))(f_m(a)-b_m)) + tq(1-H(b_c^*))/(1-q) \tag{44}$$

Solving (42) and (44) simultaneously implies

$$p_1^* = [t + d + (1/3)(f_m(a)-b_m)(1-H(b_1)) - (1/3)I(b_1) + (2/3)(tq(1-H(b_c^*))/(1-q)] \tag{45}$$

and

$$p_2^* = [t + d + (2/3)(f_m(a)-b_m)(1-H(b_1)) + (1/3)I(b_1) + (4/3)(tq(1-H(b_c^*))/(1-q)] \tag{46}$$

Substituting (45) and (46) into (41) and simplifying implies

$$\text{Profit}_1 = (t/2)(1-q)[1-y(b_1)/(3t)]^2 \tag{47}$$

where

$$y(b_1) = (1-H(b_1))\ [\ I(b_1) / (1-H(b_1)) - (f_m(a)-b_m) - 2tq(1-H(b_c^*)) / ((1-q)(1-H(b_1)))].$$

Provided profits in (47) are no more than profits in (38), merchant 1 will want to accept cards. That is, provided

$$(1-q)[1-y(b_1)/(3t)]^2\ <\ (1- qH(b_1)) \tag{48}$$

A sufficient condition for (48) to hold is if $y(b_1)$ is non-negative.

Turning to consumers' subscription decision, the marginal consumer gets a payoff of $[q(b_c^*+v-p^*+ f_c(a))+(1-q)(v-p^*)]M-r_c$ from belonging to the card payment network, and $(1-q)[v-p^*]M$ if they do not belong. Thus b_c^*, the level of b_c for which customers are indifferent between belonging and not, solves the equation

$$b_c^*=r_c/(qM) - (v-p^*) - f_c(a) \tag{49}$$

Because each individual merchant takes the number of cardholders as given when it considers deviating and not accepting cards, the determination of the marginal cardholder is determined under the posited equilibrium in which both firms accept cards. Given (37) and (49) it can be shown that provided q is not too high, then for sufficiently low a, $y(b_I)$ is non-negative.

QED

Proposition 3 delivers a very similar result to proposition 2, implying there is an equilibrium where both merchants will accept cards provided we do not get too far from the Rochet and Tirole case in which q=0. Substituting (37) into (49), we get that the marginal cardholder's type can be found by solving the equation

$$b_c^* = r_c/(qM) - (v-d-t) - a(1-q)H(b_I) / (1-qH(b_c^*)) + c_c + \pi_c$$
$$+ (c_m + \pi_m - b_m) [1-H(b_I)+q(H(b_I)-H(b_c^*))] / (1-qH(b_c^*)) \tag{50}$$

Social surplus is found by integrating the social surplus generated across different cardholders and across the two states. This is:

$$TS = M\int_{b_I}^{\infty}(b_c + b_m + v - d - c_c - c_m)h(b_c)\,db_c + M\int_{b_c^*}^{b_I}q(b_c + b_m + v - d - c_c - c_m)h(b_c)\,db_c$$
$$+ M\int_{-\infty}^{b_I}(1-q)(v-d)h(b_c)\,db_c - \int_{b_c^*}^{\infty}F_c h(b_c)\,db_c \tag{51}$$

Differentiating (51) with respect to *a* gives the socially optimal interchange fee under the no-surcharge rule. Because

$$\frac{dTS}{da} = \frac{\partial TS}{db_c^*}\frac{db_c^*}{da} + \frac{\partial TS}{db_I}\frac{db_I}{da}, \tag{52}$$

the socially optimal interchange fee will generally involve a trade-off between optimal membership (which would involve the first term in (52) equaling zero) and optimal usage in state 2 (which would involve the second term in (52) equaling zero). Generally one might expect that at the social optimum, there is too little subscription to the network, but too much usage in state 2 compared to the first-best solution. An increase in the interchange fee will lower under-subscription (and therefore raise total surplus) but increase over-usage in state 2 (and therefore lower total surplus).

To obtain specific results we calibrate the model above. In doing so we use the conditions derived above to check whether parameter values sustain the conjectured equilibrium. In order to obtain numerical results we need to know which interchange fee a card association would pick for any given set of parameter values. In the next section we derive a formula to calculate the interchange fee that roughly characterises the approach taken by VISA in Australia.

