THE REGULATION OF INTERCHANGE FEES: AUSTRALIAN FINE-TUNING GONE AWRY

Richard A. Epstein*

I. INTRODUCTION

This symposium provides a notable addition to the extensive literature that has developed on the operation of two-sided markets. Two-sided markets are generally defined as markets in which the value attributed to the goods and services received by parties on one side of an exchange depends not only on the intrinsic properties of those items, but also on the number of parties located on the other side. In the credit card context, which I shall address here, the story plays out as follows: merchants accept credit cards only

---

* James Parker Hall Distinguished Professor of Law, The University of Chicago; The Peter and Kirsten Senior Fellow, The Hoover Institution; Affiliate, LECG. The research for this paper was conducted with the financial support of Visa. My thanks to Tom Brown of Visa for his constant advice and to Howard Chang and David Evans of LECG for their continued involvement in my work. The views expressed herein are my own.

if customers are willing to use them to make purchases. Customers use credit cards only if merchants will accept them. Today, there also exists extensive literature describing the theoretical features of these markets, which, as this Article will explain, do not operate with the matchless efficiency of ordinary competitive markets.[1] Given this starting point, the question of industrial organization faced in two-sided markets is that of selecting the second best option. Because the ideal solution cannot be achieved, we must locate the closest approximation to that ideal that can be attained under practical circumstances. Stated more generally, this inquiry can be reduced to the question of whether state regulation can improve the operation of two-sided markets.

Fortunately, we do not have to examine two-sided markets solely from a theoretical perspective. We have available at least one real world example, namely the interchange fee restrictions that the Reserve Bank of Australia ("RBA") imposed on the credit card industry in July of 2003.[2] In a report entitled *Reform of Credit Card Schemes in Australia*, the RBA implemented a series of recommendations that had been announced in its comprehensive 2001 regulatory initiative.[3]

---

[1] William F. Baxter, *Bank Interchange of Transactional Paper: Legal and Economic Perspectives*, 26 J.L. & ECON. 541 (1983).

[2] For a more detailed discussion of the RBA's regulatory initiative, see Howard H. Chang et al., *The Effect of Regulatory Intervention in Two-Sided Markets: An Assessment of Interchange-Fee Capping in Australia*, REV. OF NETWORK ECON. (forthcoming 2005), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=820044 [hereinafter Chang, *Australian Interchange*].

[3] RESERVE BANK OF AUSTRALIA, REFORM OF CREDIT CARD SCHEMES IN AUSTRALIA I (2001), *available at* http://www.rba.gov.au/PaymentsSystem/Reforms/CCSchemes/IAConsultDoc/index.html [hereinafter RBA, AUSTRALIAN REFORM]. The RBA relied heavily on Michael L. Katz, Reserve Bank of Australia, *Network Effects, Interchange Fees, and No-Surcharge Rules in the Australian Credit and Charge Card Industry*, in REFORM OF CREDIT CARD SCHEMES IN AUSTRALIA II (2001), *available at* http://www.rba.gov.au/PaymentsSystem/Reforms/CCSchemes/IICommissionedReport/index.html.

To assess this report and the regulations it has spawned, it is useful to review the basic features of the credit card system. The credit card system is one component of a larger payment system, which includes cash, checks, debit and credit cards, and charge cards—the last of which provide credit only for purchases made from the business that issued the card.[4] Bankcard, an Australian operation, as well as Visa and MasterCard, employ what are known as open credit card systems, which stand in contrast to closed credit card systems, both of which are explained below. The RBA's key initiatives focused on two essential features of the open credit card systems used by Bankcard, Visa, and MasterCard. The first of these was the regulation of interchange fees. The second was a contract provision that prevents merchants in these systems from offering lower prices to their cash customers.

In order to assess these two initiatives, it is helpful first to understand how the closed, or integrated, systems of American Express and Diners Club operate. Both of these companies operate simpler three-party, or standalone, systems involving cardholders, card companies, and merchants. In closed systems, the credit card company acts as the sole intermediary between its cardholders and the merchant base, and it profits by retaining some portion of the payment from the cardholder to the merchant in exchange for its services. On the other hand, the more complex four-party systems employed by Visa and MasterCard operate a comprehensive network that consists of a large number of member banks. Therefore, these systems need to employ the services of two banks instead of a single company intermediary. The four roles in these systems line up as follows: the *cardholder* receives the card from the *issuing bank*, which transacts with the *acquiring bank,* which in turn has connections with the *merchant*,

[4] For accounts of these systems, see DAVID S. EVANS & RICHARD SCHMALENSEE, PAYING WITH PLASTIC: THE DIGITAL REVOLUTION IN BUYING AND BORROWING 1-21 (2d ed. 2005).

whose customers pay for their purchases with either credit or debit cards.

Visa and MasterCard serve as cooperative or coopetitive arrangements of the various member banks that perform recordkeeping and orchestrate the interactions between the issuing and the acquiring banks. Using the term "coopetitive" highlights that each system's member banks must cooperate in setting standards for credit and debit card transactions, even as they simultaneously compete with each other to sign up consumers and merchants on either end of the market.[5] Their "interchange fees" refer to the amounts that issuing banks collect from acquiring banks for processing a credit transaction, which the acquiring banks then pass along to merchants as additional service charges. Functionally, the interchange fee is usually deducted from the total amount owed from the issuing bank (for its customer) to the acquiring bank (for its merchant) for the relevant transaction. Before Australia enacted its regulatory scheme, the interchange fee for open credit card systems was set at 0.95% of the value of a transaction. The RBA's first initiative reduced that figure to 0.55%, a decrease of approximately 43% relative to the original charge.

The RBA's second initiative, dealing with price cuts for non-credit card users, is of lesser import than the first. Typically, contract terms in open credit card systems prohibit the merchant from offering any discounts to those customers who prefer to pay by check, cash, or debit card. In effect, these terms prevent merchants from inducing customers to favor one of these three modes of payment, each of which the RBA found to be cheaper in total cost than credit transactions.

Ultimately, these regulatory interventions are intended to address what the RBA claims are efficiency and distributional concerns, thereby reducing the implicit subsidies that cash, check, and debit card users pay for credit card users. The RBA's program is explicitly designed to improve the position of merchants (and their broad

[5] *Id.* at 161-84.

customer base) and acquiring banks relative to that of issuing banks and their cardholders (which are only a subset of all customers). This Article shall examine whether implementing these reforms has allowed the RBA to make good on its objectives. In evaluating the outcome, I do not wish to reexamine the detailed empirical issues raised by the RBA's initiatives. For these purposes, I am quite content to rely on three main sources. The first is the empirical evidence assembled by Howard H. Chang, Daniel S. Evans, and Daniel D. Garcia-Swartz.[6] The second is testimony offered by Ian Macfarlane, Governor of the RBA, in defense of the RBA's 2001 initiatives.[7] The final source consists of the various submissions made to the RBA on these matters.[8] That evidence provides a useful framework with which to reexamine the conceptual arguments that led the RBA to adopt these proposals in 2001, in light of their impact over the two years.

Accordingly, the remainder of this paper is divided into seven parts. Part II sets out the basic intellectual framework for examining the RBA's arguments in defense of its reforms. Part III applies this framework to the arguments that the RBA made in support of capping interchange fees. Part IV examines the RBA's prohibition on credit card surcharges. Part V examines the use and limits of credit card regulation. Part VI offers preliminary remarks on the relative merits of regulation and antitrust litigation for controlling the use of credit cards, should such controls be necessary. Finally, Part VII offers a brief conclusion.

---

[6] *See generally* Chang, *Australian Interchange*, *supra* note 2.

[7] Review of the Reserve Bank of Australia's Annual Report 1997-98: Hearing Before the House of Representatives Standing Committee on Economics, Finance and Public Administration (2005) (statement of Ian Macfarlane), *available at* http://www.aph.gov.au/HANSARD/reps/commttee/r2365.pdf [hereinafter Macfarlane Testimony].

[8] *See generally* Reserve Bank of Australia, http://www.rba.gov.au/PaymentsSystem/Reforms/CCSchemes/index.html (last visited November 1, 2005).

II. METHODOLOGY AND CONCLUSION

My basic thesis runs as follows: the RBA reforms have not delivered their promised benefits because they are incapable of doing so. This conclusion depends as much on the underlying intellectual framework as on the particulars of the reform effort. Quite simply, payment systems can never achieve the results attainable in perfectly competitive markets, so the best option under the circumstances is the one that best minimizes social losses. In light of this caveat, any evaluation of the RBA's reforms must be nuanced enough to ask this question: are the benefits that result from the regulations large enough to overcome the obstacles to, and costs of, achieving the stated ends? The key feature of the payment card industry is that the demands of network formation require a high industry concentration within limited affiliated groups. Even when the issuing banks do not collude with each other, prices on either side of the market could easily exceed marginal cost. But owing to the interdependence between both sides of the market, there exists no effective remedy for this imperfect state of affairs. Direct systems of price regulation create as many distortions as they remove. Moreover, antitrust or competition policy cannot fashion a useful remedy for an industry that cannot operate in pure competitive fashion. In the world of second best options, it is unwise to invest limited resources in proposing and resisting regulation, where they yield a low rate of return.

