6. Economic perspectives on exhaustion and parallel imports

*Keith E. Maskus*

I. INTRODUCTION

Parallel imports (PI) are legitimately sourced goods brought into a country, without the authorization of the entity owning some form of intellectual property rights (IPRs) to those goods, after they have been placed into circulation in another nation. Whether such trade is legally permitted depends on the geographic scope within which the IPR, whether patent, copyright, trademark or other exclusive right, is exhausted upon first sale. Typically when a commodity is first sold within a country the IPR holder cannot prevent its owner from reselling it within that market. The exhaustion doctrine governs the ability of such goods to be imported legally into another market. Some countries give the IPR owner in the target market the right to exclude such goods sold abroad from importation, a situation referred to as national exhaustion. The United States (U.S.) adopts this policy in most areas. Others permit parallel imports, which are called that because by definition they occur within a parallel distribution channel outside the control of the originator. This is the case of international exhaustion, which may be found for certain types of IPR in Brazil, New Zealand, and elsewhere. The European Union (EU) and other members of the European Economic Area (EEA) pursue an intermediate policy of regional exhaustion, under which IPR owners may exclude PI from outside the EU geographical area but such trade is fully legal inside the region.

Under the terms of Article 6 of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) at the World Trade Organization (WTO), member nations are given full latitude to select their own exhaustion regimes. It is not surprising that a global rule could not be reached, given the complexity of the subject. There are marked differences across markets in the permissibility of this trade. Moreover, for any country the law may vary across types of IPRs. For example, Japan traditionally followed national exhaustion in patents until recent court decisions opened the market in certain circumstances.

* Professor of Economics, Department of Economics, University of Colorado Boulder.
However, it is open to parallel imports in copyrighted goods except motion pictures. Australia has a mixed regime, being open to PI in trademarks but closed in patented goods (with limitations) and in copyrighted products, except for compact disks and books.

As might be surmised from this variation, decisions regarding legalized PI are complex and often arise from judicial actions, making it impossible to characterize them in a simple yet accurate way across countries. To a first approximation, however, the primary economic determinant of the regime choice is the extent to which a country has a comparative advantage in producing intellectual property of various types. For example, the U.S. is a major producer and net exporter of patented technologies and copyrighted goods. Its policy permits exclusion of patented and copyrighted items on the principle that the rights to exploit such products should be fully exclusive throughout the term of protection. Similarly, the bulk of the high-income economies in the Organization for Economic Cooperation and Development (OECD) favor limits on international exhaustion in patents and copyrights, though there are many exceptions. In contrast, India is generally open to PI because its domestic firms register relatively few patents, meaning that parallel trade competes largely with foreign-owned IPR. However, India bans PI in copyrights, reflecting the importance of its software and cinematic industries. More generally, poorer countries are more open to PI in order to promote import competition with IPR-protected goods, though again there are many detailed exceptions.

In economic terms, parallel imports are a form of cross-border arbitrage. As will be discussed below, owners of a product patented or copyrighted in multiple locations profit from setting market-specific launch dates and prices, which vary according to country characteristics. It follows immediately that banning PI is a means of sustaining this market segmentation, while permitting such trade should reduce or eliminate it. Indeed, to the great majority of economists specializing in international trade it seems intuitively obvious that PI, by generating another source of competition in high-priced markets, must be pro-competitive. That intuition carries over to official policies in the EU, where parallel trade within the region is seen as a key means of integrating the single market and generating consumer benefits.

While this idea is appealing and correct in its simplest form, it is misleading in important ways. Such is the basic conclusion of serious economic analysis of the sources and impacts of PI. This analysis is surprisingly recent, given that exhaustion has been an element of IPR policy for a long time. Growing interest in the subject may be attributed largely to three closely interrelated factors. First, the entire subject of intellectual property rights as a matter of trade policy was ignored by international economists until after

---

6 The fact that this exclusive right is limited in trademarks stems from their different legal foundations: so long as consumers are not confused regarding the true origin of goods, PI should be permitted if the mark owner placed them into circulation abroad with a related party.
7 Ganslandt and Maskus, *supra* note 5, at 267.
8 Maskus, *supra* note 1, at 1274.
TRIPS. Now it is a mainstream area of trade analysis as economists have come to see its importance for innovation, technology transfer, and market power. In this context, PI form perhaps the most interesting element of international marketing strategies affected by regulatory policy.11

Second, the intense globalization associated with the spread of multinational enterprises and international production chains, largely driven by technological advantages combined with factor-cost differences, have roused the interests of trade analysts in all forms of international technology and cultural markets.12 Whether PI are permitted matters for many strategic decisions, ranging from first-stage investments in research and development (R&D) to the timing of market entry.

Finally, the subject of price differences across markets has taken on direct urgency as a matter of development policy. Consider, for example, the issue of “tiered pricing” in essential medicines, or the idea that segmented markets should support lower prices of patented drugs in poor countries, thereby raising market access.13 While a powerful prescription in theory, the idea does not seem to hold much water in practice, raising basic questions about the efficacy of bans on parallel trade as a means of beneficial market segmentation.

Because space is limited in a single chapter, it will not be possible to do justice to what is now a deep and important economics literature. Rather, I will focus on three key questions that have been the subject of close study and highlight the central lessons emerging from that work. After a brief review of basic facts the subsequent sections cover these questions, which may be introduced as follows. First, what do our core economic models predict about the impacts of parallel trade on competition and prices under different circumstances? What little empirical evidence there is in this regard is also discussed. Second, how does the permission of PI affect dynamic decisions regarding R&D investment and product entry, and what is the role of national price regulation in this context? Third, are there important linkages between trade policy and the choice of exhaustion regime, at least in theory?

