The regional exhaustion of intellectual property

Kamal Saggi*
Vanderbilt University

July 2013

Abstract

This paper analyzes the causes and consequences of *regional exhaustion* of intellectual property, a policy regime under which a set of countries permit parallel imports from one another but not from the rest of the world. A three-country model is developed in which two high-income countries jointly choose their common exhaustion policy among national (NE), international (IE), or regional exhaustion (RE). The key result is that the two high-income countries choose to implement RE when they are relatively similar to each other and sufficiently high-income relative to the third country. We also consider a scenario where the policy choice set is restricted to non-discriminatory exhaustion regimes (i.e. NE or IE). Comparing the policy outcome of this constrained scenario with that of the core model, we show that the option to choose RE makes all countries better off.

Note: *This paper is written in honor of Bill Ethier, one of the sharpest and wittiest economists that I have had the honor of knowing. Like most graduate students at Penn, I watched in awe as Bill expounded the theory of international trade and commercial policy without ever relying on any notes or slides, crutches that most of us cannot do without! His prolific research contributions have left an indelible mark on the field. Not only have his ideas proven to be fundamental, they have been expressed with a beautiful economy of words that few can match.*

Keywords: Regional Exhaustion of IPRs, National Exhaustion, International Exhaustion, Parallel imports, Market power, Welfare. **JEL Classifications**: F13, F10, F15.

*Address: Department of Economics, Vanderbilt University, Box 1819-Station B, Nashville, TN 37235-1819, USA. E-mail: k.saggi@Vanderbilt.Edu; Phone: (615) 322-3237.

1 Introduction

By virtue of their membership in the World Trade Organization (WTO), policies of almost all major economies with respect to the protection and enforcement of intellectual property rights (IPRs) must abide by the Agreement on Trade Related Intellectual Property Rights (TRIPS). In a nutshell, this controversial agreement calls for virtually the complete harmonization of IPR policies across WTO member countries even though economic conditions and technological capabilities vary dramatically across them. While TRIPS is far-reaching in terms of what it demands of WTO member states in the realm of IPRs, there is an important set of policies that it leaves completely unregulated and unconstrained, i.e., policies pertaining to the *exhaustion* of IPRs. Article 6 of TRIPS says that "nothing in this Agreement shall be used to address the issue of the exhaustion of intellectual property rights", a statement that grants wide latitude to WTO member countries with respect to their exhaustion policies.

Exhaustion policies affect market outcomes by determining the legality of parallel imports – i.e. imports of goods protected by IPRs from foreign markets where they were originally sold by rights holders.[1] These policies can be one of three types: national, international, or regional.[2] Briefly speaking, if a country follows national exhaustion (NE), it effectively bans parallel trade since under this principle a right holder's IPR over a product is deemed to expire only in the country of first sale, making it possible for the right holder to prevent resale of its product in other markets. Under the doctrine of international exhaustion (IE), the relevant IPR expires globally with the first sale of a product anywhere so that a right holder cannot block parallel trade. Finally, under regional exhaustion (RE), the right expires upon first sale within a well-defined region comprising a group of countries but *not* outside it. As is clear, RE is discriminatory in the sense that it allows free parallel trade within a well defined region but prohibits parallel imports from the rest of the world.

The world's largest economic market – i.e. the European Union (EU) – practises regional or community exhaustion of IPRs. In a series of important decisions made over

[1] Data on parallel trade are hard to come by since it is not always possible to distinguish imports from reimports. Nevertheless, it is well established that such trade occurs in footwear and leather goods, musical recordings, cars, consumer electronics, domestic appliances, cosmetics, clothing, pharmaceuticals, soft drinks, and several other consumer products (NERA, 1999). As one might expect, both in the US and the EU parallel trade in pharmaceuticals receives the most attention from researchers and policy-makers. See Kanavos and Costa-Font (2005) for stylized facts regarding parallel trade in pharmaceuticals within the EU.

[2] Maskus (2000) provides a discussion of the observed variation in exhaustion policies across countries.

the years, the European Court of Justice has essentially argued that the free flow of parallel imports within the EU coupled with a ban on such imports from the rest of the world is the only policy that is compatible with the underlying objective of establishing a common internal market within the EU (Baudenbacher, 1998). This side of the Atlantic, the politically charged question of whether parallel imports of pharmaceuticals should be permitted from Canada and Mexico has been debated repeatedly in the US Congress over the years.[3] If such trade with its two neighboring countries were to be permitted by the US, it would essentially amount to the US implementing a policy of regional exhaustion in pharmaceuticals with its immediate neighbors. As things stand now, the US practises a policy of national exhaustion while many developing countries (but not all) favor international exhaustion.

Building on the approach of Roy and Saggi (2012a), this paper develops a three-country model that addresses several positive and normative questions: What factors determine the jointly optimal exhaustion policy of a pair of countries? When and why does regional exhaustion arise in equilibrium? What are the welfare implications of regional exhaustion for the outsider, i.e., the country from which parallel imports are not permitted?[4]

The issue of regional cooperation is especially relevant in the context of exhaustion policies because parallel trade is more likely to occur between geographically proximate countries. After all, such trade arises when retailers or other parties attempt to arbitrage away existing price differentials across international markets so that the margins earned by those engaged in parallel trade are likely to be rather small, at least relative to mark-ups earned by monopoly suppliers. Therefore, it is not surprising that policy discussions with respect to parallel trade have tended to be either about goods that have low trade costs relative to price differentials (such as software, DVDs, or pharmaceuticals) or between neighboring countries (within EU or NAFTA) or both.