This conclusion is not based, either largely or exclusively, on any appeal to public choice theory, which suggests that political factions can capture public administrative bodies. Even though key interest groups lobbied the RBA hard for their preferred positions, there is no real evidence that the RBA did anything other than seek to maximize its vision of social welfare. Although merchants as a class benefited from this proposal, the reforms were not promoted as an effort to achieve that partisan end. Rather, the RBA seems to have followed the highest standards of public responsibility in promulgating these regulations. Nevertheless, the RBA's

proposal still lacks the intellectual and institutional soundness to make it work.

The key intellectual objection to the RBA's reforms stresses the limited potential for social gains under this regime. Because the RBA chose to regulate those credit arrangements that are both "consensual" in origin and "durable" in effect, it effectively placed a sharp upper boundary on the potential for social gain under its intervention.[9] Consensual transactions routinely describe those situations in which all parties to the arrangement are better off than they were before entering into the arrangement.[10] These payment transactions therefore differ sharply from situations where the "unilateral" actions of one party have negative external effects upon a second party, such as in instances of aggression or pollution. Moreover, "durable" transactions eliminate certain risks that often infect voluntary transactions, including mistake, nondisclosure, and misrepresentation. While these risks could be present in one-shot transactions (e.g., the private sale of a used car), they are not likely to exist among repeat players, especially the sophisticated parties involved here.

In the payment industry, for example, the feedback mechanisms on information are exceedingly strong because the entire system can operate only if the parties preserve complete and accurate records of all transactions and share this information with both cardholders and merchants on a routine basis. While the element of surprise may infect some individual transactions, it is not the norm in any situation involving either checks or payment cards.[11] To be sure, most individuals do not have perfect knowledge of the full range of

[9] This conclusion holds whether such changes result from direct regulation or from private litigation, but particularly under the antitrust laws in the United States or the competition law, as it is described, elsewhere in the world.

[10] Baxter, *supra* note 1, at 552-56.

[11] A truth in lending remedy may prove appropriate in these cases. For a useful discussion concerning the topics of misrepresentation and disclosure, see Rossman v. Fleet Bank Nat'l Ass'n, 280 F.3d 384, 389-91 (3d Cir. 2002).

available card alternatives, even though competitors will constantly try to place better terms in front of them. Nonetheless, these individuals do have enough information to be sure that they are not made worse off by their particular pattern of doing business than they would have been if they had not used any credit device at all. The purpose of the RBA's regulation, therefore, was not to prevent negative sum games from emerging among the four parties to the standard open-system credit transaction. Rather, its goal was to improve the outcome of positive sum games in which all individual players shared at least some fraction of the gain under the previous institutional arrangements.

In principle, it is always possible to conceive of some regulation that can improve upon the status quo, but, as chess players concede, it is much more difficult to find that winning move on the board. In particular, there are two systemic obstacles that make this endeavor more difficult than would otherwise be the case: administrative costs and error costs. The first obstacle is that the proposed system or regulation always adds additional costs. At a minimum, these include the costs of deciding what should be done. Moreover, within the modern administrative state, new rules cannot be imposed by fiat, but instead require a period of reflection, research, comment, and debate before implementation, and in many cases, cry out for constant revision thereafter. Agencies justify this elaborate ritual as a means to acquire the information necessary to prevent their regulations from being, at best, hit-or-miss affairs. Even if the RBA's relatively simple price and term restrictions come only on the heels of this lengthy process, they are surely less costly than those setting out, for example, the requirements for workplace or environmental safety. In light of the extensive commentary and debate, the total costs have likely run into the millions of dollars on a cumulative basis, in what has proven to be an ongoing

project.[12] These costs are borne in part privately, by the firms that participate in the administrative process, and in part publicly, by all of the citizens whose tax dollars support these initiatives. Given this first obstacle, regulators should not pursue a program unless they have some reason to believe that the total gains realized therefrom are greater than the full administrative costs required to implement it.

On the other hand, evaluating the potential error costs of a proposal leads one to favor the status quo even more than evaluating the potential administrative costs. The success of a proposal depends critically on whether its ultimate design takes into account all of the relevant variables. That task is not easy to discharge because any proposed regulation, regardless of its superficial simplicity, will have ripple effects on multiple margins. More concretely, any regulatory initiative could be said to promise some form of social gain if the only consequences that mattered were those that fell within the terms of reference in the original study. But it is one thing for regulatory review to consider a limited set of consequences; it is quite another to be confident that the regulator has included within the frame of reference all of the economic consequences that matter. As I will explain, one key margin on which the RBA's proposal should be evaluated is its impact on the competitive balance between open and closed systems, which the RBA has systematically and incorrectly refused to undertake.

In conventional academic terms, the basic challenge of evaluating regulatory reform requires a close look at the persistent tradeoff between validity and reliability. The study with the greatest validity seeks to examine the impact of a regulation along as many margins as possible. For example, a study of the minimum wage law would be insufficient if it only considered possible short-term decreases in employment levels. It would also be necessary for the study to contemplate the impact on other terms of the employment relationship, including shift lengths and the

[12] *See* Reserve Bank of Australia, *supra* note 8. I do not know of any effort to sum the costs of these various efforts.

amount of training. Even a study that considered these additional facets of the problem would be incomplete if it did not also consider the impact on the level of capital improvements undertaken, the educational efforts of low income workers, the relative wages of workers in the middle of the wage pyramid, and so forth. But the further the range expands, the more difficult it is to control for the relevant variables, and the reliability of the conclusions is compromised as the set of relevant considerations expands, seemingly without limits. Finding the golden mean between the two extremes raises a persistent methodological problem that is easier to state than to resolve.

While an assessment of the minimum wage law is more complex than an assessment of many regulations governing the credit markets (e.g., what counts as an "hour" is only one of many issues to be faced), this example illustrates the necessity of considering all aspects of that problem before making any empirical judgments. Defining the proper scope for assessing the likely consequences of regulation is also difficult. The wider the net is cast, the more complete the analysis, but the less reliable the data. Nothing guarantees that we can find some combination of breadth and concreteness that makes the enterprise worth undertaking. On the one hand, the task of policymaking becomes ever more complicated. On the other, reliance on simple solutions with low administrative cost rightly becomes the implicit default position, especially when efforts are made to override consensual arrangements.

One minimum condition for this regulatory effort is to set a legislative frame of reference that considers the position of all competitors within the same overall market niche. That was not done in this instance, given the RBA's decision to refrain from regulating merchant discounts within the closed systems of American Express and Diners Club. In effect, the RBA relied on a partial equilibrium analysis with a big piece missing, namely the operation of the direct competitors of the regulated parties. If the RBA had performed a general equilibrium analysis, it would have recognized that the different forms of payment mechanisms operate at some

level of (imperfect) competition with each other. Such omissions need not be fatal, but in most cases they are likely to seriously skew the study of a regulation whose upside potential is limited by gains achievable under the present consensual and durable relationship. In this environment, poorly considered proposals are especially improvident because the benefits of a program must be large enough to cover both the anticipated administrative costs of the system and the extensive error costs driven by the artificial limitations in the study's initial frame of reference.

These general remarks are not confined to intervention in the credit system alone, but reflect upon the broader purposes of regulation within the financial markets generally. Yet before reaching the difficult issues that the RBA faced here, it is necessary briefly to consider two other types of regulation. The first type seeks to regulate overreaching and fraudulent or misleading behavior in transactions. These risks are not present here. The second type, which concentrates on the regulation of cartels and monopolies, is intended to prevent anticompetitive behavior. This risk is present here only in an attenuated fashion and is largely inapplicable to ratemaking questions.

In light of the strong background features that exist in the payment card system, any alterations to the incentive structure necessarily will affect the behaviors of the key players in the system. However, the behavioral responses are unlikely to correlate with any overall sense of social improvement because of the inability to identify a large target of opportunity: the case is not about stopping cartel conduct, for example. In addition, clearly there can be no strong Pareto improvement because the entire purpose of the reforms was to make the position of the credit card holders and issuing banks more precarious than it had been under the status quo. In itself, this criticism is hardly decisive, because the exacting Pareto standard is commonly thought to be too rigorous with respect to any social reform that does not begin with a blank slate. It is simply too much to demand that no one be made worse off when at least one person (or group of persons) is made better off. But,

conscientious social reforms should satisfy, at a minimum, the somewhat weaker Kaldor-Hicks standard. Under this standard, the RBA regulations would have to produce more gains for the merchants and their non-credit card customers—an ambitious and ambiguous objective[13]—than losses for the issuing banks and their customers, especially since there is substantial overlap among the populations of the various groups. Yet that too is an uphill battle. A closer look at the allocative and distributional consequences belies any reason to think that this standard has been achieved.

III. INTERCHANGE REGULATION

A. Merchant Choices

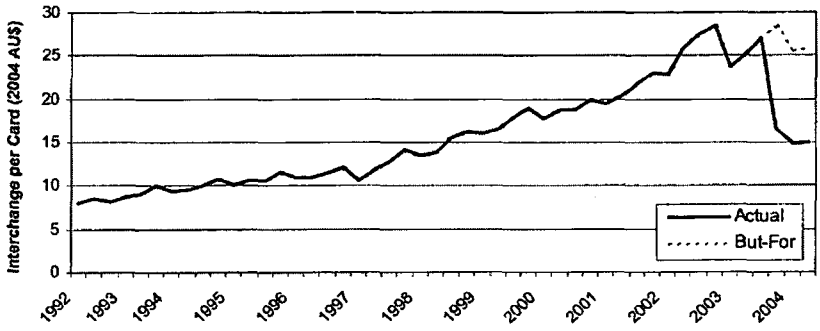
The more important of the RBA's two reforms is the sharp restriction on the level of interchange fees. The RBA established this restriction to remove the set of subsidies that led to the systematic overuse of credit cards relative to other forms of payment. The RBA argued that the high interchange fees forced *all* consumers of covered merchants to subsidize a payment system that only benefited credit card users. This conclusion contains both an allocative component (that the fees in question led to a distortion of relative prices) and a distributional component (that subsidies were extracted from non-credit card users without sufficient social justification).