As we will see, there is a great deal of ambiguity about the potential competitive and welfare impacts of parallel trade. For example, economic theory shows that it is quite possible for IPR owning firms to profit from the existence of PI when they are subject to rigorous price controls. Moreover, there are realistic circumstances in which openness to PI may actually cause prices to diverge across markets when distributors act strategically to maximize profits. Thus, exhaustion and parallel trade are far more complex than simple arbitrage theory might suggest and subtleties of this kind should be kept in mind by policy-makers and legal scholars.

12 Ibid. at 35–81, 294–312.
13 Ibid. at 258–70.
II. BASIC FACTS AND EVIDENCE ABOUT PARALLEL IMPORTS

The most frustrating feature of parallel imports for researchers is that data about them are extremely scarce and hard to locate. The reason is pragmatic: because PI are legal and entail trade in legitimate (that is, not counterfeited) goods they attract no particular attention from customs authorities despite the fact that they are traded in parallel distribution channels. Distributors of PI pay any required tariffs but there is no customs-related purpose to collect figures on parallel trade flows. The primary exception, which supports some analysis discussed below, is that public health authorities must track the origins of medicines and other goods that may affect public health. Thus, customs may be asked to identify certain import transactions as parallel imports. However, such data are almost always confidential and unavailable to researchers. Thus, there are no systematic databases on PI by commodity, source, or destination.

A. Stylized Facts

Despite this lacuna, occasional surveys have been undertaken that help characterize parallel imports. The most comprehensive one, now quite dated, was performed by National Economic Research Associates (NERA) at the request of the European Commission. The survey found that intra-EU PI accounted for different shares of sales in ten industries in 1993, though there is reason to suspect the figures were under-reported by respondents.14 PI were greatest in compact disks (10–20 percent of EU sales), cosmetics and perfumes (up to 13 percent), and soft drinks (up to 15 percent). Each of these types of commodities reflect characteristics giving rise to the potential for PI. Compact disks are copyrighted goods subject to substantial increasing returns, fashion goods are subject to heavy marketing (and potential free riding), and soft drinks intensively use trademarks. Other sectors with significant intra-EU parallel trade included automobiles, consumer electronics, clothing, and confectionery. Thus, one stylized fact is that PI can account for significant shares of trade in IPR-intensive goods within regions where they are permitted.

The NERA survey pointed out another important stylized fact, which is that the great bulk of PI are undertaken by organized parallel trading firms, who operate at the distributor level.15 A good example is soft drinks, which are far likelier to be traded in bulk form as syrup than as final packaged goods. These PI firms seek stable sources of supply among authorized wholesalers in lower-priced countries for resale to retailers in higher-priced locations. Stability and volume are important, for parallel-trading firms need to be viewed as reliable sources of legitimate products for retailers to be willing to enter into delivery contracts with them. In contrast, despite the usual depiction of parallel trade as the practice of individuals buying a car or a watch in a cheaper location for use at home,

---

15 Ibid. at 32.
individual-level and even retail-level PI is rare. The basic reason is simple: engaging in international trade requires payment of significant fixed costs for firms, in terms of finding suppliers, meeting regulatory requirements in import markets, and dealing with customs and taxes. These costs need to be spread over large volumes, implying that PI is largely done in bulk at the wholesale level.

A third important fact is that PI volumes are sensitive to international price changes. For example, an early survey found that substantial increases in such imports into the U.S. emerged after large increases in the international exchange value of the dollar. Parallel trade also seems sensitive to variations in exchange rates within East Asia. On the other hand, available evidence in the European Union, where this trade is legal among Member States, suggests that parallel imports are not sufficient to push price levels together as a result of arbitrage. This is a loose claim, for it relies on the observation that price convergence of commodities across European cities is considerably less than one might expect within a single market connected by zero tariffs and the threat of PI. This claim has not yet been subjected to serious statistical analysis, however.

**B. Empirical Evidence**

Given the data scarcity in this area, there are no rigorous statistical studies of the extent and effects of PI across industries, a real shortcoming in the literature. However, some progress has been made specifically in pharmaceuticals where, as noted above, health authorities must keep track of the sources of traded medicines. Ganslandt and Maskus offered the first statistical analysis of actual data on parallel imports and prices of original manufacturers’ patented pharmaceuticals. They gained access to detailed and high-frequency data on prescription-drug sales in Sweden from 1994 to 1998, broken down at the product level into brand-name, original manufacturers’ sales, and chemically identical products brought in through parallel channels from other markets in the EU. Sweden was an ideal natural experiment for this study because it had to switch from national exhaustion to regional exhaustion when it joined the EU in 1995. By mid-1996, after a period of eighteen months during which PI firms gained certification as suppliers of legitimate versions of patented drugs, PI began entering Sweden. By 1998 such sales comprised up to two-thirds of the market for large-volume drugs. In this regard their analysis unearthed a fifth stylized fact, which is that distributor-level parallel trade

---

17 NERA, *supra* note 14, at 68.
21 Ganslandt and Maskus, *supra* note 16.
is far more common in products with large markets and significant international price disparities.