[3]For example, in 2009 a measure to allow importation of prescription drugs from abroad fell short in the US Senate by just 9 votes due to fierce opposition from the pharmaceutical industry (*Wall Street Journal*, Dec 16, 2009). Similarly, earlier in 2000 a measure to permit drug reimports from Canada passed the US Congress but was not implemented by the then Secretary of Health and Human Services. In both instances, safety concerns were used as a justification for not allowing reimports but it is clear that the central issue for the pharmaceutical industry is the adverse impact of parallel imports (or their threat) on prices and profitability in the US market. See Goldberg (2010) and the empirical studies summarized therein for a discussion of how the practice of global reference pricing for pharmaceuticals on the part of rich country governments has adverse consequences for poor countries.

[4]While Roy and Saggi (2012a and 2012b) derive optimal exhaustion policies from the viewpoint of individual countries, the focus here is on jointly optimal policies. Furthermore, Roy and Saggi (2012a and 2012b) only consider a model with two countries and cannot therefore speak to the question of regional exhaustion, the main motivation behind the present paper.

Motivated by these considerations, this paper develops a three-country model in which international arbitrage can occur costlessly except when it is explicitly forbidden by exhaustion policies. A single patent holder/firm produces a distinct good in countries A and B and consumer tastes (or income distributions) differ across countries such that, in the absence of arbitrage, prices in countries A and B are higher than those in country C. The timing of decision making is as follows. First, governments of the two high-income countries jointly set their common exhaustion policy choosing among (a) NE under which no parallel imports of any type are permitted (b) RE under which there is free parallel trade between the two countries but not with country C or (c) IE under which there is free parallel trade in the global economy.[5] Next, taking the exhaustion policies set by governments into account, firms choose prices and trade occurs.

Two intuitive ideas drive the model. One, under IE firms may choose not to sell in the low-price market in order to sustain high prices in the two high-income markets. Two, since a firm cares only about its global profit while a welfare-maximizing government cares also about local consumer surplus, firms are more willing to eschew sales in the low-price market than is optimal from the perspective of national welfare of their home countries.

As might be expected, each firm serves all markets at a uniform price – a market outcome we call *global integration* [ABC] – only when the high-income countries are open to parallel imports and the degree of asymmetry across markets is not too large. Furthermore, by definition, if both high-income countries implement RE then markets of the two high-income countries are integrated whereas that of the low-income country is segmented from them. Under such *partial global integration* [AB;C], each firm charges a common price in the two high-income countries while selling at a lower price in the low-income country. Finally, if the high-income countries choose NE then each firm sells in all countries, charging a different price in each market.

In equilibrium, the common exhaustion policy set by the two high-income countries takes into account their joint consumer surplus and the global profits of their firms. When the degree of market asymmetry is small, the two countries implement IE and global integration [ABC] obtains. However, when countries A and B are relatively symmetric while country C is sufficiently low-income relative to them, they opt for RE and partial global integration [AB;C] obtains as the equilibrium outcome.

[5] Since prices are lowest in country C when markets are segmented globally, its exhaustion policy is inconsequential and IE on the part of the two high-income countries amounts to free parallel trade in the world.

To isolate the consequences of being able to discriminate with respect to exhaustion policies, we also address the following counterfactual question: *what if countries could only implement non-discriminatory exhaustion policies*? This counterfactual analysis provides two crucial insights. One, the degree of market integration achieved in the global economy is actually *lower* when the freedom to discriminate with respect to exhaustion policies is absent since *regional integration* [AB] – a market outcome under which firms do not sell in country C – ends up replacing partial global integration [AB;C]. Second, and more importantly, market outcomes when RE is an available option (weakly) *Pareto-dominate* those that obtain when only non-discriminatory exhaustion policies can be implemented. This is because allowing for RE leads to partial global integration [AB;C] *replacing* regional integration [AB] as the market outcome: by freeing firms from the threat of parallel imports from the low-price market, RE between the two high-price markets makes it optimal for firms to also sell in the low-price market. Why [AB;C] Pareto dominates [AB] is easy to see. Firms prefer [AB;C] to [AB] since their global profits are higher while consumers in country C prefer it because they have access to foreign goods only under the former outcome. Indeed, the prices at which consumers in country C have access to both goods under [AB;C] are the lowest over all possible market outcomes. Finally, consumers in countries A and B are indifferent between [AB;C] and [AB] since they face the same prices under the two outcomes.

Malueg and Schwarz (1994) were the first to compare the global welfare effects of various types of exhaustion policies but they took such policies to be exogenously given. In their model, when confronted with the possibility of arbitrage-induced parallel imports, a monopolist chooses to serve only markets where demand is relatively inelastic (i.e. price is high) since parallel imports from low-price markets lower its total profit. While this mechanism also plays a central role in the present model, it is worth noting, however, that openness to parallel imports does not necessarily lower a firm's profit when there is strategic interaction between the firm and another party. For example, Pecorino (2002) shows how the possibility of parallel imports can tilt the outcome of a bargaining game between a monopolist and a foreign government in favor of the monopolist by reducing its willingness to supply the foreign market. Similarly, Roy and Saggi (2012a and 2012b) show how openness to parallel imports can soften price competition under oligopoly. They also analyze how the presence of strategic interaction in the product market affects equilibrium policies implemented by governments.