The allocative claim rests on what can only be termed the odd view—especially odd for a national bank system—that credit cards provide few if any benefits over alternative payment systems. That conclusion seems to conflict with the obvious historical fact that credit cards have expanded their market share of payment transactions both before and after

[13] For discussion of these problems, see *infra* the discussion of passthroughs in Section V.

the 2003 capping of interchange fees.[14] To illustrate, Figure 1 shows a continuous upward trend that is so smooth that no one could figure out from visual inspection when the RBA imposed its cap on interchange fees.

Figure 1. Real Interchange Revenue per Card[15]



Notwithstanding this obvious trend, the RBA's basic claim is that "[h]igher sales on credit cards for individual merchants do not, of themselves, give rise to overall merchant benefits if: (i) those sales would have taken place anyway using other payment instruments; or (ii) the sales have merely diverted business from one merchant to another."[16]

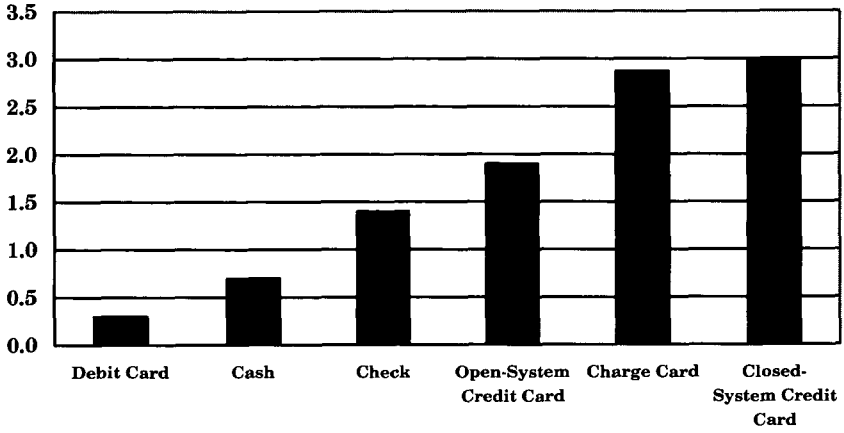
The background assumption underlying these two conclusions is that credit cards do not add much net benefit to the operation of the economy and only divert transactions into more expensive modes of payment. In support of this conclusion, the RBA notes that credit card fees comprise a much heftier percentage of sales than fees for other forms of payment, as shown in Figure 2 below.

[14] For the overall progression, see EVANS & SCHMALENSEE, *supra* note 4, at 84 (noting the steady rise in use of payment cards from about 2% to 25%).

[15] Chang, *Australian Interchange*, *supra* note 2, at 21.

[16] RBA, AUSTRALIAN REFORM, *supra* note 3, at 23.

Figure 2. Payment Costs to Australian Retailers as a Percentage of Average Transaction Value[17]



The last line, which refers to closed-system credit cards (American Express and Diners Club), was not in the RBA's original report. That bar can be added in, however. More generally, what is most instructive here is how this rate compares with that of the open-system credit cards (Visa and MasterCard), both before and after the imposition of rate caps. It seems as though the figures are derived from a different metric because the absolute values differ. But the *relative values are the same, and they speak to a real price advantage for both MasterCard and Visa.* Before the rate caps were imposed, MasterCard and Visa (the open-system credit cards) had total interchange fees of approximately 1.45%, relative to 2.58% for American Express and 2.35% for

[17] RBA, AUSTRALIAN REFORM, *supra* note 3, at 22. In this figure, the first five columns were taken from the RBA report, but the sixth column, representing closed card systems like American Express and Diners Club, has been added in because the fees for these systems are higher than those of the four-party systems. *See* EVANS & SCHMALENSEE, *supra* note 4, at 152 fig.6.1 (reporting the figure for American Express at around 3%). The implications of this omission will be discussed presently.

Diners Club (the closed-sytem credit cards).[18] The interchange fees for MasterCard and Visa were thus approximately 1% lower than for American Express and Diners Club. Once the rate caps were imposed, the interchange fees for MasterCard and Visa dropped to approximately 0.96%, while those for American Express and Diners Club remained at 2.38% and 2.31%, respectively.[19] The unregulated sector barely budged in response to the rate caps in the regulated sector, so that the gap in interchange fees between the two types of cards only increased. It seems odd indeed that the regulatory initiative is directed exclusively at that form of business which does better under the RBA's own test for relevance. The attack on the four-party, open-system, credit cards seems at best misplaced.

To bolster its attack on these four-party credit card systems, however, the RBA takes the position that the transactions involved in these cases are not "really" consensual at all. Rather, merchants, both large and small, succumb to the blandishment of credit cards because "credit card use has become so widespread that most merchants believe they have no choice but to accept credit cards."[20] I use the word "really" in quotations to stress the RBA's questionable insistence that merchants have "no choice" in this critical matter. A far simpler, albeit more cynical, interpretation of the phrase "no choice" is that merchants like to use credit cards, but, like the rest of humanity, would rather pay less for something from which they already derive net benefits.

Moreover, a closer analysis reveals the conceptual confusion that surrounds the easy invocation of this "no choice" model. Start with this question at the institutional level: Why would consumers and merchants choose to support a system of credit that is both clumsier and less

[18] *See* RBA, Bulletin table C3, http://www.rba.gov.au/Payments System/PaymentsStatistics/payments_data.html (last visited Nov. 21, 2005).

[19] *Id.*

[20] RBA, AUSTRALIAN REFORM, *supra* note 3, at i.

efficient than the systems already in place, i.e., cash and checks? The slippery phrase "no choice" only obscures the RBA's central claim. For example, one colloquial meaning of the phrase "I had no choice" refers to the common situation where a robber comes up to you and says: "Your money or your life." In one sense, you remain, to misuse a phrase from Milton and Rose Friedman, "free to choose" either alternative. Yet, anyone who decides to hand over the money can rightly say this "choice" did not count as vindication of free will, but rather provided conclusive evidence of illegal duress. The standard libertarian theory, which treats the use of force as the primary evil, has always adopted this anticircumvention principle as its corollary: to limit the effectiveness of force, the *threat* of force must be subject to the same prohibitions and coupled with the same exceptions. In other words, a threat of force in self defense is as legitimate as the use of force.

Any threat that takes the form "your money or your life" is especially dangerous because it leaves the putative wrongdoer *better off* than he would have been had he just used the force itself. You value your life more than your money. The robber values the money more than your life. Individual differences are not important in this stark calculation. So the deal is struck, leaving both sides better off than under the alternative under the circumstances. But this happy account misses the obvious rejoinder. The robber has no right to put you to that choice; he is *not* like the taxman who says: "Give us your cash or we will seize your specific assets to pay your lawful debt," which is another offer that works to the benefit of both sides. Rather, when facing the thief, you are entitled to both your money *and* your life, so, therefore, his "choice" gives you only a means of minimizing loss, not securing gain from the prethreat state of the world. Contrast that position with the ordinary business offer in which you have to choose between *your* money and *his* goods. The mere presence or absence of choice is neither here nor there. The choice given to you by the robber does not generate systematic win-win outcomes. Market exchanges do.

Against that backdrop, the phrase "no choice" in the RBA report gains no traction from the common law cases of duress. Rather, all credit transactions fall within the domain of win-win transactions, where the only quarrel is over the size of the gain for each party. Nor could it be otherwise. If the credit card system is as inefficient as the RBA supposes, then any merchant could effortlessly decide to accept only cash, checks, or debit cards and eliminate the other options. This unilateral threat (or option) to withdraw is perfectly credible no matter what decision other merchants make regarding which payment methods to accept. To avoid this conclusion, the RBA takes the position of the *marginal* merchant, who must decide whether to stay in the game or forego the enormous benefits that credit cards have "of attracting sales from, or not losing sales to, rival merchants."[21] But if the basic assumption is that credit cards are inefficient, then this dodge fails because it does not acknowledge the last crucial step: Why is it necessary for any merchant to join an inefficient system to attract customers from rival merchants or to keep them from sliding off to competitors?

To see why, think of this as a prisoner's dilemma game in which each merchant, acting out of narrow self-interest, has the option to participate in or opt out of the credit card system. Opting out is the dominant solution. If other merchants cling to the expensive system, then the single merchant that lowers prices by one or two percent has a real advantage for himself. Thus, he should defect from credit cards. Alternatively, if the other merchants opt out, then the no-credit card strategy is needed to keep pace with others who have previously adopted the no-credit card strategy. He should defect again. If the credit card system offers a more expensive payment system without offsetting benefits, then the migration from it should be fast and furious, as each merchant follows the dominant strategy of ditching credit cards in favor of cheaper cash, check, or debit card payments. There is no need for government action to eliminate the

[21] *Id.* at i.

dreaded cross-subsidy. Merchants can do this unilaterally. When they do, the people who invest in credit cards will cut them in half precisely because they have no reason to clutter up their wallets.