In their analysis the authors specified a model in which changes over time in the wholesale prices of patented drugs were affected by entry of PI firms, controlling for various determinants of pricing. The authors controlled for the obvious potential reverse causality, in which PI entry itself would vary with market conditions. In their preferred estimates they found that original manufacturers’ prices are pushed downward by competition from parallel imports, with the reduction getting stronger as more PI firms entered. Overall, the pricing power of brand-name firms was reduced by 12 to 18 percent on average due to PI competition, suggesting that this form of competition exerts a strong price-modulating impact.

Did these price effects translate into significant savings for patients and public-health budgets? In an extended paper, Ganslandt and Maskus argued that the actual cost savings were small in Sweden because the wholesale price reductions were not passed on to hospitals and patients at the retail level, rather retailers and PI firms made larger margins. This finding of a limited beneficial impact of PI on final patient costs has been found in other studies as well. For example, Kanavos and Costa-i-Font, again using detailed data, found that the gains from PI accrue largely to trading firms, which price their versions marginally below the often controlled prices in destination countries. In a related survey of six high-priced pharmaceutical markets in the EU these and other authors found that direct savings from PI totaled around 45 million Euros in 2002, ranging from 0.3 percent of the market in Norway to 2.2 percent in Denmark, though various clawback regulations raised these benefits somewhat. In contrast, the profit margins enjoyed by PI firms ranged from 46 to 60 percent, suggesting that parallel trade is lucrative, if not that effective in reducing actual patient costs, while it poses little competitive threat to incumbents. Indeed, this pessimistic finding might be considered yet a sixth stylized fact about PI, at least in medicines.

While this insight is suggestive, there are simply too few serious studies to support any confident conclusions about the impacts of PI on prices and competition in drugs, much less in other goods. Much more research should be done to develop a full picture

---

23 Ibid. at 1049–54.
27 However, another survey claimed that direct cost savings in five of these countries in 2002 amounted to 635 million euros, a significantly higher figure. This finding is questionable because the authors appear to have ascribed the entire price gap between retail prices in destination countries and distribution prices in source countries to savings from PI, a highly unlikely assumption. See Peter West & James Mahon, Benefits to Payers and Patients from Parallel Trade 67–69 (2003).
Research handbook on intellectual property exhaustion and parallel imports

of what drives parallel imports and the impacts they have on important variables. For example, there is no econometric study available that considers the potential effects of PI on decisions of IPR-owning firms to invest in R&D. This is a fundamental question, reflected in the fact that U.S. research-based pharmaceutical firms oppose opening the U.S. border to re-importation of patented and brand-name drugs from Canada, partly on the grounds that resulting lost profits would reduce their ability to develop new medicines. This ignorance represents a considerable gap in the literature on parallel trade. Policymakers concerned with PI and competition in IPR-protected markets might therefore consider directing their statistical authorities to collect more data in order to lift this veil of uncertainty.

III. INTERNATIONAL PRICE DISCRIMINATION, ARBITRAGE, AND EXHAUSTION POLICY

In this section I return to the basic economic theory of parallel trade as international commodity arbitrage and discuss its implications for how countries might choose their exhaustion policies. Again, PI arise in principle as a form of competitive pressure against firms having market power as a result of owning a territorial intellectual property right, or a basket of rights. Firms with market power maximize profits by choosing prices that reflect the economic valuation of specific consumers, thereby extracting higher amounts of surplus. Put differently, firms set higher prices where demand for a good is inelastic and lower prices where demand is elastic.

Price discrimination is ubiquitous in today’s markets, ranging from different airline ticket prices to varying discounts for purchasing electronic goods and medicines. To sustain it requires that the scope for arbitrage among consumers or markets is limited. If there were fully frictionless arbitrage among consumers facing different prices firms would be forced to set uniform prices. At the international level, frictionless arbitrage would refer to parallel trade that faces no trade taxes, transport costs, supply restrictions, or regulatory costs that would support higher prices in some markets than others. Of course, one such regulation would be national exhaustion, which effectively segments markets by barring PI.

It is useful to start this discussion by defining specific forms of price discrimination used in the economics and marketing literatures. First-degree (or “perfect”) price discrimination refers to the ability to set different prices for individual consumers arrayed along a common demand curve. Second-degree price discrimination occurs where consumers along a demand curve may be arranged into single-price blocks. Neither concept relies on a specific geographical dimension. These forms are best illustrated by the varying ticket prices paid by passengers for the same service on an airplane flight or by age-related discounts. In contrast, third-degree price discrimination refers to a situation in which groups of consumers have different demand characteristics, typically across locations,

28 Maskus, Private Rights, supra note 11, at 185–88.
29 These concepts are more precisely defined in Frederic M. Scherer, Industrial Market Structure and Economic Performance 315–17 (2d ed. 1980).
and a single price is charged in each market. The markets are separated by some kind of trade barrier or transport costs permitting oligopolistic firms to engage in “pricing to market.”30 This is the primary concept analyzed in the exhaustion context, with national exhaustion supporting such price segmentation across countries.31

Basic theory shows that profits of IPR owners are generally higher the greater the scope for third-degree price discrimination in international markets.32 Thus, we would expect original manufacturers owning the relevant IPR to prefer bans on parallel trade, as they generally do. In turn, if profitability is an important factor in setting policy we should expect IPR-generating nations to erect restrictions against PI, as noted earlier. This observation surely animates the U.S. position that permitting firms to block parallel trade constitutes a natural extension of their rights to control distribution. However, profits are not synonymous with economic welfare, which depends also on competition and price, with more competition generating greater consumer gains. In fact, the welfare results of price discrimination, and therefore of PI, are complex and the net effects on any particular country are theoretically ambiguous.33