Richardson (2002) considers the viewpoint of importing countries facing a global monopoly supplier and shows that, in equilibrium, all countries choose to allow parallel

imports since doing so ensures that the good is available locally at the lowest possible price. While his analysis delivers useful insights regarding the economic effects of IE and NE, it does not address when and why RE might arise as an equilibrium policy choice. This is important because, as was noted earlier, RE is not only the policy of the largest common market in the world (i.e. the EU) but has also been frequently considered for adoption by policy-makers concerned about the high prices of pharmaceuticals in the United States.

2 A simple model of regional parallel trade

We consider a world comprising three countries (indexed by $j =$ A,B, or C) in order to understand the causes and consequences of RE of intellectual property.

Consumers in each country consume two (patented) goods: a and b indexed by $i = a, b$. Good a is produced by a monopolist/patent holder in country A called firm a. Similarly, firm b in country B is the sole producer of good b. The quality of each good is normalized to 1 and the production cost to zero. Country C is an importer of both goods.

The retail sector in each country is assumed to be competitive with zero unit cost so that the prices set by firms/patent holders equal those facing consumers. Each consumer buys at most one unit of good i. Utility under no purchase equals zero. If consumer k in country j buys good i at price p_{ij}, the utility it derives from the good given by:[6]

$$u_{ij}^k = \theta_j^k - p_{ij} \tag{1}$$

As is well known, the taste parameter $\theta_j^k \geq 0$ can be interpreted one of two ways. Either one can view it as capturing differences in primitive preferences (such as the marginal utility of quality) or as capturing differences in income across consumers with poorer consumers having lower θ_j^k's. In what follows, we adopt the latter interpretation and assume that θ_j is uniformly distributed over the interval $[0, \mu_j]$ in country $j =$A, B, or C where $\mu_A > \mu_B > \mu_C = 1$. Countries A and B are referred to as the two high-income (or high-price) countries. It proves convenient to denote them by uppercase letters I and J where I, J = A or B and the home market of firm i by I.

The interaction between governments and firms occurs as follows:

[6] An alternative and more general formulation would be to allow consumer preferences to differ across goods so that $u_{ij}^k = \theta_{ij}^k q_i - p_{ij}$. Setting $\theta_{ij}^k = \theta^k$ simplifies the exposition by making demand functions symmetric across goods.

Stage 1: In the first stage, the two high-income countries jointly choose whether to follow (*i*) NE (*ii*) IE or (*iii*) RE of intellectual property.[7] Under NE, they prohibit parallel imports from all sources source; under IE they allow them from any source; whereas under RE they allow them only from one another. The common exhaustion policy is chosen to maximize their joint welfare.

Stage 2: After the two governments have chosen their common exhaustion regime, firms choose prices and trade occurs. Under NE, foreign retailers cannot resell products in their markets. However, under the other two regimes, foreign retailers from the relevant market(s) can engage in parallel trade if it is profitable for them to do so.

3 Market outcomes

Suppose the common policy regime of the two high-income countries is NE. Since parallel imports are forbidden, firms are free to price discriminate internationally and each firm sets a different (i.e. market specific) price in each country. In country j firm i choose p_{ij} to solve:

$$\underset{p_{ij}}{Max}\ \pi_{ij}(p_{ij}) = p_{ij}x_{ij}(p_{ij}) = \frac{p_{ij}}{\mu_j}(\mu_j - p_{ij}) \tag{2}$$

which gives firm i's optimal price in country j as:

$$p^*_{ij} = \frac{\mu_j}{2}$$

Note that p^*_{ij} is increasing in μ_j: i.e., the pattern of optimal price discrimination is such that $p^*_{iA} \geq p^*_{iB} \geq p^*_{iC}$. We refer to this market outcome as *global segmentation* [A;B;C].

Firm i's global profit under global market segmentation [A;B;C] equals

$$\pi^*_i = \sum_j \pi^*_{ij} \text{ where } \pi^*_{ij} = \frac{\mu_j}{4} \tag{3}$$

As is clear, π^*_i is the maximum (i.e. optimal monopoly) profit a firm can earn on the global market. As a result, if the two high-income countries choose NE, global segmentation [A;B;C] obtains as the market outcome.

[7] In principle, one could also consider unilateral policy choices in this model. However, since the leading example motivating the paper is the regional exhaustion policy of the EU, it is more realistic to derive optimal policies under the assumption that there is coordination between the two high-income countries.

Consider market outcomes under IE. When both high-income countries permit parallel imports, firms are more constrained in their pricing behavior. Each firm faces a trade-off between charging optimal market specific prices and the number of markets served: the more markets a firm serves, the further away it gets from its optimal price in each market. Thus, when parallel imports can flow freely in the global economy, as they do under IE, each firm chooses between the following three pricing strategies:

(a) Sell in all markets at a common price.

(b) Sell only in the two high-income markets at a common price that maximizes its profits in those two markets.

(c) Sell only in its home market at its optimal monopoly price.