Yet, that is just not in the (credit) cards. Instead, we see the reverse flow taking place as lower transaction costs, in the form of faster computers, more reliable scoring methods, and other improvements, have led the credit card market to expand into niches that it had not previously penetrated. Five or ten years ago, credit cards were not accepted at fast food outlets like McDonald's and Subway. The sums were too small and the per transaction costs were too high. If the RBA's assumptions were correct, this equilibrium should have proven stable because every firm could stand pat without having to play follow-the-leader. However, the opposite result ensued. Special price deals, one suspects, brought this segment of the market into the credit card orbit. It is unlikely that the decision was just a bad mistake. Clearly something is deeply amiss in any analysis whose major premise predicts the demise of credit cards in the midst of an explosion in their use. To see what this problem is, we have to examine more closely the proposition that credit cards have no advantages over other systems of immediate payment. The key is finding benefits that differentiate all types of credit payments from all other payment systems.

B. The Not So Subtle Credit Card Benefits

To highlight the benefits of credit cards, divide the landscape of payment methods into two categories. Charge and credit cards, including those issued by both open and closed systems, are on one side of the line. Checks, cash, and debit cards, which I collectively term "debit instruments," are on the other side. The credit and charge instruments, which are more expensive than the debit instruments, also perform additional functions. Because the debit instruments have no built-in credit component, they only allow the purchaser in a given transaction to spend the money that he or she already has in his pocket or salted away in some bank

account. These stores of wealth do not allow purchasers to buy goods or services that are desired today with a promise to pay for them tomorrow out of future earnings. Nor do they give their users the prospect of free float for a month, which is attractive to high-income borrowers who have sufficient funds to pay off their bills when they become due. With debit cards, it is as though one wanted to use savings to do the work of life insurance. Yet, it takes little imagination to realize that most people would like to make their consumption patterns smoother than their earning patterns.[22] No profound demonstration of this point is needed at this time. Just recall that, even as adults, people borrow while in school yet simultaneously save for retirement. Thus, the earnings in an individual's peak period are moved both forward and backward in time.

Debit instruments cannot perform this function of tapping future income. In contrast, credit and charge cards facilitate this goal on a shorter time frame by allowing individuals to pay over time for goods and services, whose value to the consumer is captured over time. Few people buy homes with cash; instead, they take out home mortgages, where the monthly payments closely track the benefits derived from the initial purchase during each pay period.

A question then arises over which type of card is superior, the store charge card that is valid at only one merchant, or the credit card from either a closed or an open system that may be used at any merchant who has agreed in advance to accept the card. Clearly, if one method was dominant over all others, then we would see either credit or charge cards driven from the marketplace. But the simple fact that Neiman Marcus does not accept general credit cards provides evidence that each card has its own distinct advantages enabling it to survive. Charge cards, for example, may entitle their users to certain privileges that non-charge card holders cannot receive, including the receipt of information

[22] For a basic discussion of the permanent-income hypothesis, see Robert Schenk, *Permanent-Income Hypothesis* (2002), available at http://ingrimayne.saintjoe.edu/econ/FiscalDead/PermIncome.html.

about various sales and special offerings; free services such as gift wrapping and holding or delivering parcels; the ability to purchase, or purchase more quickly, by phone; and the ability to go to the head of the line at the (quaintly termed) cash register. The obvious point here is that the charge card signals a form of loyalty that generates loyalty discounts in return. In a competitive marketplace, there is no reason whatsoever to regulate the pricing structure of these charge cards any more than there is to regulate the price of pajamas or alarm clocks. After all, reductions in the price of pajamas are said to create an implicit cross-subsidy from purchasers of alarm clocks to those of pajamas because of an implicit shift of some joint costs from the former to the latter.

Nonetheless, there is a simple but compelling explanation that explains why credit cards have gained ground in recent years relative to charge cards: the credit card in question may be used at any one of a number of stores, which allows its holder to compare prices and goods across different merchants in an effort to obtain the best available combination of price and quality.[23] It is a mistake to assume, as the RBA did, that the net result of any product or service substitution should be scored as a wash. Each substitution should generate some additional level of consumer surplus. Merchants, therefore, find themselves in an unhappy position. If they stand aloof from the network, they lose the ability to compete for credit purchases that may benefit them. So, they join the greater number of merchants who accept credit cards and thus make the market more competitive for everyone else, because they sensibly expect that more shoppers will come their way. A similar reason explains why jewelers or diamond dealers all like to work in the same district. Their close proximity generates higher overall traffic that more than compensates for the stiffer competition. The consumer benefits in both cases are unambiguous.

Credit cards also have other advantages for which debit instruments are an ineffective substitute. Try to rent a car

[23] *See* EVANS & SCHMALENSEE, *supra* note 4, at 122-24.

at the airport or check into a hotel without a credit card. Usually, it is necessary to provide a credit card to cover the expenses of a trip. The protection here is imperfect because the card limit may be lower than the customer's potential tab. This recognition explains why most consumers seek to raise their credit limits, as opposed to keeping them low (in an effort to guard against binge buying, which people can indulge in on store credit). But any card is likely to prove superior to leaving wads of cash on the table or signing over a blank check that may or not be backed by sufficient funds. Credit cards, therefore, also have their distinct market niche. The question remains, though, how the regulation of interchange fees in open systems will manifest itself.

C. The Impact of Interchange Fees

1. Debit Instruments

The obvious implication of the RBA report is that the removal of the cross-subsidy will increase, to some extent, the frequency of use for all debit instruments. The extent of this effect depends in large part on the relative price differences observed by consumers. In this regard, it is critical to look back to the relative costs of the two transactions as outlined in Figure 2 above.[24]

Here, the costs portrayed are not the total costs to the user, but only the *transfer* payments that users make to merchants in order to participate in these transactions. But the situation from the customer's point of view is decidedly different than this figure indicates because cash and checks, in particular, have costs that are not reflected in these transfer payments.

The private costs associated with the use of cash and checks are plentiful and should not be ignored. Most people withdraw their cash from an ATM. When cash is drawn from a machine owned by another bank, consumers frequently pay a flat fee that, for most withdrawals, will be

[24] See *supra* note 17 and accompanying text and graphic.

equal to about one percent of the money received.[25] In addition, to maintain ready access to their funds, consumers must keep their money in checking accounts that pay little or no interest. Often there are also fees for maintaining checking accounts or for cashing individual checks. Another private cost is risk: both cash and checks can be lost or stolen.[26] With regard to cash, what is gone is gone; there is simply no way to stop payment on this ultimate negotiable instrument. All of these aforementioned costs must be added back into the cost of transacting by cash or check.

One response to this concern argues that the amount of money lost or stolen is trivial in relation to the total amount of cash in circulation. While this may be true, it is not a decisive criticism because the amount of cash lost is typically small because most people choose to carry only small amounts of cash. They are able to do this because they have credit cards at their disposal for most transactions, especially for big ticket items. Few people will opt to carry around the thousands of dollars that it would take to purchase a plasma television when they can simply take a credit card instead.

The costs associated with the risk of theft also applies to the use of checks in stores. Anyone who has ever had his or her checkbook stolen knows how difficult it is to stop payment on forged checks. A theft may go undetected for

---

[25] Greg McBride, *Checking Study: ATM fees keep climbing*, BANKRATE.COM, May 11, 2005, http://www.bankrate.com/brm/news/chk/20050511c1.asp. The study reports that the average fee for withdrawal from a "wrong" bank ATM is now about $1.35 and that estimated total annual fees exacted from these transactions approximate almost four billion dollars in 2005. A second recent survey placed the average ATM service charge at $1.55. Mike Licker, *NYPIRG Plans to Warn Students About Expensive ATM Service Fees*, THE DAILY ORANGE, Apr. 26, 2004, http://www.dailyorange.com/media/paper522/news/2004/04/26/News/Nypirg.Plans.To.Warn.Students.About.Expensive.Atm.Service.Fees-671154.shtml. In some cases, a percentage fee is charged depending on the size of the withdrawal, but these surveys did not report as such.

[26] *See* EVANS & SCHMALENSEE, *supra* note 4, at 30-31. Evans and Schmalensee report that $850 million in cash was stolen from 84,000 robberies and pickpockets in 1995 alone.

hours or perhaps even days. In the meantime, large amounts of money can be charged against a consumer's account, particularly if merchants fail to seek verification of the check writer's signature. Thus, the obvious strategy is to leave the checkbook at home because it is safer to carry a credit or debit card. This analysis suggests that debit cards, due to their inherent safety value, have an even greater transactional advantage than the overall numbers suggest and may explain why their use has increased so rapidly.

Moreover, future developments will continue to influence the mix of payment instruments for debit transactions. One issue worth examining is how the use of online bill payment fits into the overall scheme of payments.[27] While these online transactions do not count as credit payments, they do transfer some portion of the payment volume from checks into electronic exchanges, since the same banks that run the credit systems support the applicable networks. The benefits of online payments are manifest: time savings, instantaneous financial position reporting, and other such personal and financial benefits. One important advantage of this ongoing development is that it allows individuals to better manage their credit card balances and, as a result, to control their interest rates. Even if banks had preferred the older systems of payment, competition would nonetheless have brought about a shift towards online payment because of these kinds of advantages for the consumer. And, in fact, most banks also prefer online banking transactions because of their own cost savings. Furthermore, the whole system contains within it no element of government subsidy because it more closely resembles a private form of money, albeit one parasitic on the general currency. It is thus clear that checks, once considered universal payment options, are becoming obsolete. Why perpetuate such obsolete systems?

