Pharmaceutical Innovation and Parallel Trade

Anna Rita Bennato
&
Tommaso Valletti

CCP Working Paper 12-9

Abstract:
This paper investigates the effects yielded by the interaction between government regulation policies and parallel trade, with a particular focus on the pharmaceutical sector. We provide a complete welfare analysis that accounts for both global investment decisions in R&D as well as local costly distribution of drugs. We study the patent holder’s decisions when a foreign government can introduce a direct price control to lower the price of patented drugs. We show that, under parallel trade, investment can rise only when the foreign government takes into full account its impact both on investment and on the firm’s decision to supply the regulated country. This arises because of a complete withdrawal from price regulation. The regulated country is however better off under an intermediate form of commitment whereby the foreign government anticipates its effect only on local distribution and delivery, but not on global R&D investment. In this case, the government resorts to some price regulation, which reduces investment in particular under parallel trade.

May 2012

Keywords: IPRs, parallel trade, pharmaceutical R&D, price regulation.

JEL Classification codes: F13; L12; O34

Acknowledgements: We thank Emeric Henry, Richard Friberg, Margaret Kyle, Yossi Spiegel, Stefan Szymanski, and seminar participants in Bristol, Mannheim, Paris, Rome, Stockholm, Tel Aviv, Toulouse and Zurich.
**Contact Details:**
Anna Rita Bennato: Centre for Competition Policy, University of East Anglia.

Tommaso Valletti: Imperial College London, University of Rome “Tor Vergata” and CEPR.
Pharmaceutical Innovation and Parallel Trade∗
Anna Rita Bennato† Tommaso Valletti‡
May 7, 2012

Abstract
This paper investigates the effects yielded by the interaction between government regulation policies and parallel trade, with a particular focus on the pharmaceutical sector. We provide a complete welfare analysis that accounts for both global investment decisions in R&D as well as local costly distribution of drugs. We study the patent holder’s decisions when a foreign government can introduce a direct price control to lower the price of patented drugs. We show that, under parallel trade, investment can rise only when the foreign government takes into full account its impact both on investment and on the firm’s decision to supply the regulated country. This arises because of a complete withdrawal from price regulation. The regulated country is however better off under an intermediate form of commitment whereby the foreign government anticipates its effect only on local distribution and delivery, but not on global R&D investment. In this case, the government resorts to some price regulation, which reduces investment in particular under parallel trade.

Keywords: IPRs, parallel trade, pharmaceutical R&D, price regulation.
JEL classification: F13; L12; O34.

∗We thank Emeric Henry, Richard Friberg, Margaret Kyle, Yossi Spiegel, Stefan Szymanski, and seminar participants in Bristol, Mannheim, Paris, Rome, Stockholm, Tel Aviv, Toulouse and Zurich.
†Centre for Competition Policy, University of East Anglia.
‡Imperial College London, University of Rome “Tor Vergata” and CEPR.
1 Introduction

Parallel imports are genuine goods produced under protection of a patent, trademark, or copyright, placed into circulation in one market, and then imported into a second market without the authorization of the owner of the intellectual property right (IPR). Parallel trade exists when there are significant price differences between countries, making this trade attractive. International price differences can be sustained only if IPRs are fully protected, making the creator the exclusive owner of her innovation: it seems quite obvious for the patent holder to exert market power by charging for the same good (or similar items) a different price in different markets. This form of third-degree price discrimination yields ambiguous welfare effects in the short run. The key aspect is typically whether price discrimination causes more markets to be supplied compared to a uniform pricing regime.\footnote{See Varian (1985), and Danzon (1997) for an application to the pharmaceutical industry.}

Policies at the international level support parallel trade when conducted among a group of relatively homogeneous countries (Malueg and Schwartz, 1994). Nevertheless, there is no unequivocal view about the implications of parallel trade, especially because of the trade-off between static and dynamic efficiency (Valletti and Szymanski, 2006). Reflecting this, the question of whether parallel trade should be permitted or inhibited still triggers hot policy debates in many countries. The pharmaceutical industry, which relies mostly on the patent system to protect its returns on innovation, emphasizes that parallel trade, or any reduction in the degree of protection of IPRs, could cause a fall in the pace of innovation, due to decreased private incentives to invest in R&D. For sure, the empirical relevance of parallel trade is undisputed.