To derive conditions under which it is profit-maximizing for a firm to adopt each of these pricing strategies, we first derive optimal prices under each strategy. If both firms choose to sell in all markets under IE, the market outcome is referred to as global integration or [ABC]. Under this outcome, assuming $p_i < \mu_j$, firm i chooses its global price p_i to solve

$$\underset{p_i}{Max} \sum_j \pi_{ij}(p_i) = \underset{p_i}{Max} \sum_j p_i x_{ij}(p_i) = \underset{p_i}{Max} \sum_j \frac{p_i}{\mu_j}\left(\mu_j - p_i\right) \tag{4}$$

Solving this problem gives firm i's optimal price under global integration [ABC]:

$$p_i(\text{ABC}) = \frac{3}{2}\frac{\Pi_j \mu_j}{\sigma} \text{ where } \sigma \equiv \sum_{j \neq k} \mu_j \mu_k \tag{5}$$

It is straightforward to show that firm i's optimal uniform price $p_i(\text{ABC})$ is increasing in all of its arguments (i.e. in μ_j where $j =$A,B, or C). Firm i's profit under [ABC] equals

$$\pi_i(\text{ABC}) = p_i(\text{ABC}) \sum_j \frac{1}{\mu_j}\left(\mu_j - p_i(\text{ABC})\right) \tag{6}$$

Now suppose each firm decides to sell only in the two high-income markets at a common price, an outcome that is referred to as regional integration or [AB]. Under this outcome, firm i chooses its common price p_i in the high-income countries to solve

$$\underset{p_i}{Max} \sum_J \pi_{iJ}(p_i) = \sum_J p_i x_{iJ}(p_i) = \sum_J \frac{p_i}{\mu_J}(\mu_J - p_i) \text{ where } J = A, B$$

Solving which gives firm i's optimal price under regional integration [AB]:

$$p_i(\text{AB}) = \frac{\Pi_{j \neq C} \mu_j}{\sum_{j \neq C} \mu_j} \tag{7}$$

Firm i's total profit under regional integration [AB] is given by

$$\pi_i(\text{AB}) = \sum_J \frac{p_i(\text{AB})}{\mu_J}(\mu_J - p_i(\text{AB}))$$

If firm i sells in its home market and in country J while firm j sells only in market J, then we get *partial regional integration* $\langle \overrightarrow{\text{IJ}} \rangle$. We have $p_{iJ}(\overrightarrow{\text{IJ}}) = p_i(\text{IJ})$ whereas $p_{jJ}(\overrightarrow{\text{IJ}}) = p^*_{jJ}$. Finally, firm i's profit when it only serves its domestic market is given by $\pi^*_{iI} = \frac{\mu_I}{4}$.

We can now state:

Lemma 1: *Since* $\mu_A > \mu_B > \mu_C$ *we have:*
(i) $p^*_{iA} > p^*_{iB} > p^*_{iC}$.
(ii) $p_i(\text{AB}) = \sum_{j \neq C} \lambda_j(\text{AB}) p^*_{ij} > p_i(\text{ABC}) = \sum_j \lambda_j(\text{ABC}) p^*_{ij}$ *where* $0 < \lambda_A(\text{ABC}) = \frac{\mu_B \mu_C}{\sigma} < \lambda_B(\text{ABC}) = \frac{\mu_A \mu_C}{\sigma} < \lambda_A(\text{AB}) = \frac{\mu_B}{\mu_A + \mu_B} < \lambda_B(\text{AB}) = \frac{\mu_A}{\mu_A + \mu_B} < 1$.

Lemma 1 has four main messages. First, when free to price discriminate internationally, each firm charges its highest price in country A and its lowest price in country C. This pricing behavior simply reflects the differing willingness to pay on the part of consumers in different countries. Second, if price discrimination is not possible so that a firm must charge a common price in markets it serves, then its common price when it serves only the two high-price markets is higher than its price when it serves all markets. The intuition is straightforward: if the least profitable market is also served, then the firm lowers its common price to take account of weak demand conditions in that market. Third, the common price a firm charges in markets that it serves is a weighted average of its optimal market specific prices for those markets. Note also that $\lambda_A(\text{ABC}) < \lambda_B(\text{ABC})$ and $\lambda_A(\text{AB}) < \lambda_B(\text{AB})$: in other words, the weights determining the common price are inversely proportional to the relative importance of each market. This simply reflects the optimization involved in setting a single profit-maximizing price for all markets served. Fourth, the inequalities $\lambda_A(\text{ABC}) < \lambda_A(\text{AB})$ and $\lambda_B(\text{ABC}) < \lambda_B(\text{AB})$ say that the weight given to the optimal price in each high-price market is lower if a firm serves all three markets at a common price relative to when it serves only the two high-price markets.

This is also quite intuitive: the constraint that the inability to price discriminate puts on a firm's pricing behavior is more binding when it serves three asymmetric markets as opposed to only the two high-income markets.

Note that

$$\pi_a(\text{AB}) \geq \pi^*_{aA} \Leftrightarrow \mu_A \leq 3\mu_B \tag{8}$$

i.e. if the asymmetry between the two high-income markets is sufficiently large then firm a finds it profitable to serve only its home market as opposed to serving both markets A and B when parallel imports can flow freely between them. To rule out this uninteresting scenario, for the rest of the analysis we assume that $\mu_A \leq 3\mu_B$. Together with the assumption $\mu_A > \mu_B$ this implies that our analysis will be restricted to the region where $\mu_B < \mu_A \leq 3\mu_B$.