To appreciate this ambiguity at its simplest level, consider the effects of price discrimination across markets on global well-being. Differential prices distort global consumption decisions in the sense that they frustrate consumers in higher-priced locations who would buy the good if it were available through trade at a lower charge. At the same time, total global consumption could be higher with differentiated prices as additional consumers in more price-sensitive markets can afford to buy the good. In pragmatic terms, market segmentation offers incentives to originator firms to sell their goods at lower prices in poorer markets, where demand is likely to be more elastic. If parallel trade shifted products from these markets to higher-priced ones, consumers in the source countries would be harmed. Indeed, even the threat of PI could deter firms from engaging in otherwise beneficial price segmentation.

In this context, it is important to note that where parallel trade is legal we may not see any actual trade arise. The reason is that the competitive threat of PI would induce originator firms to set uniform prices, or more accurately prices that differ across borders only by the costs of organizing and shipping goods, leading to no net trade. It is ironic that legalized PI would actually generate little or no actual trade, an important observation for scholars seeking to understand the phenomenon.

The basic condition under which third-degree price discrimination reduces total global welfare is that total consumption would be at least as large without segmentation, that is, under a globally uniform price.34 Put differently, price discrimination can increase total welfare if it raises total world consumption by serving higher-elasticity consumer groups.

31 Ganslandt and Maskus, supra note 5, at 274.
32 Krugman, supra note 30, at 16.
33 Scherer, supra note 29, at 317 regarding a simple welfare analysis and Ganslandt and Maskus, supra note 5, at 274–75 concerning parallel imports.
and markets.\textsuperscript{35} It should be noted that this primary result is static and does not consider any potentially negative effects of uniform pricing (under PI) on R&D and growth. Moreover, it depends on the ability of PI to equalize prices at low costs. If arbitrage wastes resources, as noted in the next section, PI can be harmful.

More broadly, price discrimination may encourage firms to open new markets in lower-income countries, which they would be willing to do if the price they command exceeds the marginal costs of supplying those locations. If they were forced to charge an equal international price it may not be worthwhile to serve such markets. Thus, in this context market segmentation may be strictly preferred to market integration through PI, because the new markets raise both consumer benefits and profits, without damaging consumers in existing markets. Note further that where there are multiple markets, and therefore multiple potential prices and entry decisions, the possibility that uniform pricing could restrain welfare is yet higher.

This complex nature of the exhaustion doctrine was captured well in a canonical model by Malueg and Schwartz.\textsuperscript{36} They present a theory with many countries and compare three potential regimes regarding exhaustion: price discrimination (segmented markets), a uniform price (integrated markets), and a mixed situation with identical prices in subsets of markets but potential price discrimination between different groups of markets. In their model total output does not depend on the degree of price discrimination as long as all countries are served, meaning that price discrimination is only beneficial if it opens new markets. This implies that demand conditions must be quite different across countries for price discrimination to increase world welfare. They show that a mixed regime, with uniform prices within groups of markets with similar demand patterns and price differences across country groupings with dramatically different patterns, raises global economic well-being compared to either a uniform price or price discrimination across all countries. This result suggests that it could be globally optimal to have regional IPR exhaustion within groups of countries with similar demand characteristics but to permit price segmentation across groups. This insight offers support for the EU’s policy of regional exhaustion, while pointing out the potential importance for similar developing countries to consider permitting PI among them, perhaps in a preferential trading arrangement.

Such observations motivated Richardson\textsuperscript{37} to ask a fundamental question about global exhaustion policies. He set out a multiple-country model in which each country engages in Nash bargaining to maximize its own welfare, assuming the policies of other nations to remain fixed. His basic question was whether countries of the world, characterized by different demand conditions and comparative advantage in IP-sensitive goods, could jointly agree either on total market segmentation (national exhaustion in all) or globally uniform prices (international exhaustion in all). His finding was striking: neither extreme


\textsuperscript{37} Martin Richardson, \textit{An Elementary Proposition Concerning Parallel Imports}, 56 J. Int’l Econ. 233 (2002).
could be sustained as an equilibrium outcome in bargaining among disparate nations. This finding is significant for it explains why TRIPS negotiators could not agree on an international exhaustion rule. Rather, Article 6 of TRIPS leaves the choice up to each individual country or region, which is eminently sensible in economic terms.  

IV. CURIOS ECONOMICS OF PARALLEL IMPORTS

The analysis to this point views PI as simple and low-cost arbitrage against international price differences. While generating important insights this approach is misleading for it assumes that originator firms simply place their goods into circulation and do not react when there is a change in the exhaustion regime, other than to set prices that differ by the trade-cost margin when PI are permitted. Surely, however, IPR-owning firms, which by definition embody market power, would adopt different distribution and pricing strategies in a situation where PI are banned compared to where they are legal. This basic observation generates considerably wider insights, sometimes counterintuitive, about the impacts of parallel trade.