This paper aims to analyze pros and cons of parallel trade, with a particular emphasis on the long-run implications for the pharmaceutical sector. We study how different regimes of IPRs interact with specific features of government intervention, namely price cap regulation. We develop an analysis based on the strategic interaction between a single innovative firm, based in the unregulated North, and a foreign government, located in the South. We adopt a framework where international exhaustions have real effects only when combined with other regulatory instruments, as demand elasticities between countries do not differ in our model. We demonstrate how the commitment propensity of the South government matters dramatically when its policy maker engages in drug price control. Indeed, large costs are associated with developing drugs and, by the time drugs reach the market, these costs are sunk. With that regard, identifying the effects of the government’s choices on the pace of innovation, we deal with the well known hold-
up problem. We investigate the advantages that the South government might obtain following a commitment strategy, distinguishing between different degrees of ability to commit. We draw a distinction between an ‘R&D investment’ stage, and a subsequent ‘market access’ stage where there is a further and costly delivery of drugs to the South. This distinction also captures the fact that the pharmaceutical company engages in some investment decisions that affect the markets globally (e.g., R&D) as well as some other decisions that are applicable only locally to some specific countries (e.g., delivery to the South). The extent of the impact of a price regulation scheme hinges crucially on its actual timing vis-à-vis these two stages.

The conventional wisdom that parallel trade is detrimental to profits and investment has recently been challenged by Grossman and Lai (2008). They show that, in a world where international exhaustion is permitted, the pace of innovation often is faster than in one with national exhaustion. More precisely, they consider that, where parallel trade is allowed at the international level, a foreign government has incentives to apply a less stringent price control of pharmaceuticals, because it recognizes that its policy has a global impact and fosters investments. In a world with two countries, both the innovative country and its trading partner can achieve benefits from parallel trade in terms of increased consumer surplus and a boost in the pace of innovation.

In our model, we also share the same feature as in Grossman and Lai (2008) that, under an international exhaustion regime which allows parallel imports, the price control in the South affects the price in the North as well, which can induce the South government to increase its controlled price. In addition, we introduce the idea that the consumers in the South are costly to serve, and therefore the South government will want its consumers to get access to the drug and be adequately served. Without any commitment (i.e., when the South government sets its regulated prices last, i.e., without taking into account neither the global nor the local investment decisions of the firm), parallel trade has no impact, since in any case the firm does not supply its good to the South. More interestingly, we show that parallel trade unambiguously reduces investment in a regime of ‘partial’ commitment, whereby the South government regulates the price prior to drug delivery in the South - but after investments have already occurred. It instead can increase investment only under ‘full’ commitment, when the South government moves first, before both global R&D and local delivery choices are made. When it moves first, the South government finds it so costly to elicit investment under parallel trade that it always prefers to withdraw from any price regulation. This leads to higher investment compared to when the South is insulated from the North and some price regulation would be applied. Our results therefore make precise the conditions and assumptions
that are needed for parallel trade to have beneficial long term effects. Importantly, we also find that a regime of ‘partial’ commitment yields the highest consumer surplus in the South as well as the highest global welfare. It follows that the South should find some credible way to achieve commitment to ensure local delivery, but not to the extent to anticipate its full effects on global R&D.

The remainder of the paper is as follows. In the next section we discuss international exhaustion and the derogation from IPRs. In Section 3 we present our model assumptions and describe the benchmark situation without parallel trade. Parallel trade is considered in Section 4. In Section 5 we extend the benchmark by studying the impact of price regulation. Finally, in the last section we summarize our results and conclude.

2 International exhaustion and parallel trade

In this section we analyze the economic issues concerning the exhaustion of property rights under the TRIPs agreement. The term “parallel” emphasizes the fact that genuine but unauthorized products are imported across country borders creating a parallel channel to the manufacturers’ authorized distribution. Even though parallel trade does not refer either to illegal or informal sector activities, or to trade in pirated or counterfeit goods, it is commonly referred to as “grey market”. Parallel trade represents one of the most controversial issues in the international trade-policy ground, and has raised difficult questions, especially in the pharmaceutical industry.