Lemma 2: *Suppose the policy regime chosen by countries A and B is international exhaustion. Then, the following hold:*

(i)

$$\pi_i(\text{ABC}) \geq \pi_i(\text{AB}) \Leftrightarrow \mu_A \leq \mu^c_A = 5\mu_B\mu_C/(4\mu_B - 5\mu_C) \tag{9}$$

where $\partial\mu^c_A/\partial\mu_B < 0$*;* $\partial^2\mu^c_A/\partial^2\mu_B > 0$*; and* $i = a, b$*.*

(ii)

$$\pi_i(\text{ABC}) \geq \pi^*_{iI} \Leftrightarrow \mu_I \leq \mu^*_I = 8\mu_J\mu_C/(\mu_J + \mu_C) \tag{10}$$

where $\partial\mu^*_I/\partial\mu_J > 0$ *and* $\partial^2\mu^*_I/\partial^2\mu_J < 0$*.*

Part (i) of Lemma 2 says that each firm prefers to serve all markets at a common price to serving only the two high-price markets if $\mu_A \leq \mu^c_A$. The threshold μ^c_A is common for both firms since each firm is considering the same decision: whether or not to drop market C. The fact that $\partial\mu^c_A/\partial\mu_J < 0$ means that from firm i's perspective, an increase in the profitability of the market in the other high-income country makes it more attractive to serve only the two high-income countries as opposed to serving all three countries.

Part (ii) says that if a firm's home market is sufficiently lucrative (i.e. exceeds the threshold μ^*_I), then it prefers to sell only at home as opposed to serving all markets since the high degree of asymmetry with respect to the other markets forces it to set a price that is too far away from its optimal home price. The fact that $\partial\mu^*_I/\partial\mu_J > 0$ is quite intuitive: as the attractiveness of other markets' increases, firm i's preference gets tilted in favor of serving all markets as opposed to only its home market.

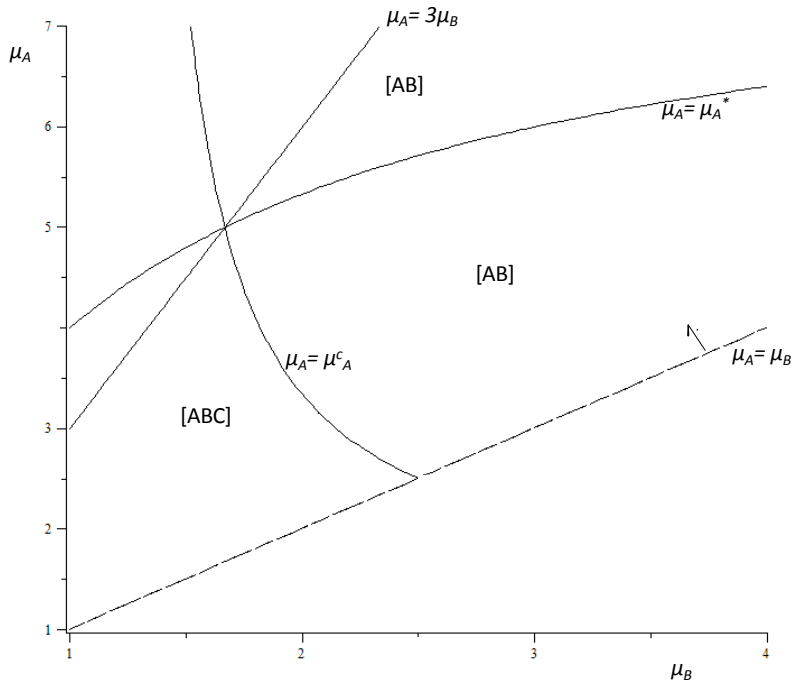

Figure 1: Market outcomes under international exhaustion

The first major result can now be stated:

Proposition 1: *Suppose the two high-income countries choose international exhaustion. Then, global integration* [ABC] *obtains as the market outcome if* $\mu_A \leq \mu_A^c$ *whereas regional integration* [AB] *obtains otherwise.*[8]

Figure 1 illustrates market outcomes under IE in the (μ_B, μ_A) space. Recall that only the parameter space above the 45 degree line and below the $\mu_A \leq 3\mu_B$ is relevant. The downward sloping curve plots μ_A^c. When the three countries are relatively similar to each other, as they are when $\mu_A \leq \mu_A^c$, firms in the two high income countries find it profit maximizing to sell in all markets and global integration [ABC] obtains. When optimal prices for countries A and B are similar to each other but differ substantially from those in country C, then firms choose to sell only in the two high-price markets and regional integration [AB] obtains.