A. Profit-Increasing Parallel Imports

It seems evident that firms would gain maximal profits if they can separate markets and price according to local demand. However, this simple observation rests on the notion of third-degree price discrimination, in which there is a single price in each market because all consumers are homogeneous. In markets where consumers have heterogeneous demands, meaning different valuations for a particular good, and those consumer types can be segregated the situation changes. In these cases firms can set a menu of prices in different markets to separate consumers into types, such as those who value higher quality or want the good quickly and those who prefer lower quality or prefer to wait. Consumers sorting through this form of second-degree discrimination makes it possible for originator firms to see higher profits if international arbitrage through PI is permitted. Moreover, parallel imports actually may exist in equilibrium as they flow to where consumers have higher valuations.

Consider a simple example supporting PI that are profitable for the originator firms. Suppose that in one market a firm can set a high price to a price-inelastic group, the members of which find it too costly to engage in arbitrage. It can also set a low price to a price-elastic group, which for some reason faces low arbitrage costs. The firm now has the option of opening a second market abroad at an even lower price and permitting PI between the countries. Under these circumstances it is possible for PI to increase the firm’s global profits by raising price in the second (source) market while not diminishing profits derived from the high-price segment of the first market. This scenario seems consistent

---

38 This analysis was extended to more general circumstances by Kamal Saggi, Regional Exhaustion of Intellectual Property, 10 Int’l J. Econ. Theory 125 (2014).
with the common observation that pharmaceutical prices in high-income (and inelastic) private market segments are sustainable in poor countries despite the potential threat of parallel trade.  

A completely different example of how PI can profit the originator firm was set out by Raff and Schmidt. Consider a case where demand levels for a good are both unknown and different across countries, while unused inventories have little direct value to the manufacturer. In this situation, parallel trade potentially increases competition among retailers that sell the good but it also raises the expected value of inventories because they can be sold to other markets. Banning PI causes prices to fall when retailers have large stocks on hand and demand is low. This effect diminishes the retailers’ willingness to purchase large volumes and can reduce the manufacturing firm’s profits. However, allowing parallel trade may keep retail prices from falling sharply when demand falls, which sustains order volumes and increases the originator’s profits. Logic of this kind may explain why some industries, such as automobiles, take a relatively benign attitude toward parallel trade.

B. Vertical Pricing and Distributor-Level Parallel Trade

It is natural to conceive of PI as arbitrage against retail price differences across countries. However, as noted earlier, surveys suggest that the bulk of such trade happens at the wholesale level. This is true of both consumer goods, such as fashions, cosmetics, pharmaceuticals, and automobiles, and intermediate goods, including machinery and chemicals. There are three practical reasons why parallel trade by direct consumers is small. First, there are significant costs associated with finding reliable supplies and getting goods across borders, deterring small volumes. Second, there may be complementary services sold with the physical goods, such as calling plans for mobile phones, which are sacrificed through trades outside authorized channels. Third, consumer-level parallel trade may be illegal in order to protect public health. This is the case for pharmaceuticals in the EU, for example, and for alcoholic beverages in the U.S. Thus, the simple idea that manufacturers discriminate in prices among international retailers and individuals easily trade against these differences is wrong. Instead, firms sell to (often independent) wholesale distributors, which then deal with retailers. The bulk of PI entails specialized trading firms buying at the distributor level in cheap markets.

In this context, IPR owners generally wish to maintain a large degree of vertical control over its licensed distributors. In the U.S., for example, firms are permitted to establish

---


41 The most recent example is Keith E. Maskus & Frank Staehler, Retailers as Agents and the Limits of Parallel Trade, 70 EUR. ECON. Rev. 186 (2014). These authors analyze a case in which there is uncertainty about which consumers have high valuations and which have low valuations for a good. Firms can hire independent local retailers, which can help resolve this problem, but it is still possible for no PI to actually emerge.


43 Ganslandt and Maskus, supra note 16, at 946–47.
exclusive territories for their distributors, which can be efficient in terms of serving consumers across disparate locations. European firms also often set up national or regional distributors, attempting to control trading among them through contractual terms. However, this sets up the possibility that differential wholesale prices set across markets could generate incentives for PI between them, sourced at the distributor level. One means that original firms can use to manage this problem is through vertical price control (VPC), or setting wholesale prices that can endogenously limit or deter PI.

The primary economic theory of PI and VPC was developed by Chen and Maskus. In this framework, a manufacturer protected by IPR in both a home and foreign market has one independent distributor in each location. The firm needs to set its wholesale prices to induce profit-maximizing retail prices, which vary between markets according to demand conditions. If PI are legal, however, these distributors compete with each other in the higher-priced location.

This basic setup reveals several interesting tradeoffs. First, parallel trade generates more competition, reducing profits of the original firm. Second, this firm, in maximizing its profits, will choose to limit or even eliminate PI by setting a menu of wholesale prices. While these prices deal with the PI problem, they reduce overall profitability compared to the case where no PI are threatened. Thus, in VPC models parallel trade follows the usual intuition in that it reduces profitability of IPR owners. Finally, and this is an important point, parallel trade itself incurs real resource costs, which is costly in welfare terms. This latter observation might be considered yet another stylized fact: PI can be competitive but the simple act of hauling goods across borders uses up labor, capital, and other inputs, solely to arbitrage against prices.

This analysis was extended by later authors, who took fuller account of the strategic effects of market power at the distributor level. They closely studied the effects of diminished PI trade costs, which could arise from greater market integration or more efficient transportation systems. Working through the various cases, they demonstrate the curious possibility that, as the costs of shipping goods fall toward zero and the volume of PI increases, the manufacturer would react by raising its wholesale price in the import market even as it reduces the wholesale price in the export market. This happens because it becomes profitable at some point to push one of the distributors out of the market by raising the price it is charged, leaving the joint region open for the remaining distributor.