The legal status of parallel trade differs worldwide. Within the European Union parallel imports are a legitimate trade, despite that all European members recognize IPRs as established at the international level. The U.S. does not allow parallel trade in pharmaceuticals, while many Asian countries do, particularly in copyrighted products (Kyle, 2009). At the international level, a first attempt to find a solution to this disputed matter has been done during the Uruguay Round negotiations. Article 6 of the TRIPs agreement states that it is possible to resort to parallel trade by the exhaustion of IPRs, however ultimately the WTO has left each member country the possibility to fix its own regime for such exhaustion.

Parallel imports are in fact part of the “free trade” policy. Official European statistics show that in 2002 the total share of parallel imports reached 20% of the high-price pharmaceutical markets (Kanavos and Costa-Font, 2005).

This aspect has been stressed with the particular aim to provide developing countries affected by endemic diseases, such as HIV/AIDS, and malaria, the necessary policy to tackle their health problems. On the other hand, the U.S. government has recognized the possibility to prevent parallel trade from specific countries (Australia, Morocco, Singapore) by contractual means (Fink and Reichenmiller, 2005). This is also controversial, as preventing parallel trade by means of private contracts could be considered...
Some studies argue that parallel trade, where it is permitted, has not yielded the expected results in terms of convergence in price.\textsuperscript{4} Although several policy papers have been written, less attention has been paid on the long-run economic implications of parallel trade on IPRs.\textsuperscript{5} Scholars who believe that such arbitrage could erode IPRs, weakening the incentive for investment (e.g., Chard and Mellor, 1989; Barfield and Groombridge, 1998; Danzon and Towse, 2003; Li and Maskus, 2006), prefer Ramsey-type differential pricing as the best way to improve access to low-price drugs while still preserving investment in R&D. Complementary to this perspective, cross-national drug price differentials may not be based on demand elasticity, but on differences in other relevant demand factors (Maskus, 2000; Scherer, 2003). The interference of national governments in private markets by way of regulation of drug prices is a factor causing price differences at the international level (Pecorino, 2002; Jelovac and Bordoy, 2005; Bardey et al., 2010).

A more recent strand of the literature, to which our paper belongs, reassesses the role of parallel trade and focuses on the willingness to invest in R&D. This is particularly relevant, since normative results regarding parallel imports should ideally come from models in which innovation is accounted for. In the presence of parallel trade, welfare either increases or decreases depending on whether dynamic effects of parallel trade are examined (Valletti and Szymanski, 2006; Valletti, 2006; Grossman and Lai, 2008; Kyle and Rey, 2010).

An important aspect emerging from the literature is that the patent holder’s decision to export is endogenous. Therefore pricing regulations have a significant influence on the entry of firms into foreign markets, especially into less developed countries (Goldberg, 2010). These entry decisions depend on entry costs, as well as on the impact that local regulations might have globally. In our model we introduce explicitly the notion of local delivery costs in the South. The system by which drugs are supplied within a country is an aspect that has a key impact on the final price of drugs, and on their accessibility\textsuperscript{6} (WHO, 2002). Chaudhuri et al. (2006) stress the importance of weak distribution networks in India. They show that, even when multinational patent-holders

\textsuperscript{4}Parallel trade does not imply necessarily price convergence if consumers do not believe that the original drug and the parallel imported drug have the same value (Jelovac and Bordoy, 2005). Vertical distribution arrangements also can play a key role (Maskus and Chen, 2004). Empirical studies in the EU include Ganslandt and Maskus (2004), Kanavos and Costa-Font (2005) and Kyle (2007).

\textsuperscript{5}For a review of the literature see Szymanski and Valletti (2005).

\textsuperscript{6}Lack of public health infrastructures and services constitute an important barrier to the access to drugs for many developing countries (for more details see http://www.unmillenniumproject.org/documents/TF5-medicines-Complete.pdf).
enter developing countries, the distribution and marketing networks of multinationals are limited and costly, so that their products may not be reaching remote rural areas. They also argue that access to drugs and distribution coverage should be a crucial part of any welfare analysis. We follow their spirit and, by modelling (costly) access to drugs in the South, we are able to show the conditions under which price regulation and parallel trade interact strongly.

3 Model assumptions

There are two countries that we denote respectively as the North (N) and the South (S). In each country, consumers are heterogeneous, with preferences à la Mussa and Rosen (1978). Specifically, a consumer of type \( \tau \) that buys a product of quality product \( u \) at a price \( p \) enjoys a net utility given by:

\[
U(\tau) = \tau u - p, \tag{1}
\]

where \( \tau \) measures the consumer’s marginal valuation of quality. The taste parameter \( \tau \) is distributed uniformly over the interval \( \tau \in [0,1] \). Consumers can also decide not to buy any supplied good, and in this case they obtain their reservation utility, which is independent of type and normalized to zero. Since the lowest type is 0, in both countries there will be always someone who does not buy any product, unless it is offered for free.