Now consider market outcomes under RE. It is clear that when the two high-income countries implement RE, both firms necessarily sell in country C. This is because each

[8]It is straightforward to show that if $\mu_A \leq 3\mu_B$ then we must have $\mu_A^c \leq \mu_A^*$.

firm is free to earn π_{iC} in country C without affecting its profit in other markets since parallel imports cannot flow into those markets from country C. Thus, the only question is whether to serve all markets or only the home market and country C. Firm i prefers selling in all markets to selling in only countries I and C iff

$$\pi_i(\text{AB};\text{C}) \geq \pi^*_{iI} + \pi^*_{iC} \Leftrightarrow \pi_i(\text{AB}) + \pi^*_{iC} \geq \pi^*_{iI} + \pi^*_{iC} \Leftrightarrow \pi_i(\text{AB}) \geq \pi^*_{iI} \Leftrightarrow \mu_I \leq 3\mu_J \quad (11)$$

Since $\mu_B < \mu_A \leq 3\mu_B$ it must be that $\pi_i(\text{AB}) \geq \pi^*_{iI}$ so it follows that under RE, both firms prefer selling in all markets as opposed to only in their respective home markets and country C. Thus, under RE, in addition to the four types of market outcomes described under non-discriminatory exhaustion policies, we need to consider only one additional outcome, i.e., partial global integration [AB;C] under which each firm sells in the two high-income countries at a uniform price and in country C at its optimal monopoly price for that market. Market prices and profits under this outcome can be recovered using the preceding derivations. We have $p_{iJ}(\text{AB};\text{C}) = p_{iJ}(\text{AB})$ and $p_{iC}(\text{AB};\text{C}) = p^*_{iC}$.

We can now state:

Proposition 2: *If the two high-income countries choose regional exhaustion as their policy, then partial global integration* [AB;C] *obtains as the market outcome.*

4 Optimal policy

To derive equilibrium policies, we assume that the objective of countries A and B is to maximize their joint welfare. Therefore, while choosing their common exhaustion policy, the two governments take into account consumer surplus over both goods and global profits of the two firms.

Welfare of country I where I=A,B under market outcome M is given by

$$w_I(\text{M}) = \sum_i cs_{iI}(\text{M}) + \sum_j \pi_{Ij}(\text{M})$$

where $i = a, b$; j =A,B, or C; I,J=A,B, and M= [AB], [AB;C], [ABC], or [A;B;C] whereas welfare of country C equals

$$w_C(\text{M}) = \sum_i cs_{iC}(\text{M})$$

where

$$cs_{ij}(\text{M}) = \frac{1}{\mu_j} \int_{p_{ij}(\text{M})}^{\mu_j} (\theta - p_{ij}(\text{M}))\, d\theta = \frac{1}{2} \frac{\left(\mu_j - p_{ij}(\text{M})\right)^2}{\mu_j} \quad (12)$$

Plugging in the relevant prices into the above formula yields consumer surplus in a country under a particular market outcome. Using the relevant expressions for consumer surplus and firm profits under alternative outcomes, the following proposition can be established by direct calculations:

Proposition 3: *Each country's welfare under alternative market outcomes is as follows*:

(*i*) $w_A(\text{ABC}) > w_A(\text{AB;C}) > w_A(\text{A;B;C})$.

(*ii*) $w_I(\text{IJ;C}) = w_I(\text{IJ}) + \pi^*_{iC} > w_I(\text{IJ})$ *for* I,J=A,B.

(*iii*) $w_B(\text{A;B;C}) > w_B(\text{AB; C})$ *and* $w_B(\text{A;B;C}) > w_B(\text{ABC}) >$ *iff* $\mu_A > \mu^w_A$ *where* $\partial\mu^w_A/\partial\mu_B > 0$.

(*iv*) $w_C(\text{A;B;C}) = w_C(\text{AB;C}) > w_C(\text{A;BC}) > w_C(\text{ABC}) > w_C(\text{AB})$.

Part (*i*) of Proposition 3 simply reflects the fact that prices are lowest in country A's market under global integration [ABC] whereas they are highest under global segmentation [A;B;C]. Part (*ii*) clarifies that from the viewpoint of the two high-income countries, the welfare of each under partial global integration is the sum of its welfare under regional integration and the local firm's optimal monopoly profit in country C. Indeed, partial global integration maximizes the joint welfare of countries A and B subject to the constraint that parallel imports from country C are forbidden. The threat of parallel imports from country B to A ensures that firms charge the same prices in both markets while a ban on parallel imports from country C ensures that both firms export to country C because each firm is free to charge its optimal (low) price in country C without having to lower its prices in countries A and B.

Part (*iii*) says that, unlike country A, country B's most preferred regime is not necessarily global integration. Furthermore, country B prefers global segmentation [A;B;C] to regional integration [AB]. The intuition for this is clear. Freeing parallel trade with country A has two consequences for country B, both of which are negative. First, it raises prices in country B's market and therefore lowers the welfare of its consumers. Second, it reduces the total profit of firm *b* since it is unable to charge its optimal prices in all markets, as it does under global segmentation. Therefore, country B loses from having free parallel trade only with country A. The intuition for the comparison of [A;B;C] and [ABC] is now easy to see. While firm profits decline if global segmentation is replaced by global integration, consumer surplus in country B increases provided that country A is not so large that the downward pressure on local prices that results from permitting parallel trade with country C is swamped by the upward pressure caused by integration with country A.

The intuition for why country C's most preferred outcome is global segmentation is obvious: local prices are the lowest under this outcome whereas they are the highest under global integration [ABC]. Finally, the worst outcome for country C arises when its market is simply not served by firms, as is the case under [AB].