[27] Terry Savage, *Why online bill payment is booming*, MSN.COM (2005), http://moneycentral.msn.com/content/Banking/Betterbanking/P38218.asp (anticipating that this year forty million Americans will pay bills online to avoid the inconvenience and cost of writing checks).

Aside from efficiency and theft concerns, another important issue relates to the number of subsidies that are involved in the respective payment systems. The entire clearance system for credit cards is privately financed; it is as though the dream of a private money system has been realized by indirection. The same cannot be said for either the cash or the check system, both of which receive extensive government subsidies.[28] The national treasury bears the cost of printing, distributing, and protecting currency. The entire system of check exchanges relies upon the operation of the various federal banks. Hence, one way to look at the RBA initiative on its own terms is to see that it substitutes one implicit cross-subsidy (for cash and check users) for one explicit cross-subsidy (for credit card users). The net resource implications of such subsidies are not clear. How is the state subsidy, which falls on all persons alike, better or worse than a cross-subsidy that applies within an elaborate industry network of consenting participants? It is far from clear whether the RBA's dismal view on credit cards is an accurate one.

The final issue relates to the utility of cash and checks. Cash represents a national currency, and persons who travel overseas incur significant conversion costs that can easily consume one to two percent of the money's overall value. Checks, for their part, have an even narrower scope of use. They are of little use in foreign countries and, even within the United States, have limited value outside of the community served by the issuing bank. It is harder to run a credit check on a checking account than on a credit card, and people do not want to suffer the costs associated with bad checks. Credit and debit cards are both immune from these relative disadvantages. Given all of the concerns associated with the use of cash and checks, we should expect that, as

[28] I have not been able to find exact estimates of the cost of these subsidies, but the number is commonly stated to be in the billions. For one estimate, see Océ, *Check Clearing for the 21st Century Act: Printing Implications for Financial Institutions and Service Bureaus* 4 (working paper, *available at* http://www.ondemandjournal.com/whitepapers/Check_21.pdf).

technology continues to improve over time, the use of both credit and debit payment options will further increase relative to traditional modes of payment.

2. Closed-System Payments

The substitution from the open systems, Visa and MasterCard, to the closed systems, Diners Club and American Express, is more critical to this systematic analysis. The economic dislocations here stem from the truncation and inadequate frame of reference of the RBA's *interchange* inquiry. Interchange fees arise only in a credit card network that utilizes both acquiring and issuing banks. In contrast to the open systems, the closed systems utilize a single company that serves as the sole intermediary between its customers and its merchants. The more compact nature of a closed system's business obviates its need to levy any interchange fees. However, such a system is not free from the underlying economic issue that prompted the RBA's initial effort: Were merchants as a group—and, indirectly, their non-credit card customers—being asked to foot too much of the cost of running the entire credit card system, thereby creating an implicit subsidy for credit card users? That question arises solely because the merchant, through the payment system, receives only a portion of the face amount of the credit card charge. The issue is actually more acute for merchants (and their customers) transacting with the closed-system players, who charge a *larger* percentage of the face amount of the transaction than the open-system players, while avoiding the use of interchange fees.

Why then limit the inquiry to interchange fees when the merchants' complaint applies equally to both forms of credit transactions? Surely merchants cannot object to the high costs of their own store cards. At this point, moreover, the rate of substitution between the two rival card types (open-system and closed-system) should be greater than that observed between credit cards and any debit system. If these credit advantages prove tangible, then we would expect closed systems to garner a larger share of the overall market after the cap on interchange fees than they had prior to the

cap's imposition. If this swing does occur, then the ostensible subsidy that the closed systems' cardholders receive from non-credit card users should increase accordingly.

This prediction seems to have been borne out by the recent data collected by the RBA, which shows an increase in the use of closed-system cards relative to open-system cards, even though the gap in the merchant discount has grown steeper since the advent of the RBA's selective administrative rule. It is a fact that three-party (closed-system) cards now account for 17% of transactions by cash volume as of August 2005, up from 13.7% as of September 2003.[29]

In the face of this criticism, the RBA retreats to a set of unpersuasive bromides. The RBA claims that "at this stage, such regulation [of both types of card systems] would not improve the overall efficiency of the payments system," that it would have "relatively little effect on merchant charges," and that any effort to reach this problem could "only be addressed through considerably more extensive regulation than that currently existing in the credit card schemes."[30] The RBA advances these defensive claims without support from any empirical or theoretical evidence. In particular, the RBA does not explain why *no* regulation of *either* closed or open systems is better than the *partial* regulation of one type of competitor with its attendant distortions. Nor does the RBA offer any explanation as to why the admitted shift in transaction volume has taken place, if not for the competitive boost.

The simple explanation for the problem lies in the peculiar economics of two-sided markets, which include

---

[29] Letter from Leigh Clapham, Senior Vice President, MasterCard Australia, to Dr. John Veale, Head, Payments Policy of the Reserve Bank of Australia (Aug. 25, 2005), *available at* http://www.rba.gov.au/PaymentsSystem/Reforms/CCSchemes/SubmissionsDCCCIStd/mc_25082005_1.pdf [hereinafter Clapham Letter].

[30] *See id.* at 3; *see also* Press Release, Reserve Bank of Australia, Payments System Reform (Feb. 24, 2005), *available at* http://www.rba.gov.au/MediaReleases/2005/Mr_05_02.html.

credit card platforms. Both closed-system merchant charges without regulation and open-system merchant charges with regulation must satisfy this minimum condition: all relevant parties must desire to remain in the network. Two-sided markets are distinctive in that the charges levied on all players in the system do not equate with these players' marginal costs, as they generally do in competitive markets where marginal costs are constant. In competitive market situations, there are plenty of bargains, but no bargaining, because of the unique competitive equilibrium that sets price equal to marginal cost.

That unique equilibrium does not exist in two-sided markets. Hence, a pricing problem arises that is not dissimilar to the problem that arises in connection with the general marginal cost controversy: How does one price an item when the cost of the first unit is exceedingly high (e.g., a new bridge or a new drug), but the cost of each subsequent unit is very low?[31] The decision to charge only marginal cost for all units means that the developer cannot recoup its upfront costs over the useful life of its asset. Some form of subsidy, therefore, must be administered without any clear sense that the ultimate social value of the product will justify the subsidy's imposition. The only alternative pricing method, which in most cases is thought to yield fewer distortions, is to allow prices on subsequent units to be set above marginal cost, allowing for some loss of the user base, so that the company may recover its upfront fixed costs. The point here is that even in this relatively simple scenario, there is no optimal solution that reaches the happy outcome of an ordinary competitive market where—in the limiting case—marginal costs are constant for all levels of production.

At this point, it should be clear that the two-sided market problem shares a common property with the high fixed, but

[31] For the original proposal, see Harold Hotelling, *The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates*, 6 ECONOMETRICA 242 (1938), as criticized in R. H. Coase, *The Marginal Cost Controversy*, 13 ECONOMICA 169 (1946), and, in relation to intellectual property, in John F. Duffy, *The Marginal Cost Controversy in Intellectual Property*, 71 U. CHI. L. REV. 37 (2004).

low marginal, cost situations. Given the excess wealth within the system, it is not possible to set all prices equal to marginal costs, which explains why everyone (merchants included) jockeys to shift a larger fraction of the costs of the system to other market players. To see why this is so, think of the familiar problem facing the pub that knows that its success depends upon its ability to attract male and female customers in roughly even proportions. It is a safe assumption that the cost of service for both classes of customers is the same, yet charging the same price to both sexes will not get either men or women their desired mix of goods and services. Men do not want beer and pretzels as such; they want beer and pretzels with women present in sufficient numbers. Uniform charges based only on the cost of the provision of service thus results in too few women relative to men; the men withdraw because there are not enough women, and the few remaining women withdraw because there are too many men. The market shuts down on both sides. The only way to overcome this problem is to charge different prices to each sex based on the intensity of their demand. Men will remain in the market at the higher price, and more women will enter it at the lower price, thus reviving the underlying market.

The credit card market shows the same imbalance, with the merchants as the eager participants and the customers as the reluctant ones. We should therefore, in most markets, see an outcome in which a greater fraction of the common costs are borne by the merchants, which, for open-system cards, are mediated through the interchange fee and, for closed-system cards, are expressed through a lower reimbursement rate for merchants. The RBA's imposition of regulation makes a difference because it leaves the closed-system card companies with more money "in the till" to recruit customers to their payment systems. Since these closed-system cards specialize predominantly in higher-end customers, the observed outcome is predictable—a larger proportion of transactions are funneled through the closed-system cards, representing a still higher proportion of the total dollar amount of credit transactions.

3. Distributional Consequences

The RBA's concern with unregulated interchange fees also has an important distributional component. Let us suppose (contrary to fact, I believe) that credit cards are a less efficient way of doing business. There is an inherent unfairness in asking users of cash or checks to subsidize the use of these instruments. Thus, the critical question is to ask to what extent, even accepting the RBA's basic assumptions, do these cross-subsidies actually exist?