Notice that, in contrast with previous literature, preferences in each country are identical, so that parallel trade cannot exploit differences in willingness-to-pay per se. We are therefore assuming that there are great disparities of income both in the North and in the South. Similar results would arise with alternative specifications that still resulted in the same elasticity of demand in both countries. This modelling choice is made to abstract from other aspects that have already been investigated by the literature, and thus make our contribution more transparent.\(^7\)

North and South differ in three important respects. First, the good is supplied by the patent holder who is based in the North. This is the only firm authorized to provide the patented good, both in the North and in the South market. By spending resources on R&D, the monopolist can improve the quality of its good, with the cost of quality, denoted as \( C(u) \), increasing at an increasing rate, \( C'(u) > 0 \) and \( C''(u) > 0 \). These are “global” costs which are incurred only at the investment stage, while all other costs at

\(^7\)Instead of North and South, we could have equally used West and East, or Domestic and Foreign. We keep the \( N \) and \( S \) notation because it is quite common in the literature this paper belongs to, as well as to highlight differences in distribution costs which are quite natural in the North/South context.
the manufacturing stage are set equal to zero.

The second difference between the North and South stems from distribution costs and access to health services. While the North has a system already in place for distributing, selling, and administering drugs, this does not hold for the South. In particular, we assume that, when a mass \( x \) of consumers is supplied in the South, there are some corresponding “local” costs defined as \( L(x) \), increasing at an increasing rate, \( L'(x) > 0 \) and \( L''(x) > 0 \). To obtain closed-form solutions, we employ the following function:

\[
L(x) = k \frac{x^2}{2},
\]

where \( k \) is a parameter that allows us to describe how costly it is to supply the South. In the North there is a unit mass of customers, while the mass \( x \) of consumers supplied in the South varies and is endogenously determined in equilibrium, as it depends on coverage costs. We will show that the impact of parallel trade depends quite crucially on the relative market size of North and South. Hence the specific role of \( k \) is to parametrize coverage costs and generate differences in the market size: a small value of \( k \) corresponds to a “large” South relative to the North, and vice versa.\(^8\)

In other words, we have in mind that, in the South, there is a certain mass of potential consumers who live distributed over a line, with unit density. The line represents how easy or difficult it is to supply and market drugs at that location (e.g., geographic access). Consumers at \( x = 0 \) are those in the biggest city, where it is very easy to supply them (e.g., because basic services are already in place), while to reach, supply and administer drugs to more people in remote regions becomes progressively more expensive for the monopolist. At each location, there is heterogeneity of taste according to (1) (i.e., rich and poor people live both in cities and in rural areas). A multi-dimensional screening problem, whereby \( \tau \) and \( x \) were somehow correlated, is beyond the scope of this paper.

The third difference concerns the role of governments. We assume that the government in the North does not regulate any aspect of drug production and consumption. The North has adopted a strong system of IPR that grants a patent to the monopolist for reasons that we do not model but just take as given. In contrast, we consider different approaches of the South government in relation to drug price control that we will further specify below. Hence the strategic players in our model are the monopolist firm and the South government.\(^9\)

\(^8\)Other than this, there is nothing peculiar about the specific functional form chosen for \( L(x) \) which we mainly keep for analytical simplicity.

\(^9\)We do not study a strategic trade game between the North and the South (see, e.g., Roy and Saggi, 2012a and 2012b), abstracting also from tariffs and subsidies. While the North government is
We proceed in developing the model in several steps. We study two different regulatory regimes on the exhaustion of IPRs. If parallel trade is banned, the firm can, in principle, set a different price in each market, because market segmentation is possible. However, if parallel trade is permitted, the firm is forced to set an identical price both in the North and in the South market, as it would otherwise attract arbitrageurs. Arbitrage is perfect and reimportation costs do not exist.

In the following Section, we first examine the simplest model where the South government is also passive and does not regulate drug prices, which are therefore freely set by the patent holder.