5. VISA's setting of interchange in Australia

For the purposes of modeling the level of interchange fees, the approach adopted by VISA Australia can roughly be characterized as the balancing of profitability between acquiring and issuing banks, where profits are scaled by revenue. To see what this implies using our model, as in previous sections let $b_c{}^*$ denote the type of consumer with the lowest b_c that belongs to the card network and denote by b_1 the type of consumer with the lowest b_c that uses their card in state 2. Note that for some parameter values $b_1 > b_c{}^*$, while for other sets of parameter values $b_1 = b_c{}^*$. Let $T_1 = [1-H(b_c{}^*)]M$ and let $T_2 = [q(1-H(b_c{}^*)) + (1-q)(1-H(b_1))]M$. T_1 is a measure of the number of transactions that cardholders make if they always use their cards, while T_2 is a measure of the number of transactions that cardholders will actually make with their cards. (For some parameter values, $T_1 = T_2$). The revenue and costs that issuing banks face are:

$$\text{Rev}_I = T_1(r_c/M) + T_2 a \qquad \text{Cost}_I = T_1(F_c/M) + T_2(c_c + f_c(a)),$$

while the acquiring banks face revenue and costs of

$$\text{Rev}_A = T_2(f_m(a) - a) \quad \text{Cost}_A = T_2 c_m$$

When VISA's method of setting the interchange is interpreted in terms of the model, the centralised interchange fee a will solve

$$(\text{Rev}_I - \text{Cost}_I)/\text{Rev}_I = (\text{Rev}_A - \text{Cost}_A)/\text{Rev}_A$$

This implies

$$a_{\text{VISA}} = \pi_c - (T_1 F_c)/(T_2 M) + (c_m/\pi_m)(\pi_c + (T_1/T_2)((r_c - F_c)/M)) \qquad (53)$$

Note if π_c is increasing in a and π_m is decreasing in a, this expression does not represent an explicit solution for a_{VISA}, but rather an implicit solution from which the interchange fee can be determined. We will use this formula below to compute the interchange fee that VISA sets given particular parameters of the model.

6. A Numerical Example

To illustrate the welfare effects predicted by the model we use the following numerical values: $v=100$, $d=80$, $t=10$, $M=100$, $\underline{r}=25$, $F_c=25$, $\pi_c=0.5$, $\pi_m=0.5$, $c_c=0.25$ and $c_m=0.55$. Thus, we assume each transaction (absent payment costs) is worth \$100 to the consumer, has a cost of \$80 to the merchant, and that in the Hotelling version of the model, the profit margin set by the merchant is \$10.

To realise a transaction requires a payment mechanism and this is not without cost. We assume that the technological cost to the issuing and the acquiring banks is 25 cents per-transaction. We assume that c_m is 30 cents higher than c_c since we suppose it includes the system cost that is paid to VISA through the interchange fee. We assume the issuing banks face a per-cardholder cost of \$25, which it passes on directly to its cardholders (so $r_c=\$25$). In addition, we assume both issuing and acquiring banks earn margins over their incremental costs of 50 cents per-transaction, which help cover fixed and common costs and possibly provide a return on capital.

We assume the cardholders' types (b_c) are distributed according to the Normal distribution, with a mean value (μ) of \$1 per-transaction and a standard deviation (σ) of \$1. This implies over two-thirds of consumers have valuations b_c between 0 and \$2 per-transaction. We also assume the merchants' transactional benefits from being able to accept a card (b_m) is equal to \$1. Finally, we assume that each consumer will potentially want to make 100 transactions in a year (although the actual number of transactions depends on q and whether the particular consumer holds a card or not). Noting that the measure of consumers in the model is always set equal to one, all welfare results are measured per-consumer (per-capita). These parameters are chosen so that, when cardholders always use their cards in both states, the interchange fee that would be paid by acquiring banks is \$1.10 per-transaction.[20] Given the mean values adopted here for π_m and c_m, the merchant discount implied by these parameters is \$1.85.

---

20   *Note that the difference between the interchange fee paid by an acquiring bank and the interchange fee received by the corresponding issuing bank, which is determined by equation (53), is the VISA system charge.*

One of the most important parameters of the model is q, which measures the fraction of times that a customer requires their card to complete transactions. It reflects the probability a consumer is constrained by funds at the particular time they want to make a purchase. Because in practice there are other (albeit imperfect) ways consumers can adjust to a lack of funds other than using credit, q is lower than a literal interpretation of the model implies.[21] A plausible value of q is likely to be between 0.05 and 0.1. Figure 1 shows the welfare impact of relaxing the no-surcharge rule for a range of values of q between 0 and 0.15.