The answer is that these subsidies, while they do exist, tend to be small. The distributional point is usually framed to say that credit card holders as a class receive a subsidy from individuals who use cash or checks. But one mistake in framing the point this way is that it assumes that the various users fall into discrete categories without overlap in class membership. This assumption is most decidedly not valid. Most people use all three systems of payment in different proportions and, hence, appear on both sides of this imaginary line. Assume, in the limit, that all persons, regardless of economic status or consumption patterns, used the six common payment systems—cash, check, debit, charge, and the two kinds of credit—in identical proportions. No one person would receive a subsidy from any other person. Each would benefit in his or her role as a subsidized person, while each would also be harmed in the role of a subsidizing person, but always *in the same proportion.* In this situation, the only concerns hearken back to the efficiency question. If the systems are mispriced relative to each other, then some systems will be overconsumed and everyone would benefit from a shift in prices that captured the relevant cost variables.

The actual situation on the ground is, of course, more muddied. There are some people who do not use credit cards and still others who use them in greater or lesser proportions. It seems that people with both credit and debit cards tend to favor one over the other. Accordingly, it would be necessary to figure out the transfer payments that some consumers on balance receive from others. But the *net* size of those transfers among consumers has to be decidedly

smaller than any reduction in the collection of fees, and, certainly as between credit and debit, many consumers have the easy option to shift, at least in part, to their favored system. Thus, the most recent testimony from Governor Macfarlane calculates the annual savings to non-cardholders from capping interchange fees at $580 million.[32] But of course much of the wealth is simply a transfer from one pocket to another among persons who use all of the different types of payment systems simultaneously. We know that in Australia, consumers use the various instruments—cash, check, credit card, and debit card—in roughly equal proportions. Allowing for differential spending patterns, it seems unlikely that even half of the purported savings represent net transfers to persons who predominantly utilize debit instruments.[33]

There is one final irony in this argument concerning the extent of transfers: there are two ways in which the RBA could eliminate the ostensible transfer payments. One option, which is the RBA's dominant strategy, would be to crack down on the use of credit cards. A more sensible way to achieve this result, however, would be to allow market forces to *expand* the use of credit cards, so that fewer transactions would run through cash and checks. If that route were taken, the extent of the transfer would necessarily shrink as the percentage of credit card payments rose. Ironically, if all transactions took place via credit cards, then interchange fees could not possibly contain an element of cross-subsidy. In light of these considerations, it may well be that the public systems of cash and checks are inferior technologies propped up by the imposition of the RBA's interchange rules.

---

[32] Macfarlane Testimony, *supra* note 7, at 29-32.

[33] Although the calculations would be difficult, it seems likely that most people who record the greatest dollar volume on credit cards also make extensive use of other payment systems. For these people, redistribution is merely from one pocket to another. In contrast, the gain to those consumers who do not use credit cards, or use them sparingly, is likely to be small.

These criticisms do not, of course, apply with respect to debit card systems, which share many of the efficiency features of credit card systems, but are cheaper to operate precisely because they do not have to quantify or respond to credit risk. Indeed, debit cards often have uses that cash and checks do not (e.g., as payment in e-commerce) and, as a result, we may well expect their use to grow. Indeed, one reason for Visa's stronger market position relative to MasterCard was its early and forceful moves into the debit card business—yet another example of how successful firms can survive only by cannibalizing their own high-margin businesses.

What then of the possible cross-subsidies between debit and credit cards, which are often run by the same banks from the same platform? The answer in many ways parallels the cross-subsidy answer given above in relation to cash and checks. Both systems are part of a consensual network in which there is no unique fee structure that prices all transactions at marginal cost. If it turns out that debit card users as a class are unhappy with the association of these cards with credit cards, even when they carry both, then some form of market segmentation will occur. New companies will offer debit-only cards at lower rates than those at which existing debit card users receive the blended product, and the incumbents will have to follow suit—unless the synergy of the two cards has advantages that debit-only plus credit-only cards could not equal.

Nor would it be necessary that the pressure for market segmentation come from customers. It might also come from merchants who wish to negotiate debit card-only deals with their banks at proportionately lower rates. In California, for example, payment at the gasoline pump is by debit card only (a true nuisance!) for precisely this reason. Given this range of responses, it is at best an open question as to how many people using both systems simultaneously will migrate from the current platforms. But as long as that possibility remains—and any third person could start a debit-only network—then regulation is not needed as a form of

industrial policy to force market institutions to march down an intelligent path.

IV. THE NO SURCHARGE RULE

The second major RBA initiative, disallowing surchages, is surely less important than capping interchange fees, but it is nonetheless worthy of comment. The basic objection in this case is that standard contractual prohibitions on surcharges prevent merchants as a whole from adjusting prices for their various classes of users in order to reflect the relative costs that each class imposes on the overall system. As with interchange fees, the potential gain from any such regulation is strictly limited because we know from the outset that the differential is one that does not meet with stern market resistance—which would surely be the case, for example, if the differential cost of service were closer to five percent instead of the one or two percent that we see. In part, there is no resistance because customers often prefer to pay with credit cards than with cash, and these key customers would be angry with the small surcharges that occupy their time and do little else. The customers who pay by check and cash do on occasion ask for and receive certain discounts, particularly on larger purchases, and thus have little bloc interest—assuming that they are a bloc—to override the clause. Thus, the porous nature of this rule accounts for part of its durability.

Another objection is that the argument for a surcharge rests on the assumption that non-credit card users gain no particular advantage because others use credit cards at the same place of business. There is some reason to think that this assumption is overstated. Consider the following thought experiment that could, in principle, become the basis of systematic empirical research. Suppose that you are intent on paying for a product by cash or check and could enter into either of two competitive establishments, one that announces that it will accept credit cards at no additional cost and another that announces that it will not take any cards at all. If one takes the RBA assertions seriously, then the cash purchaser should, *ceteris paribus*, make a beeline

for a place that, on average, will charge about one percent less for each individual purchase.

There are three reasons why the behavior observed will not prove to be all that simple. First, from the outside, the two stores are in some sense black boxes, especially to tourists, who have no knowledge of what types of goods they carry or the clientele that they serve. In these situations, the credit card logo over the door operates as a useful signaling device that tells the potential buyer that this business has sufficient stability and organization to participate in a credit card operation. This type of information is valuable not only to credit card holders, but also to cash purchasers who see it as a general indication of the scope and reliability of the business' full range of operations. There is no direct way for credit card companies to collect for this bonding function, in essence a form of co-branding, from non-credit card users. The imperfect maintenance of the surcharge helps counteract such a free riding effect.

The second reason is more restrictive, and it applies only to those people who are happy to use cash, but carry credit cards as well. When they enter a store, they may well plan on making some modest purchase. However, if they then decide to buy an unexpected and expensive item, they may prefer to switch to a credit card for payment. Thus, they will gravitate to the place that provides them with the added flexibility of paying by credit card. Once again, allowing the credit card company to impose a surcharge is the only way to overcome free riding by cash-only customers.

Finally, there is the standard concern with the ease of transacting, which looms especially large in any setting with a high volume of low-cost transactions. Customers hate to wait in lines, and anything that complicates the checkout process will often result in a reduction in sales, with attendant detriments to both sides. Multiple price schedules are sure to slow down business, even if the delays only come from customers who have to check first to see whether they have sufficient cash in their pockets or funds in their checking or debit accounts. In any individual case, the time

costs look trivial, but their aggregate sum over the course of days and weeks, like that of interchange fees, is not. If multiple price schedules serve to slow down lines by even one or two percent—a conservative estimate—this would translate into higher labor costs and longer waiting times, which might equal or exceed the surcharge differential.

These costs are, of course, apparent to the merchants themselves, which is why in many instances they avoid the adoption of separate price tiers even when given the opportunity to do so. In this regard, it is instructive that the surcharge arrangements found in the Australian context do not support the view that similar arrangements are essential to deal with the implicit cross-subsidy of cash customers. The best data suggests that the percentage of merchants that use surcharges is between 2.3% and 7%, and that, of these merchants, none impose surcharges in all cases.[34] In addition, in some instances merchants impose surcharges on credit transactions that exceed the interchange fee, a strategy that suggests that these surcharges are imposed with a modest eye toward price discrimination. Credit card customers are, on average, likely to have higher demand for goods than are non-credit card customers, and the extra surcharge is a way to single them out. Again, therefore, the no surcharge rule does not promote competition, but instead allows merchants in the Australian markets (which are more concentrated than U.S. markets) to exercise some modest degree of monopoly power.

The general unease with surcharges at the merchant level also provides the best explanation of why issuing banks have not resorted to tacking on variable charges to individual credit transactions, even when interchange fees are capped. The main argument in favor of having issuing banks impose surcharges on their cardholders is that it eliminates the ostensible cross-subsidy that credit card users

[34] *See* Chang, *Australian Interchange*, *supra* note 2, at 19. Note that in the Netherlands only about 10% of transactions involve merchant surcharges. The figure is just 5% in Sweden, and is also small in Great Britain.

receive when the interchange fee is obscured by the uniform prices charged to all purchasers. In such a case, bank-imposed surcharges should create uniform prices for products and services precisely because the merchants need not build interchange fees into their cost structure. Yet that alternative system has never been used in the credit card business, even during its infancy. Credit card systems that were devoid of any semblance of market power used interchange fees in order to help issuing banks to woo credit card customers. This history is consistent with the view that two-sided markets require these types of internal adjustments, wholly without regard to grand considerations of monopoly and market structure. Similar to the way small bars may charge different prices to men and women, small (and large) credit card companies have used the same strategy for predominantly the same efficiency-driven reasons.