4 A benchmark: the irrelevance of parallel trade

Without parallel trade, perfect market segmentation is possible. Both in the domestic and in the foreign market, the patent holder behaves as a monopolist. We solve a two-stage game where the monopolist first decides on R&D, and then it sets the price in each market, as well as the coverage in the South.

In each market, there is a marginal type who is just indifferent between buying and not buying, defined as

$$\tau_i = p_i/u,$$

where \(i = N, S\). For future reference, it is also convenient to define consumer surplus in both countries, which is respectively

$$CS_N = \int_{\tau_N}^1 (\tau u - p_N) d\tau = \frac{(u - p_N)^2}{2u},$$

$$CS_S = x \int_{\tau_S}^1 (\tau u - p_S) d\tau = x \frac{(u - p_S)^2}{2u}. \tag{2}$$

In the last stage, the monopolist sets a price \(p_N\) in the North and a price \(p_S\) in the South to maximize its profits

$$\pi_N + \pi_S = \int_{\tau_N}^1 p_N \cdot \tau_N + x \int_{\tau_S}^1 p_S \cdot \tau_S - L(x)$$

$$= p_N (1 - p_N/u) + p_S (1 - p_S/u) x - k x^2 / 2.$$
It follows immediately that
\[ p_N = p_S = p^* = \frac{u}{2}, \]
with different profits in each country due to coverage differences. Indeed, in the North the monopolist makes a profit equal to \( \pi_N = \frac{u^4}{4} \) and in the South its profits are \( \pi_S = \frac{u^4}{4}x - k\frac{x^2}{2} \).

The optimal coverage of the South is also immediately derived and equal to
\[ x = \frac{u}{4k}, \]
which is increasing in quality, as gross profits at each location in the South also increase in quality.

In the first stage, the patent holder maximizes its global profits
\[ \Pi = \pi_N + \pi_S - C(u) = \frac{u}{4} + \frac{u^2}{32k} - C(u). \]

The monopolist thus offers both in the North and in the South a good having the same optimal quality \( u^* \), implicitly defined by
\[ \frac{1}{4} + \frac{u}{16k} = C'(u). \quad (3) \]

Since the monopolist already sets the same price everywhere, we obtain our first result: parallel trade, despite forcing the monopolist to set a uniform price in every market, has no impact. Thus the monopolist still charges \( \tilde{p} = p^* = u/2 \) everywhere, where we use the tilde sign for any regime with parallel trade.

**Proposition 1.** In the benchmark, parallel trade does not affect the investment decision, and consumer surplus and welfare also are invariant to the exhaustion regime.

The above analysis establishes our benchmark. Notice that we framed our approach in terms of a realistic two-stage game where investment choices are prior to the price setting. This timing is inconsequential though, since all decisions are taken by a single decision maker, and parallel trade does not affect optimal pricing. In the next Section we show how parallel trade and the precise timing of moves have instead real effects when the government in the South engages in drug price control.

---

10 We assume that the second-order condition is satisfied, for which we require that \( C'' > 1/16k \).
5 Price cap and commitment

In this section we analyze the effects produced by the introduction of price regulation in the South. Quite often governments regulate prices with the final aim to benefit consumers while still providing incentives to innovate.\(^\text{11}\) We develop our analysis by assuming that the South government has the ability to set a price cap in its own market. The price cap is chosen by a benevolent government with the aim to maximize the welfare of consumers in its own country alone. What will turn out to be critical for the analysis is the order of moves, which reflects also the South government’s commitment when setting the price cap. The complete sequence of moves is shown in Figure 1, where we already anticipate the three different levels of commitment the South government might have, corresponding to its intervention at different points of the timeline.

No commitment (NC) We start with the starkest example, where the South government has no commitment at all, and sets its regulated price in the last stage of the game, without anticipating its effects neither on global nor local investment decisions. Thus, as it is shown in the right branch of Figure 1, we consider the following timing of the game: first the firm invests in R&D, and successively decides the coverage of the South country, as well as the price in the North. Then, in the last stage, the South government sets the price in its own country.

It is immediate to show that the South government, once the good has been invented and delivered to the South, will always have an incentive to set its price as low as possible, that is, \(p_S = 0\), as we normalized to zero all manufacturing costs. The monopolist anticipates that no profits will be made in the South, so it decides not to cover any part of it. Global profits are made only from the North, \(\Pi = u/4 - C(u)\). The monopolist still invests, but an amount lower than before, as it is now \(1/4 = C'(u^{NC})\) and thus, compared to (3), it is obvious that \(u^{NC} < u^*\). Profits and consumer surplus decrease everywhere, especially in the South where there is no supply at all.