Let the joint welfare of countries A and B under market outcome M be denoted by $w_{AB}(\text{M})$. We can now state one of our key results:

Proposition 4: *The joint welfare of countries* A *and* B *under alternative market outcomes is as follows*:
(i) $w_{AB}(\text{ABC}) > w_{AB}(\text{AB;C}) > \max\{w_{AB}(\text{A;B;C}), w_{AB}(\text{AB})\}$.
(ii) $w_{AB}(\text{A;B;C}) > w_{AB}(\text{AB})$ *iff* $\mu_A < \mu_A^J$ *where* $\partial\mu_A^J/\partial\mu_B > 0$.

Provided that firms sell in all markets, free flow of parallel imports among all countries is the most desirable market outcome from the perspective of the two high-income countries since the threat of parallel imports from country C helps lower local prices. As was noted earlier, while opening to parallel imports from both sources necessarily lowers prices in country A, the situation is more mixed from country B's viewpoint: the openness to parallel imports on the part of country A raises prices in country B whereas its own openness to parallel imports from country C lowers prices. Furthermore, note that when comparing partial global integration [AB;C] to global integration [ABC] the increased openness to parallel imports from country C under the latter regime necessarily lowers prices in country B. However, since openness to parallel imports can induce firms to not export to country C, partial global integration [AB;C] is more desirable than both market segmentation [A;B;C] and regional integration [AB]. Intuitively, partial global integration [AB;C] is an attractive 'middle ground' between lowering prices in the highest price market by allowing parallel imports to flow between countries A and B while also preserving the export incentives of firms and thereby collecting rents from country C's market. Insofar as the comparison between market segmentation [A;B;C] and regional integration [AB] is concerned, part (*ii*) of Proposition 4 informs us that market segmentation is better for the joint welfare of countries A and B than regional integration iff $\mu_A < \mu_A^J$. The intuition for this is clear: when $\mu_A < \mu_A^J$ market prices are fairly similar in the two high-income countries even in the absence of integration and it is important to induce firms to export to country C's market, making global segmentation a more desirable alternative than regional integration.

We can now derive the jointly optimal parallel import policy of countries A and B. From Figure 1 we know that if they choose IE as their common policy, global integration

[ABC] obtains whenever $\mu_A \leq \mu_A^c$ and regional integration [AB] otherwise. Given that global integration confers the highest joint welfare on the two countries (see Proposition 4) it is immediate that whenever $\mu_A \leq \mu_A^c$, the optimal policy of countries A and B is IE. Also note from Proposition 4 that if global integration cannot be induced as the market outcome then the next best option for countries A and B is partial global integration [AB;C]. From Proposition 2 we know that a policy of RE delivers [AB;C] as the market outcome whenever $\mu \leq 3\mu_B$.

The following result can now be stated:

Proposition 5: *If $\mu_A > \mu_A^c$ then the jointly optimal policy of the two high-income countries is regional exhaustion and the resulting market outcome is partial global integration* [AB;C]*; otherwise their optimal policy is international exhaustion and the resulting market outcome is global integration* [ABC].

To clarify the implications of RE for market outcomes and welfare, suppose countries must choose between the two non-discriminatory exhaustion policies of NE or IE so that they cannot implement RE. Under such a constraint, we can state:

Corollary 1: *If regional exhaustion is not a feasible policy option, equilibrium policy choices are the same as before except when $\mu > \mu_A^c$: over this region, the two countries implement international exhaustion and regional integration* [AB] *replaces partial global integration* [AB;C] *as the market outcome.*

Thus, when the discriminatory exhaustion policy RE is unavailable or simply infeasible to implement, an outcome where global markets are more integrated is *less* likely to obtain. Furthermore, and more importantly, aggregate world welfare under partial global integration [AB;C] is higher than that under regional integration [AB]:

$$ww(\text{AB;C}) = ww(\text{AB}) + w_C(\text{A;B;C}) > ww(\text{AB})$$

Indeed, *partial global integration* $\langle$AB;C$\rangle$ *is Pareto-improving over regional integration* [AB]. Intuitively, countries A and B are strictly better off under [AB;C] relative to [AB] because their consumers face the same prices under the two regimes whereas their firms fare strictly better (since each earns its optimal monopoly profit in country C's market). Note, however, that country C is also strictly better off because its consumers are supplied both goods under [AB;C] whereas they are supplied neither good under [AB]: consumer access at monopoly prices is better than no access at all.[9]

[9] Although the context is different, this result is reminscient of Ethier's argument that regional trade agreements have a useful role to play in the multilateral trading system – see Ethier (1998).

5 Concluding remarks

This paper develops a simple model to shed light on the economics of regional exhaustion, a discriminatory policy that permits parallel imports from some trading partners but not others. The analysis is motivated by the experience of the European Union that practises this policy and of the United States, where discussions regarding the merits of permitting parallel imports from neighboring countries such as Canada and Mexico seem to resurface in Congressional debates at regular intervals.

One of the few, and perhaps the only, exception to non-discrimination available to WTO members with regard to their IPR policies is that they can pursue discriminatory exhaustion policies. This exception appears to conflict with the widely held view among policy-makers and researchers that the principle of non-discrimination underlying the multilateral trading system is generally a good idea. This paper has shown that, contrary to what common intuition might suggest, the freedom to discriminate with respect to exhaustion policies does not lead to beggar-thy-neighbor outcomes. In fact, an important result of this paper is that if countries were required to adopt only non-discriminatory exhaustion policies, the resulting outcomes would be (weakly) Pareto inferior: either the welfare of each country would be unaffected or all countries would be made worse off. This result argues in favor of the wide latitude available to WTO members with respect to their national exhaustion policies.