Any surcharge imposed by issuing banks on cardholders has two distinct disadvantages. First, it forecloses the possibility of merchant subsidies for cardholders: once the lines of communication are cut, it is impossible to run cost subsidies. Second, this new fee structure creates yet another level of uncertainty and confusion. In principle, a potential purchaser in a store would like to know the total transaction price at the moment that goods or services are purchased. That determination is not possible if there is a subsequent service charge of unknown size on the credit transaction of which the merchant is ignorant. In such a case, the customer has to wait until he receives his bank statement to be sure of the exact charges. If there are multiple transactions, there is a greater likelihood of squabbles over individual bills, especially if the surcharge in question is not denominated as a simple fraction of the total bill. Yet imposing a flat fee reduces the ability to differentiate between different types of payment arrangements adopted by merchants—such as electronic and non-electronic processing. Furthermore, even if the issuing bank instantaneously disclosed the fee to the merchant, it would merely replicate the unpopular system of surcharges at the

merchant level. The argument in favor of the current contract provision is not that the no surcharge rule is perfect. Rather, it is that the set of advantages and disadvantages are so difficult to untangle that these fees are not an ideal target of regulation.

V. THE USES AND LIMITS OF CREDIT CARD REGULATION

In one broad sense, both the interchange and the surcharge regulations are efforts to regulate the payment systems industry. However, the two regulations differ in dramatic ways in their institutional implications. There are no obstacles preventing the implementation of surcharge regulation on a permanent basis. The only change required is to strike surcharge clauses from contracts. There is no need for any continuing oversight or enforcement of the provision, as nature will take its own course in the adjustment of other terms and practices.

Setting the interchange fees, however, is an entirely different proposition that raises all of the nasty complexities associated with ratemaking procedures that have traditionally been invoked to deal with natural monopolies. The first step of the analysis is to determine those categories of expenses that should be included in the rate base. The RBA, in its initial findings, took the hard line that the only expenses covered were those that related to either the processing of transactions or the prevention of fraud. The cost of an "interest-free period," which it makes sense to include under a sound analysis of two-sided markets, was excluded from the base on the grounds that it "is a matter exclusively between individual card issuers and their customers."[35] The same arguments apply to "loyalty programs" that issuing banks use to attract and keep customers[36] and, more dubiously, to various fees for starting

[35] RBA, AUSTRALIA REFORM, *supra* note 3, at 49.

[36] *Id.* at 49-50.

and maintaining a network, even though these relationships have no exclusivity features.[37]

In effect, the RBA's solution fixed the interchange fee and thus eliminated it as a source of competition between the various payment systems. At that point, the operating costs had to come out of bank profits or cardholder fees, whether on a per transaction or annual fee basis. The interchange fee had to account for the differential demands found in two-sided markets in order for the merchants to capture the full range of the surplus. The RBA thought that further adjustments would take place on both sides of the markets. Merchants would have some incentive to pass their lower costs through to consumers in the form of lower prices. Likewise, issuing banks would have an incentive to raise annual fees and transaction fees to cover the shortfall in revenue. However, dollar-for-dollar passthroughs are unlikely on both sides of the market, in part because of the high concentration in both the retail and banking sectors of the Australian system. In fact, the Chang, Evans, and Swartz study estimated that about one half of the merchant savings could be expected to be passed through to consumers, if—and it is a large if—the rigidities of the price system do not overwhelm the small sums involved.[38] There is no systematic guarantee that when all of these changes are netted out, one could point to any net allocative improvement in the overall operation of the payment system.

Choosing this limited rate base is only the first step in the overall process. The RBA is also required to set the actual rates, and that determination gives rise, in turn, to two difficult questions. The first involves the *duration* of the rate decision. Presumably, the 0.55% fee chosen represented a good initial estimate of the average cost of running the interchange system. Such a presumption is accorded notwithstanding the fact that the RBA has offered no clear

---

[37] *Id.* at 51-52.

[38] *See* Chang, *Australian Interchange*, *supra* note 2, at 18 (noting that a 50% reduction in a 0.21% merchant saving works out to 0.105%, which on a $40 transaction is about $0.04, an amount easily lost).

articulation of why it chose this number, much less why, given the differences in customer mix and service systems, this number should be uniform across all payment systems. But with the rapid movements in technology, or changes in the mix of borrowers, nothing guarantees that the system's two key cost drivers will remain constant over time. For example, there is a real risk that an increase in credit card fraud could require additional infrastructure security measures that would exceed the allowable costs, especially expenditures that would be incurred today but would have to be amortized over future years. By the same token, greater efficiencies in processing transactions could, depending on scale effects, reduce the costs of running this system. The balance is not possible to determine in *a priori* fashion.

These uncertainties suggest that some rate adjustments, either up or down, might be appropriate over time. In figuring out what these adjustments might need to be, however, the regulator has to guard against the danger that regulated firms (or some more than others) will pad their expenses in order to increase their net profit under regulation. The major tradeoff in regulation is between the size of the rate base and the rate of return. When the base includes all investment expenditures, regardless of their prudence, the rate of return should be lower because the firm bears less of the economic risk than when such expenditures are excluded. However, that relationship does not normally supply sufficient information regarding how either of these tasks should be undertaken. It is useful to exclude certain combinations, but no ironclad test points to a single unique solution.

These problems only arise in regulated markets. No firm in an unregulated market has the same type of perverse incentive to control the costs of its operations. Instead, in an open market, a firm will work to reduce its costs in order to increase its profits. Yet, by the same token, the risk of confiscation inherent in any ratemaking scheme has generated in the American setting a substantial body of

"takings law."[39] The danger here is that the regulation will be so restrictive that the regulated firm will not be able to recover its initial investment or make a reasonable return on it. There is, therefore, no "safe" direction in which the regulator can move. Too little regulation yields monopoly profits; too much results in confiscation. An unregulated market does not pose these difficulties.

A second problem is parasitic on the first. At a first approximation, the operation of one credit card payment system looks more or less similar to another, but the technology chosen or the book of customer business could easily vary across firms. The cost structures of two or more competitive firms could, therefore, differ in small ways that add up in a high-volume business. In the absence of regulation, each firm is saddled with its own cost structure. Accordingly, there is little doubt that if system *A* offered lower interchange fees than system *B*, we should expect merchants, *ceteris paribus*, to gravitate toward the system with the lower cost—either by taking only one card or by steering their customers to the preferred card when both are accepted. If such migration does not take place, then the best explanation for that stable equilibrium is that each company is serving the clientele in its own niche in the best possible way. In the end, all card companies have an incentive to reduce interchange costs to maximize profits. There is no obvious need or way for a credit card company to act opportunistically in an unregulated system.

The incentives under a system of regulation are much less clear. The first question is whether the rate structure should allow different interchange fees for payment systems that bear different costs. (Of course, in unregulated markets, the inefficient firm just fails to recover its full costs.) This issue

[39] The courts have long been divided over the best approach for preventing monopoly profits without allowing confiscation. For the most recent Supreme Court decision on the matter, see Duquesne Light Co. v. Barasch, 488 U.S. 299 (1989). *See also* Jersey Cent. Power & Light Co. v. FERC, 810 F.2d 1168 (D.C. Cir. 1987). For earlier Supreme Court landmarks in this area, see Smyth v. Ames, 171 U.S. 361 (1898) and Fed. Power Comm'n v. Hope Natural Gas, 320 U.S. 591 (1944).

arose in the last round of the Australian discussion, when the representatives of both Visa and MasterCard noted that there is no good solution to regulating rival schemes with different cost structures.[40] Any decision to use a "blended rate" in setting the interchange fee would effectively serve to distort the incentives of both firms. The firm with higher costs will receive a rate that will not quite cover its expenses, even if the costs it incurs are justified. The firm with the lower rate will receive an extra bonus that bears no relationship to its own costs. Yet, adopting a system that allows the interchange fee to rise in response to higher costs does nothing to control the operational efficiency of the system or to rein in any component of monopoly pricing. In the end, therefore, no form of interchange regulation is likely to generate efficiency gains. Cost-based regulation forces the firm that cuts costs to lose revenue if rates are calculated with reference to some fixed rate of return. The regulated firm would do better by raising costs, not by cutting them. In unregulated markets, regardless of structure, the opposite is true. The firm that lowers costs makes more profits because it can internalize some portion of the relevant gain. Those forces remain in the interchange fee market.

On balance, therefore, the overall situation is as follows. The RBA's basic perception is that the high concentration in the credit card industry suggests that the interchange fees observed are higher than those that could be sustained in a competitive market. If this is so, then the rate regulation does not look like some special intervention tailored to interchange relationships, but rather like typical rate-of-return regulation on a natural monopoly thought to be a product of the new technological age. Regulation is used because a competitive equilibrium is not attainable in this network industry. The regulatory initiative, therefore, is best explained as an effort to offset some perceived monopolistic imbalance stemming from the small number of independent networks, but that is not the explicit rationale

[40] *See* Clapham Letter, *supra* note 29, at 1-21.

of the RBA.[41] At this point, however, the RBA initiative runs into the same objection that was raised many years ago by Harold Demsetz in his classic article, *Why Regulate Utilities?*: the public expenses incurred will not achieve the static competitive solution.[42] At the same time, the rate structure operates as a systematic drag on long-term innovation by robbing the innovative firm of the additional revenues it otherwise deserves.[43] The criticisms of the RBA interchange regulations, therefore, must be tempered because the credit card industry does not operate in a perfectly competitive market. As a result, some regulation in principle *might* create some overall improvement. Nevertheless, one might wonder whether major regulatory initiatives ought to rest solely on that kind of existential theorem in light of the formidable obstacles that stand in their path. The reforms that are necessary to correct one admitted market imperfection introduce other costs, errors, and distortions in their stead, which, in all likelihood, are greater than those costs that regulation has eliminated. The proper institutional response should take its cue from the Hippocratic Oath: *primum non nocere*. First, do not harm.