Notice that, once again, there is an irrelevance result for parallel trade. In fact, under parallel trade, if the firm supplied the South, the price regulated at zero would apply to the North as well, cannibalizing profits everywhere. Thus, under parallel trade, the firm will decide \textit{not} to supply the South market, hence achieving the same outcome as without parallel trade, though for a slightly different reason.

\(^{11}\)For an overview of theory and practice of price regulation in the pharmaceutical sector, see Danzon (1997) and Danzon and Chao (2000).
**Partial commitment (PC)** The previous case points to the fact the South government has to give incentives to the firm to be present in its own market, both with and without parallel trade. These incentives arise from restraining its ability to regulate prices and avoid hold-up problems. Therefore we now alter slightly the timing of the game, which is again in three stages. First, the firm chooses R&D. Then the government of the South sets its regulated price. Finally, the firm decides the coverage of the South market, as well as the price in the North. This timing endows the foreign government with some commitment capabilities, as in the second stage it acts anticipating the monopolist’s local coverage decision. The complete timing of the events corresponds to the middle branch of Figure 1.

Figure 1: *Sequence of moves for different commitment levels*

We start first with the case without parallel trade. In the last stage, the firm sets the monopoly price $p_N = u/2$ in the North, while coverage in the South is decided from maximizing $\pi_S = p_S(1 - p_S/u)x - kx^2/2$, that is

$$x = p_S \frac{u - p_S}{ku}.$$  \hspace{1cm} (4)

In the second stage, the South government sets the price cap $p_S = p_C$ that maximizes its consumer surplus, given by (2), anticipating the firm’s coverage reaction:

$$CS_S = x \frac{(u - p_C)^2}{2u} = \frac{(u - p_C)^3 p_C}{2ku^2},$$

which results in a price-cap of

$$p_C = \frac{u}{4}.$$
that is obviously greater than zero (as otherwise coverage would also be zero), but also lower than the unrestricted monopoly price. From (4), coverage is then \( x = 3u/16k \).

In the first stage, the monopolist maximizes the global profit

\[
\Pi = \pi_N + \pi_S - C(u) = \frac{u}{4} + \frac{9u^2}{512k} - C(u),
\]

from which we obtain a level of quality \( u^{PC} \) that satisfies \( u^{NC} < u^{PC} < u^* \), \(^{12}\) i.e., investment is higher compared to the previous case without any commitment, but lower than that in the unregulated case. It also follows that the cap is set at \( p_C = u^{PC}/4 < p^* \).

We now turn to parallel trade. In the last stage, the monopolist anticipates that the price set in the South will determine the price globally, and thus maximizes

\[
\pi_N + \pi_S = p_S(1 - p_S/u)(1 + x) - kx^2/2
\]

with respect to the coverage in the South, which still gives (4), exactly as in the case without parallel trade.

In the second stage, the South government in principle should still set the same price cap as without parallel trade, that is, \( p_S = p_C = u/4 \). However, as the cap applies everywhere, the government of the South must, in addition, ensure that the monopolist is willing to supply the product there. Recall that, in stage 2, quality has already been chosen and investment is sunk. Should the monopolist block sales to the South, it will then sell only in the North at a price \( p_N = u/2 \) with associated gross profits of \( u/4 \). Therefore the South government maximizes

\[
\max_{p_C} CS_S = \frac{(u - p_C)^3p_C}{2ku^2}
\]

s.t. \( p_C (1 - \frac{p_C}{u})(1 + x) - \frac{kx^2}{2} = \frac{p_C}{2} (1 - \frac{p_C}{u})[2 + \frac{p_C}{k} (1 - \frac{p_C}{u})] \geq u/4. \]

We can now state the following result.