The figure presents the welfare losses arising from relaxing the no-surcharge rule for each of the different models of merchant's behaviour that we considered in the paper. In each case we assume that under the no-surcharge rule, the interchange fee that would be paid by acquiring banks is $1.10 per-transaction (note given the size of each transaction is approximately $90, this corresponds to a percentage rate of 1.22%).

The solid black line shows the welfare losses that arise from allowing surcharging when all merchants are monopolists. Because monopolistic merchants leave no surplus to cardholders, they destroy the payment network, and so a fraction (q) of all transactions are lost as a result of surcharging. As section 4.1 outlines, regardless of the parameter values there is no equilibrium involving cardholders when there is surcharging. Under the no-surcharge rule, there may also be no equilibrium involving cardholders for certain parameter values. Our numerical example shows that given the other parameter values assumed, this is the case for q<0.04. For these parameter values welfare is unaffected by whether there is surcharging or not. For higher values of q, welfare losses are between $200 and $300 per-person per-year. These losses increase with q, as the value of the VISA network that is put at stake by surcharging also increases with q.

21 Delaying a purchase, making smaller purchases, or arranging alternative (but more costly) forms of credit are three alternatives. All these imply some loss in value for consumers and firms.
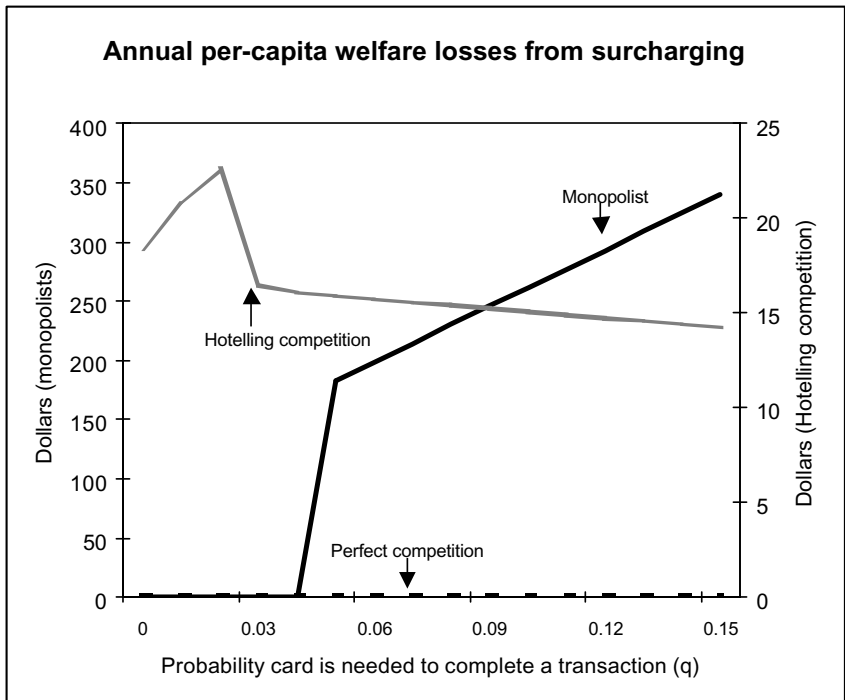
112

T.1

Annual per-capita welfare losses from surcharging

Figure 1

Under Hotelling competition, whereby merchants set prices as markups over costs, surcharging leads to smaller but still significant welfare consequences. This case is represented by the grey line in figure 1, which first increases in q and then decreases in q. Despite the non-monotonic behaviour of the welfare effects, welfare is always lower under surcharging.[22] For plausible values of q, welfare losses are at least $15 per-person per-year. For the Australian economy, assuming 15 million consumers and a discount rate of 5%, this implies a present discounted welfare loss of at least 4.5 billion dollars.

---

22 *The welfare losses are first increasing in q because a higher q means more value is lost when consumers do not hold cards as a result of surcharging. As q increases beyond a certain point this effect is offset by an opposing effect. With high q more consumers will start to hold cards despite surcharging since the value to being able to use the card in state 1 becomes sufficiently great. Thus the size of any welfare effects arising from allowing surcharging start to decline.*

The third case, of perfectly competitive merchants, is represented by the dashed line which coincides with the horizontal axis. In this case welfare remains the same regardless of the surcharging policy.