While the analysis provides several interesting insights about regional exhaustion, it abstracts from the effects of this policy on incentives for innovation. In this context, it is worth noting that whether national or international exhaustion provides stronger incentives for innovation is far from a settled question. For example, while Li and Maskus (2006) find that national exhaustion encourages incentives for cost reducing process innovation by a monopolist, in a North-South framework Grossman and Lai (2008) show that incentives for product innovation can be higher when North's policy is international exhaustion and the Southern market is subject to an endogenous price control.[10] The relationship between regional exhaustion and innovation is a topic worthy of future research.

[10] Valletti (2006) shows that whether the incentive for quality improvement is higher or lower under international exhaustion depends upon whether price discrimination in international markets is cost or demand based. See also Valletti and Szymanski (2006) for related welfare analysis.

6 Appendix

Proof of Proposition 3

Part (i): $w_A(\text{ABC}) > w_A(\text{AB;C}) > w_A(\text{A;B;C})$. First note that

$$w_A(\text{AB; C}) - w_A(\text{A; B; C}) = \frac{(\mu_A - \mu_B)(\mu_B^2 + \mu_A\mu_B + 2\mu_A^2)}{4(\mu_A + \mu_B)^2} \geq 0 \text{ since } \mu_A \geq \mu_B$$

It is straightforward to show that $w_A(\text{ABC}) - w_A(\text{AB;C}) > 0$ iff

$$2\mu_A\mu_B - \mu_A\mu_C - \mu_B\mu_C > 0 \Leftrightarrow \mu_A(\mu_B - \mu_C) + \mu_B(\mu_A - \mu_C) \geq 0$$

which always holds.

Part (ii). We have

$$w_B(\text{AB; C}) - w_B(\text{A; B; C}) = \frac{(\mu_A - \mu_B)(\mu_A^2 + \mu_A\mu_B + 2\mu_B^2)}{4(\mu_A + \mu_B)^2} \geq 0 \text{ since } \mu_A \geq \mu_B$$

Part (iii): We can show $w_B(\text{A;B;C}) > w_B(\text{ABC}) >$ iff $\mu_A > \mu_A^w$ where $\partial\mu_A^w/\partial\mu_B > 0$ as follows. Let μ_A^w be defined by $\Delta w_B \equiv w_B(\text{ABC}) - w_B(\text{A;B;C}) = 0$. Differentiation establishes that $\frac{\partial \Delta w_B}{\partial \mu_A} > 0$ and $\frac{\partial \Delta w_B}{\partial \mu_B} < 0$. This implies that μ_A^w increases in μ_B since

$$\frac{\partial \mu_A^w}{\partial \mu_B} = -\frac{\frac{\partial \Delta w_B}{\partial \mu_B}}{\frac{\partial \Delta w_B}{\partial \mu_A}} > 0$$

Part (iv): This follows from $p_i(\text{ABC}) > p_{iC}(\text{A;BC}) > p_{iC}^*$ and the fact that country C's welfare is zero under regional integration [AB].

References

[1] Baudenbacher, C., 1998. Trademark law and parallel imports in a globalized world – recent developments in Europe with special regard to the legal situation in the United States. Fordham International Law Journal 22, 645-695.

[2] Ethier, W. J., 1998. Regionalism in a multilateral world. Journal of Political Economy 106(6), 1214–1245.

[3] Goldberg, P. K., 2010. Intellectual property rights protection in developing countries: the case of pharmaceuticals. Journal of the European Economic Association 8, 326-353.

[4] Grossman, G.M., Lai, E., 2008. Parallel imports and price controls. Rand Journal of Economics 39, 378-402.

[5] Li, C., Maskus, K.E., 2006. The impact of parallel imports on investment in cost-reducing research and development. Journal of International Economics 68, 443-455.

[6] Malueg, D.A., Schwartz, M., 1994. Parallel imports, demand dispersion, and international price discrimination. Journal of International Economics 37, 167-195.

[7] Maskus, K.E., 2000. Parallel imports. World Economy 23, 1269-1284.

[8] National Economic Research Associates, 1999. The economic consequences of the choice of regime of exhaustion in the area of trademarks. Final Report for DG XV of the European Commission.

[9] Pecorino, P. , 2002. Should the US allow prescription drug reimports from Canada? Journal of Health Economics 21, 699-708.

[10] Richardson, M., 2002. An elementary proposition concerning parallel imports. Journal of International Economics 56, 233-245.

[11] Roy, S., Saggi. K, 2012a. Equilibrium parallel import policies and international market structure. Journal of International Economics 87(2), 262-276.

[12] Roy, S., Saggi. K, 2012b. Strategic competition and optimal parallel import policy. Canadian Journal of Economics 45(4), 1369-1396.

[13] Valleti, T.M., 2006. Differential pricing, parallel trade, and the incentive to invest. Journal of International Economics 70, 314-324.

[14] Valletti, T.M., Szymanski, S., 2006. Parallel trade, international exhaustion and intellectual property rights: a welfare analysis. Journal of Industrial Economics 54, 499-526.