This is an important finding in policy terms. It suggests that an open regime of PI, coupled with declining trade costs, could cause greater concentration of the wholesaler market while inducing retail price divergence rather than convergence. Indeed, this may be one reason for the observed failure of retail prices within the EU to move together as much as anticipated. A deeper implication is that legalized PI, in an environment where trading partners actively seek to reduce the costs of trade, can be anticompetitive in its impacts at the distributor level. In fact, this problem can be exacerbated by a well-meaning competition rule requiring that manufacturers set uniform wholesale prices across

---

44 Ganslandt and Maskus, supra note 5, at 269–70.
46 Ganslandt and Maskus, supra note 16, at 946.
international markets, which is found in EU policy. Predictions of this kind, along with others that will emerge as this literature deepens, are ripe for empirical testing.

V. PARALLEL TRADE, R&D, AND POLICY SPILLOVERS

Static economic analysis shows that PI ordinarily reduce the profits made by an original manufacturer owning some form of protected IPR. How does this insight carry over to dynamic markets? To an important degree, expected profitability determines how much firms are willing to invest in new technologies and products. Accordingly, innovative firms would take the legal treatment of PI into account in determining their R&D programs.

One of the primary arguments against legalizing parallel trade is that it would diminish investments in new technologies. This claim is prominent in the U.S. debate over whether to legalize PI (called “re-imports”) of pharmaceuticals from Canada and Western Europe. Research-based pharmaceutical firms have lobbied successfully to defeat such legislation or its implementation. More generally, originator firms in any IPR-based industry tend to oppose the opening of markets to PI. For example, major music publishers strongly opposed the liberalization of import barriers in Australia in 1998, though they lost that battle.

Despite its importance, the theoretical economics literature on this point is limited and empirical analysis is absent. A few insights are worth highlighting, however, including the strong linkages among price controls, R&D, and parallel trade, especially in pharmaceutical products.

A. Models of R&D and Parallel Trade

Li and Maskus published the first theoretical model linking PI to investments in R&D in the vertical control model described above, adding an R&D stage prior to the competition between distributors. Thus, in a first stage the manufacturer decides how much to invest in a technology that, if successful, will reduce its marginal production costs. Depending on the amount of R&D, the firm chooses its vertical price structure either on the basis of a high marginal cost or a low marginal cost. The differences in these costs and prices determine the volume and impacts of PI. Their basic conclusion is that PI would reduce final-stage profits of the manufacturer, leading it to invest less initially. However, the extent of this impact on R&D depends on numerous factors, including the transportation cost in parallel trade. Thus, legalizing PI does not necessarily reduce economic welfare even though it diminishes the equilibrium expenditures on innovation. Indeed, it can raise

47 Maskus, Private Rights, supra note 11, at 183–85.
49 Maskus, Global Economy, supra note 10, at 211.
well-being if trade costs are small. At larger transport costs, however, the combination of resources used up in parallel trade and diminished R&D reduces welfare.\footnote{Similar results were found for product-improving innovation in Changying Li & John Robles, 

In a different theoretical framework, Valletti\footnote{Tommaso M. Valletti, 
\textit{Differential Pricing, Parallel Trade, and the Incentive to Invest}, 70 J. Int’l Econ. 314 (2006).} showed in a basic arbitrage model with a single monopolist and segmented markets that a policy mandating a uniform international price, such as a rule for global exhaustion of intellectual property rights, increases welfare \textit{ex post} but may reduce \textit{ex ante} investment in quality-enhancing R&D. More specifically, suppose differential pricing is based on variations in demand characteristics, meaning that some consumers in high-price markets go unserved. In this model the uniform price from unimpeded PI ultimately raises static welfare but reduces prior innovative investments due to diminished profitability. Whether overall welfare, in the \textit{ex ante} expected sense, rises or falls when costless PI are permitted, depends on market parameters. In a related paper a similar model is used to illustrate a welfare tradeoff in which international exhaustion and uniform pricing can raise static welfare but reduce investments in product quality.\footnote{Tommaso M. Valletti & Stefan Szymanski, 

Overall, then, while parallel trade tends to diminish incentives for R&D, the full impacts cannot be readily characterized. On these grounds I hesitate to predict that exhaustion policies have an identifiable causal impact on economic growth.

\section*{B. Price Controls and Parallel Trade}

I noted earlier the obvious point that government price controls in different countries can generate parallel imports. Things could go both ways, of course, for PI in turn could impact the ability to sustain these regulated pricing regimes. Consider, for example, the fact that in the EU each country has full authority over its public health policy, including choosing price regulations in pharmaceuticals that vary markedly across Member States.\footnote{Ganslandt and Maskus, \textit{supra} note 22, at 1036; see also \textit{Kanavos et al.}, \textit{supra} note 26, at 26–27.} Some countries, generally in lower-income southern Europe, set price ceilings that are substantially lower than prices in higher-income northern Europe. Despite this sovereignty in health policy, EU law is committed to permitting free circulation of goods as part of the Single Market. In consequence, parallel trading firms respond to these price differences and medicines move from countries with lower regulated prices, such as Spain, Portugal, Italy, and Greece, to markets with laxer regulations, such as the United Kingdom (U.K.), Netherlands, Germany, Sweden, and Denmark. It is reasonable to claim that the great bulk of PI in pharmaceuticals within the EU stem from such differences in policy.