For these numerical calculations, the mean (μ) and standard deviation (σ) for the distribution of the cardholder benefits, as well as the merchants' transactional benefits from being able to accept a card (b_m), were all set to one for sake of any better assumption. Given the arbitrary nature of this assumption, together with the arbitrary degree of product differentiation ($t=10$) in the Hotelling model, it is important to consider how sensitive are the results to these assumptions. Table 1 below presents annual per-capita welfare losses from surcharging under some alternative scenarios around the benchmark case.

Table 1: Welfare losses arising from surcharging for the Hotelling model

	q=0	q=0.02	q=0.04	q=0.06	q=0.08	q=0.1
Benchmark	18.39	22.76	16.08	15.74	15.41	15.07
μ=0.5	18.55	25.44	18.27	17.89	17.51	N/A
μ=1.5	14.37	15.92	11.15	10.92	10.68	10.45
b_m=0.5	17.04	27.39	17.12	16.75	16.40	16.04
b_m=1.5	9.52	10.13	7.61	7.45	7.30	7.14
σ=0.5	31.78	32.07	22.07	21.61	21.15	20.69
σ=1.5	12.66	16.31	11.70	11.40	11.16	10.91
t=5	18.39	16.43	16.08	15.74	15.41	15.07
t=15	18.39	26.55	33.52	18.23	16.01	15.24

Note: Welfare losses are measured in dollars per-capita, per-year. The benchmark case is $\mu=1$, $b_m=1$, $\sigma=1$, $t=10$.

With the exception of the case with $\mu=0.5$ and $q=0.1$ the conditions in propositions 2 and 3 apply and the parameter values are such that it is still an equilibrium for merchants to accept cards. The table shows that the qualitative results are not affected by varying the benefits arising to cardholders and merchants from using cards. Quantitatively, the results show that the welfare losses from dropping the no-surcharge rule are greater when the transactional benefits of using cards are smaller, and when there is less variation in these benefits across consumers. These comparative static results reflect how the distribution of consumers, over which the private benefits under the no-surcharge rule are more closely aligned with social benefits than under surcharging, changes as a result of changes in one of the parameters. To the extent policy makers believe the transactional benefits of using credit cards are smaller for cardholders and merchants than our benchmark case, these results suggest if anything the welfare losses from allowing surcharging could be larger than those reported here.

The model also allows us to examine different approaches to setting the interchange fee. If instead of using VISA's rule for setting the interchange fee, the socially optimal interchange fee is used,[23] per-capita welfare could be increased by a modest 23 cents when $q=0.05$ and 22 cents when $q=0.1$. Alternatively, suppose the interchange fee was set at zero. Compared to VISA's rule this would lead to a reduction in per-capita welfare of 14.11 dollars when $q=0.05$ and 12.77 dollars when $q=0.1$. Thus, our model suggests that regardless of whether merchants will surcharge in the absence of the no-surcharge rule, a policy of regulating the interchange fee to zero and removing the no-surcharge rule will reduce welfare significantly. When firms do surcharge in the absence of the no-surcharge rule, then the welfare losses are best estimated at between \$15 and \$16 per-person per-annum, reflecting the reduction in card membership as a result of surcharging.[24] When firms do not surcharge, even when they are able to, the welfare losses are best estimated at between \$13 and \$14 per-person per-annum as a result of setting the interchange fee to zero.

---

23   *For the benchmark model above this corresponds to an interchange fee received by the issuing banks of 0.96 for both q=0.05 and q=0.1.*

24   *Regulating the interchange to zero does not affect the outcome under surcharging since as we showed the interchange fee no longer affects consumers' decision to hold cards or not when there is surcharging.*

T.1

7. Conclusion

In this paper we have provided a formal model of a card payment system and analysed the welfare effects of dropping the no-surcharge rule from this system as well as some properties of the socially optimal interchange fee. The model highlighted a fundamental problem with allowing merchants to surcharge card transactions – merchants with market power can use surcharging to expropriate the value of credit card transactions that cardholders would otherwise enjoy. This leads to too little membership of the card network and not enough usage of cards for transactions. Although merchants value having cardholders who will be able to make purchases which they otherwise might not make, each individual merchant takes as given the number of cardholders, and so ignores the effect that its own price discrimination has on a consumer's decision to join the card network. Likewise, each potential cardholder ignores the benefit they will provide to merchants if they carry the card, despite the fact they value more merchants. To maximise the benefits provided to cardholders and merchants from the network, an interchange fee can be set which ensures benefits are appropriately shared between cardholders and merchants. However, when merchants surcharge for cards, any interchange fee will be passed directly back to cardholders, and so no such transfer is possible. Thus eliminating the no-surcharge rule eliminates the ability to maximise the value of the network by promoting card membership, and generally leads to too few cardholders and too little use of cards compared to the level desired by a social planner.