VI. ANTITRUST IMPLICATIONS

In sum, the RBA's well-intentioned efforts to counteract imperfections in the interchange market through regulation may be worse than the disease itself. It does not follow, however, that administrative actions are the wrong way to regulate *if* some state intervention is required to address some perceived monopolistic (or at least oligopolistic) power inherent in these credit card networks. If some public response is indeed required, then the administrative approach, for all its failings, might be preferable to its most common rival—the enforcement of antitrust laws through private rights of action, as allowed in the United States

[41] *Id.*

[42] Harold Demsetz, *Why Regulate Utilities?*, 11 J. LAW & ECON. 55 (1968).

[43] *Id.*

under the Clayton Act, for damages resulting from another firm's improper business practices.[44]

In the antitrust setting, as in the regulatory arena, there is good reason to be cautious. Even under the best of circumstances, it is difficult to design an ideal antitrust law. The strongest case for intervention on that score is presented by "hardcore" cartels in ordinary businesses that raise prices, reduce quantity, and precipitate overall social loss.[45] Yet, even a straightforward question, such as whether to allow a merger of two firms, is fraught with difficulty, for it is not easy to trade off the deadweight losses that result from the extension of monopoly power (if any) against the efficiency gains that result from the change in operations. Once one moves beyond these cases, the efficiency/restraint tradeoffs are as hard, if not harder, to measure in the judicial setting as they are in the administrative one.

In dealing with interchange fees, the administrative law approach has major advantages. At the outset, combining efforts of the various banks to work out one unified payment system has huge social advantages that could not be achieved by requiring each separate bank to set up a standalone credit card system for its own customers, like the one Bank of America ("BOA") originally tried.[46] BOA quickly refocused its effort when it realized that some collaborative effort was necessary. Moreover, imposing remedies through professional bodies improves the chances of finding a decisionmaker who is cognizant of the costs and benefits of alternative institutional arrangements. Most critically, an administrative body is less likely to adopt an extreme position than a runaway jury that operates free of all long-term institutional restraints.

---

[44] Clayton Act § 4, 15 U.S.C. § 15 (1992) ("any person who shall be injured in his business or property by reason of anything forbidden in the antitrust laws may sue therefore in any district court of the United States").

[45] For a recent discussion, see HERBERT HOVENKAMP, THE ANTITRUST ENTERPRISE: PRINCIPLE AND EXECUTION 26-28 (Harvard Univ. Press 2005).

[46] EVANS & SCHMALENSEE, *supra* note 4, at 61-67.

This last point is of exceptional importance because the entire edifice of antitrust law works best when independent firms, facing constant marginal costs, conspire together to raise prices above the competitive level, often by secret agreements. In those cases, the harsh penalties offset the risk of nondetection and respond to the imposition of a clear social loss. But why bring the antitrust laws to bear against any network industry whose internal structure is transparent to all and whose configuration is driven heavily by efficiency considerations? It is well understood that the member banks in any credit card network compete vigorously to sign up cardholders or merchants. Must they also compete over the interchange fees that are passed through to merchants? The basic premise of the four-party, open-system arrangement is that any merchant can take a branded card with comfort, even without knowledge of which bank has issued it. The uniform fee means that the merchant knows its costs, without having to learn which of hundreds of possible banks issued the card in question. The standard fee thus allows for seamless completion of high-volume, low-margin transactions, which would stop dead in their tracks if subjected to any requirement of individuated negotiation. The competition comes as the issuing banks seek to sign up customers, and as the credit card intermediates attempt to adjust their interchange fees to cover costs and provide the ideal balance between merchant and user fees.

In this institutional setting, antitrust law is of little or no use. In and of itself, it cannot be used to set rates, especially those that are subject to variation over time, like interchange fees. Its role, therefore, is much the same as in all industries that have some rate regulation component. It serves as a backstop to prevent illicit collusion above and beyond that needed to keep any network intact. To use a familiar example, the 1996 Telecommunications Act imposes extensive regulations on the setting of interconnection fees

throughout the entire telecommunications network.[47] The 1996 Act also contains an explicit provision preserving the applicability of antitrust laws. Recently, however, the Supreme Court in *Verizon v. Trinko*[48] took the correct position that the extensive FCC regulation of the telecommunications industry precluded the use of antitrust laws to second guess decisions made in the regulatory arena, especially at the instance of a plaintiff (Trinko) who did not purchase its services directly from the regulated party (Verizon).

The interchange fee regulation differs from the telecommunications cases, at least in the United States, insofar as there is no system that caps these fees like the Australian model. Even so, if some effort to deal with any concentration issue (of less concern in the United States than in the more concentrated Australian market) is desired, administrative regulation offers the proper forum in which to raise those issues, as it does in all ratemaking contexts. Given the strong efficiency properties of the current arrangements, any breakup of the network would be wholly misguided because of the massive dislocations that it would create. Furthermore, a huge treble damages award (over a nonexistent baseline) could easily gobble up a company's entire revenue. For example, trebling the 0.40% reduction ordered by the RBA would equal 1.20%, which is higher than the original interchange fee. In short, so long as some interchange fees are necessary to keep the network functioning, antitrust laws should not apply in the absence of price collusion between the different networks.

The critique of the use of antitrust law in this arena must also consider the fundamental difference between the four-party and three-party systems. MasterCard and Visa, out of necessity, operate through combined action which is *justified* in light of the network coordination issues that payment

[47] Telecommunications Act of 1996, Pub. L. No. 104-104, 110 Stat. 56 (1996) (codified as amended in scattered sections of 47 U.S.C.).

[48] Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 412-14 (2004).

systems are designed to resolve. Nevertheless, they are always faced with the charge that they operate as a combination in restraint of trade notwithstanding the manifest need for a uniform interchange fee. Firms like Amex and Diners Club, which operate standalone three-party systems, can set their own fees unilaterally, thereby escaping the charge of collusion, regardless of their (limited) market power in a concentrated payment card market. In general, it creates major distortions to subject two competing modes of business organizations to radically different antitrust regimes. Right now, MasterCard is busy turning itself into a proprietary organization, in part to minimize the potential application of the antitrust laws.[49] Other reasons may also account for the shift, such as an improved ability to raise capital. But even so, potential exposure to the antitrust laws should not place a thumb on the scale, which in some situations at least will motivate parties to adopt inefficient business forms. If some form of intervention is required, then a uniform regime of interchange fees, which may allow for some interfirm variation in actual rates, is far superior to subjecting the market to antitrust laws that are unable to do the fine tuning necessary to undertake so delicate an enterprise.

VII. CONCLUSION

This examination of the RBA's proposal demonstrates how difficult it is to make and implement government policies that improve upon the outcome of those arrangements brought into operation by the ingenious application of the principles of voluntary association. The development of the payment card system owes nothing to government support or government regulation. The system, as it emerged, developed from private networks which, at

[49] Robin Sidel, *MasterCard Aims to Shift Ownership*, WALL ST. J., Sept. 1, 2005, at C3 ("Several people also said that the partial sale of the card association could help shield MasterCard from a series of recent antitrust lawsuits over fees that merchants pay to accept and process cards.").

their inception, appeared to be of limited scope and utility. But the growth of networks that link individuals together can take place in a gradual way. Individuals who have no information about how networks are organized can sign on with a bank. The bank then will enter into transactions with other financial institutions in ways that promote their mutual gain. Within the past fifty years, the transformation in payment systems has taken place along multiple dimensions, so much so that it is easy to lose track of the gains that these innovations have wrought. But in thinking back on my own life, I recall that in 1973 my wife and I used a mix of American Express travelers' checks—We didn't leave home without them!—and a letter of credit to pay for a month's holiday in Europe. But that was in an age before large credit card networks, ATMs, online banking, immediate credit verification, and a thousand other conveniences, large and small, became universally available.

The processes that generate these major gains always look chaotic, and their results often leave short-term losers that call out for further government regulation. Right now, in the wake of Hurricane Katrina, gasoline dealers have mounted a concerted protest against the rise in (fixed) interchange fees that have cut into their profit margins with the spike in gasoline prices.[50] That being the case, there is no need for piecemeal intervention, especially when these same firms will see their interchange fees sink as prices work their way back to lower levels. The self-help remedy of taking only cash or debit cards—checks won't cut it in this market—is far preferable to another round of litigation or regulation that could be replicated in any other submarket that experiences substantial price variation.

As this timely example illustrates, the case for state intervention is not made out by a showing that someone can devise a hypothetical system of fee restraints that is said to respond to some higher test of rationality. All working systems contain imperfections that are difficult to defend in

[50] *See* Margaret Webb Pressler, *Card Companies Are Filling Up At the Station*, WASH. POST, Sept. 25, 2005, at F01.

some nonexistent first-best world. The basic social principle, however, overlooks these blemishes and starts from a very different premise: don't mess with success. The rapid expansion of payment systems has exceeded the wildest expectations possessed by anyone even a decade ago. This expansion will not continue into the next decade if state regulation in Australia or anywhere else expands its hold over a complex system that has succeeded thus far without the guidance of an all too visible hand.