On its face this remarkable juxtaposition of national sovereignty in health policy and the inability of governments to regulate PI seems inconsistent. If parallel imports withdraw supplies of medicines from highly regulated economies toward less regulated markets they should reduce the price gaps between markets. This would frustrate the
intentions of the regulations in source countries and push economies toward policy harmonization. Further, shifting from a regime of no PI to legalized trade would change incentives for innovative firms to undertake R&D in pharmaceuticals. Indeed, the presence of rigorous price regulations is sometimes blamed for declining investments in medical R&D in Europe relative to the U.S. in recent decades, a factor that should be exacerbated by unrestricted PI. Thus, PI in markets subject to price regulations could have markedly different effects than arbitrage in cases where the IPR-owning firm can adjust prices without worrying about price limits.

Once again, however, general conclusions are elusive, even in theory. Regarding the pricing issue, there is little evidence in the EU that PI has the effect of reducing international price differences, which remain large in pharmaceuticals. As noted above, studies suggest that, despite the elastic responses of PI to price differentials, the primary effects are to afford significant price-cost markups for parallel traders within the EU. In turn, there is little effective pressure on governments to change their regulatory systems, which are markedly stable. The reasons for this insensitivity are straightforward. First, price regulations in higher-price countries tend to stabilize retail and pharmacy prices, meaning they do not necessarily fall in the presence of PI. Second, PI volumes are rarely sufficient to place real pressure on price controls. Third, firms are permitted under EU law to set contractual limits on supplies sent to their distributors in various markets, so long as they do not explicitly aim to restrict PI. Combined with differential packaging and a natural tendency for patients and pharmacists to consider PI goods to be of lower quality, this permits sustainable brand-name pricing.

The issue of PI and R&D investments in the context of pharmaceutical price regulation has attracted attention among economists. The simplest analyses accord with intuition: international exhaustion should reduce incentives for R&D in the face of rigorous price controls. For example, in a horizontal arbitrage model Rey takes government policy as given and shows that parallel trade should cause retail prices in unregulated markets to fall and thereby reduce the global incentives to invest in R&D. Similarly, Danzon argues forcefully that permitting PI, at least among advanced economies, is unwise in a dynamic sense. In her view, international price discrimination is beneficial in that it reflects differences in elasticity of demand, making it an efficient means of allocating the costs of R&D among markets. Such discrimination, supported by market segmentation, is

55 This possibility explains the considerable concern on behalf of Canadian public health authorities whenever the U.S. contemplates removing its restrictions on re-imports; see Maskus, Private Rights, supra note 11, at 183–85.
57 Maskus, Private Rights, supra note 11, at 179–82.
60 Here she refers to market segmentation as supporting “Ramsey pricing,” which means permitting differential pricing across users of the services of a public good as the most efficient means of financing its development. This interpretation is problematic for many reasons.
Economic perspectives on exhaustion and parallel imports

therefore globally optimal from a social point of view. Such thinking militates against international exhaustion, at least where R&D costs are high and the innovative products of research bear global welfare spillovers. However, Danzon’s analysis is limited in that it fails to consider that price differences in medicines are generally associated with price regulations, not profit-maximizing price setting by pharmaceutical companies. In this decidedly second-best world of multiple distortions and policies, defining optimality is rarely that straightforward.

The most rigorous theoretical analysis of these tradeoffs is found in Grossman and Lai. These authors develop a two-country (or region) North-South model in which the South government chooses its optimal price-control policy, given the prior choice by the North government of an exhaustion regime. This is the first paper in economics to recognize explicitly that price controls and world exhaustion regimes are jointly determined, given fundamental welfare tradeoffs. They show that the North’s choice of PI regime can change the South’s government policy equilibrium through its induced impacts on innovation. Suppose that innovation of new medicines occurs solely in the North but these products are consumed in both locations. The South faces a basic tradeoff: by relaxing its price controls this country gains from access to new products but suffers higher consumer welfare costs from higher prices on patented goods already in the market. This tradeoff is altered by a shift in the North’s exhaustion policy. When the North chooses to permit PI, it becomes the potential destination of parallel exports of local supplies from the South. Because this reduces profits of the Northern pharmaceutical firms they may choose not to serve the South at all if prices in that market are capped below certain threshold levels. This lack of access is detrimental to the South, so its government reacts by relaxing its price controls, which in turn raises profits of the Northern firms. Under some market parameters it is possible for this profit gain to be larger than the costs of competition from PI.

In this way parallel trade weakens the incentives of lower-income economies to impose the aggressive price regulations that would maximize static consumer welfare in the absence of such trade. In turn, a low price cap in one country reduces incentives for R&D investment worldwide, which is harmful for future consumers even in the low-price countries. Arbitrage through PI thus makes government policies interdependent and forces every government to consider the consequences of its price regulation on global incentives to invest in new products. In their model this effect can be so important that permission of PI reduces free riding (that is, the ability of low-priced nations to benefit from research abroad) and ultimately strengthens the incentives for innovation. In equilibrium, global investment in research increases. The authors conclude that legalized parallel trade is actually pro-innovation, rather than a factor that restrains R&D in controlled goods.