Imposing the no-surcharge rule and setting a sufficiently high interchange fee either leads to the first-best outcome, or provides a trade-off between network membership and card usage, depending on the exact parameters of the model. In the latter case, at the optimal interchange fee there will still be too few cardholders compared to the first-best solution, but some cardholders will use their cards even when cash would be more efficient. Thus a trade-off exists between the membership and usage of the card system, compared to the levels that would be chosen by a social planner who knows exactly when cardholders do and do not need to use their card to complete a transaction.

We calibrate the model to illustrate the magnitude of welfare losses that arise if the no-surcharge rule is eliminated. In each case we look at, dropping the no-surcharge rule reduces social welfare.[25] Depending on the likelihood that consumers require a card to complete any given transaction and the extent merchants expropriate surplus from cardholders, the welfare losses we calculate from dropping the rule range from a few dollars per-person per year to several hundred dollars per-person per-year. For our benchmark case using the Hotelling model our best estimate of the welfare losses is that they are at least 15 dollars per-person per-year, or for the Australian economy, at least 4.5 billion dollars in present value terms.

---

25 *The only exception is when there is perfect competition between merchants, in which case welfare is unchanged.*

The interchange fee implied by VISA Australia's rule is quite similar to the socially optimal fee calculated from the Hotelling model. The welfare gains that can be achieved by adopting the socially optimal interchange fee rather than the existing interchange fee are small (less than 25 cents per-person per-year). This is in contrast to the social losses that arise if the interchange fee is dropped altogether (these are between $13 and $14 per-person per-year).

These welfare calculations provide a useful first indication of the scale of losses that might arise from dropping the no-surcharge rule or eliminating the beneficial role played by the interchange fee. A number of additional effects will likely compound these welfare losses in practice. While analytically difficult, one could consider a range of merchant types when modeling merchants' participation in a card network at the same time as cardholders' participation. We suspect that doing so would reveal that to the extent the interchange fee and the no-surcharge rule help balance the gains to using cards between merchants and cardholders, they also serve to maximise the value of the card network to merchants. Although each individual merchant would prefer a lower interchange fee to be applied to its acquiring bank, as well as the ability to price discriminate, in aggregate, merchants value a system of rules that ensures cardholders will also join in sufficient numbers. Thus, the welfare losses of dropping the no-surcharge rule could well be larger if the model also incorporated partial participation by merchants. Similarly, if fixed network costs were explicitly incorporated into the model, one would expect the welfare losses associated with under-subscription and under-usage of the card network when the no-surcharge rule is eliminated could be substantially greater than we estimate, as with less transactions the margins on these transactions would have to increase. We have also assumed that the card network would remain open and non-discriminatory despite surcharging. In practice, large issuing and acquiring banks may be able to reach deals with merchants to set surcharges in discriminatory ways against cards issued by an entrant bank, which to the extent it creates barriers to competition in the issuing market would again tend to raise the margins on these transactions.

One relatively straightforward extension to the model would be to explicitly incorporate a cost to merchants of handling cash.[26] We assumed cash was costless, despite the fact some merchants incur heavy costs to handling cash transactions. With costs of handling cash and transactional benefits of accepting cards, it is no longer obvious merchants will want to surcharge, even if allowed. However, even if merchants do not want to surcharge, the analysis of the paper is still relevant - both because if costs change merchants may then want to surcharge, and because it provides an analysis of the optimal setting of the interchange fee absent surcharging.

---

[26] In the Hotelling model this plays the same role as the transactional benefits b_m of accepting cards (which, in part, can be thought of as the cost saving from not handling cash).

T.1

References

Baxter, W.F. (1983) "Bank Interchange of Transactional Paper: Legal Perspectives," *Journal of Law and Economics*, 26, 541-588.

Coase, R.H. (1960) "The Problem of Social Cost," *Journal of Law and Economics*, 3, 1-44.

Frankel, A.S. (1998) "Monopoly and Competition in the Supply and Exchange of Money," *Antitrust Law Journal*, 66(2), 313-361.

Rochet, J and J. Tirole. (1999) "Cooperation among Competitors: The Economics of Credit Card Associations," Centre for Economic Policy Research, Discussion Paper No. 2101, March.

Schmalensee, R. (1998) "Payment Systems and Interchange Fees," mimeo, MIT.