There are other mechanisms under which exhaustion policies interact with other IPR

62 This paper is an extension of the seminal theory in Gene M. Grossman & Edwin L.-C. Lai, International Protection of Intellectual Property, 94 AM. ECON. REV. 1635 (2004). In that model the North and South non-cooperatively choose optimal patent lengths, which depend on their innovative capacities, market sizes, and demand patterns.
standards and generate international policy spillovers. One important question is the degree to which exhaustion policies affect the strength of IPRs themselves. Consider a situation in which the South must decide whether or not to provide patents, depending on the exhaustion regime of the North. If it does not protect patents the South permits local imitation of the North’s good, but this is likely to be a low-quality version. If the South has stronger preferences for the original, higher-quality version, it may have to offer patent protection in order to get the firm to provide it in their market. But its willingness to do so depends on the North’s exhaustion regime. If there is free PI into the North the degree of patent protection in the South would have to be correspondingly larger to induce entry of new goods. Again, we see that the WTO-provided flexibility of Northern governments to choose an exhaustion regime, while sensible on many grounds, is not innocent when it comes to inducing patent law changes around the world.

As in the case of vertical price control models of PI, the literature on arbitrage, dynamic innovation incentives, and policy interdependence is in its infancy. Considerably more research could be devoted to fleshing out relevant market characteristics that would affect the returns to R&D under parallel trade and how responsive IPR policies are to shifts in exhaustion regimes (and trade policy more generally). This analysis is important to sort out the likely relationships among IPR, competition and price regulation, parallel trade, and innovation policies. This research agenda remains important despite the absence to date of sound empirical evidence that PI in fact has such impacts. Indeed, one of the most pressing unanswered questions in the literature on parallel trade is its role in encouraging or discouraging innovation. While this project would pose formidable issues to resolve in terms of identifying the relevant effects, it is well worth pursuing.

VI. CONCLUSION

The exhaustion doctrine is a legal construct with complex implications, as noted in the other contributions to this volume. As a matter of economics, it may be construed simply in its essence as a regulatory decision regarding whether to keep a market open or closed to parallel imports. Countries with a strong comparative advantage in innovation and creation, such as the U.S. and the European Union as a whole, tend to see national exhaustion as an important component of an IPR owner’s right to control distribution across borders for the duration of her protection. Other countries tend to view openness to PI as an important means of sustaining competition and access to goods.

This chapter has reviewed the primary outcomes of economic theory analyzing PI as a form of international commodity arbitrage. At its simplest level PI imposes, in principle, considerable discipline on the ability of firms to separate markets and establish differential prices. Whether this outcome is globally and nationally beneficial or harmful depends on various features of market size, demand, and production capacity. Even in its most

---


elemental form, the welfare effects of parallel trade are ambiguous. When additional complications arising from market power and the ability to set strategic prices to limit or deter PI are added to this mix the story becomes yet more involved. In theory, moving from national to international exhaustion could be procompetitive in its price effects or anti-competitive in its impacts on distributor margins and concentration. This policy change could reduce profits and R&D incentives, thereby diminishing prospects for innovation and growth. Perhaps surprisingly, it could actually raise R&D incentives under important circumstances, by placing pressure on countries to relax their price controls in key commodities in order to sustain access to goods. These complexities and ambiguities arise everywhere in the analysis of parallel trade, itself a new scholarly endeavor with much more to learn. This is why one keen observer refers to the economics of exhaustion as a “jungle,” with little in the way of clear lessons though much to think about for scholars and policy-makers.65

In making this review I have left aside some important questions that embody broader policy elements than IPR exhaustion but are clearly related to this issue. Two questions in particular should be mentioned, which have received only the briefest of treatments here. First, some analysts argue that, at least in the area of ensuring access to essential medicines, a global policy of national exhaustion (or perhaps regional exhaustion with a limited geographical scope) would be beneficial.66 The notion is that strict market segmentation would induce pharmaceutical originators to engage naturally in tiered pricing, with extensive discounts for poor countries. This might be of particular utility in so-called “Type 2” diseases, such as heart ailments and cancer, which are prevalent in both rich and poor nations. A companion suggestion would be for wealthier governments to abandon their linkage of price controls to “reference pricing” in which domestic price ceilings depend partially on the lowest prices found elsewhere, including in developing countries. This policy is thought to discourage firms from charging low prices in the latter markets for fear of cannibalizing their price negotiations in larger and more lucrative locations. Our evidentiary basis to assess such claims is limited.

A second related area is the extent to which exhaustion policies, perhaps in conjunction with pricing regulations, affect decisions of firms to delay the launches of new products in various markets or alter their R&D incentives in particular goods. Available evidence indicates that such impacts could be significant, raising consequent concerns about the ultimate impacts of parallel trade on public health status.67 Again, however, far more research is called for.

There are two central lessons to be drawn from this review. First, there are numerous complex and diverse potential impacts of exhaustion polices and PI, which could in principle be of considerable importance for the global economy. This situation argues for far more close empirical work, tied to economic theory, in order to sort out just which factors

65 Fink, supra note 3, at 174.
66 Various arguments are reviewed in detail in Maskus, Private Rights, supra note 11, at 263–65, including a critical discussion of the empirical basis for this claim.
are important and how responsive market forces and government policies are to parallel trade. The unfortunate implication arises from the second lesson. The extreme scarcity of solid data on parallel trade places considerable limits on our abilities to answer such basic questions. Thus, it seems important to conclude the chapter by appealing to authorities and international organizations to devote more effort to collecting such databases, at least in sectors of considerable importance for public policy, and making them available for research.