**Proposition 2.** Imagine the South government can partially commit. (i) When the cost of supplying the South is low (\( k \leq k^{PC} \)), investment is higher than without commitment but lower than in the benchmark. Parallel trade reduces both investment and the price cap: \( u^{NC} < \tilde{u}^{PC} < u^{PC} < u^* \) and \( \tilde{u}^{PC}/4 = \tilde{p}_C < p_C = u^{PC}/4 < p^* \). (ii) When the cost of supplying the South is high (\( k > k^{PC} \)), parallel trade further reduces investment down to the same level as without any commitment, despite possibly setting a more lenient

\(^{12}\)The second inequality derives from comparing \( \partial \Pi/\partial u \) with (3), and noting that \( 18u/512k < u/16k \).
cap: $u_{NC} = \tilde{u}_{PC}^{P C} < u^{PC} < u^*$ and $\tilde{u}_{PC}^{P C}/4 < \tilde{p}_{C} = u^{PC}/4$, with $\tilde{p}_{C} > p_{C}$ for very high values of $k$.

Proof. The solution to (6) is simply the unconstrained solution $\tilde{p}_{C} = u/4$ (which is the same expression as without parallel trade, though the equilibrium qualities might differ) if the constraint is not binding, which can be rewritten as $\frac{3u}{16} + \frac{9u^2}{512k} \geq \frac{u}{4}$. This does not bind if $k$ is low enough. Otherwise it amounts to setting the lowest price that makes the firm’s participation condition just binding. The solution then is

$$\tilde{p}_{C} = \begin{cases} \frac{u}{4} & \text{iff } k \leq \frac{9u}{32}, \\ \frac{1}{2}(u - \sqrt{4ku + u^2 - 2u\sqrt{2k(2k + u)}}) & \text{iff } k > \frac{9u}{32}. \end{cases} \quad (7)$$

Under the parallel trade regime, $p_{N} = p_{S} = \tilde{p}_{C}$ implies a global profit

$$\Pi = \pi_{N} + \pi_{S} - C(u) = \begin{cases} \frac{u}{4} - C(u) & \text{iff } u < \frac{32}{9}k, \\ \frac{3u}{16} + \frac{9u^2}{512k} - C(u) & \text{iff } u \geq \frac{32}{9}k. \end{cases}$$

Compared to (5), the solution therefore is $u_{NC} < \tilde{u}^{PC} < u^{PC}$, and thus $\tilde{p}_{C} = \tilde{u}^{PC}/4 < p_{C} = u^{PC}/4$, when $k \leq k^{PC}$. When $k > k^{PC}$, it is $\tilde{u}^{PC} \equiv u^{NC}$. The function $\Pi$ is single-peaked in $u$ for low and for high values of $k$. Instead, for values of $k$ close to the threshold $k^{PC}$, it is double peaked. The optimal investment is found on the right branch of the profit function, until $k^{PC}$ is reached, where $k^{PC}$ is defined from $\frac{u_{NC}}{4} - C(u_{NC}) \equiv \frac{32u^{PC}}{16} + \frac{9\tilde{u}_{PC}^{P C}^2}{512k} - C(\tilde{u}_{PC}^{P C})$. For sure, $k^{PC} > \frac{9u_{NC}}{32}$. We also note that, close to the threshold value $k^{PC}$, then from (7) $\tilde{p}_{C}$ is set above but still very close to $u/4$ and the prevailing effect is the reduction in quality arising from lower profits in the North: $\tilde{p}_{C} < p_{C}$. When instead $k$ is very high the inequality is reversed. To show this, take the limiting case $k \to \infty$: from (5) without parallel trade it is $u^{PC} \to u^{NC}$, with $p_{C} = u^{PC}/4$. Under parallel trade, from (7), it is $\tilde{p}_{C} \to \tilde{u}^{PC}/2$, with $\tilde{u}^{PC} = u^{NC}$. Hence $\tilde{p}_{C} > p_{C}$ for high enough values of $k$. QED

Proposition 2 shows that there are now real effects from parallel trade, which are further investigated with the help of Figure 2. The four panels plot the differences of several key variables, with and without parallel trade, as a function of $k$, under a quadratic investment function, $C(u) = u^2/2$. These variables are respectively invested quality, price cap, consumer surplus in the South, and global welfare (defined as the sum of consumer surplus in the two countries and firm’s profits). Recall that the role of the parameter $k$ is to change the South coverage costs, and therefore vary endogenously the market size of the South relative to the North: low (high) values of $k$ corresponds to a
large (small) South in equilibrium.

We start the discussion when supplying the South is not very costly \( k \leq k_{PC} = \frac{9}{112} \approx 0.08 \)\(^{13}\) and participation is not at stake, since the monopolist is willing to supply the South even at the price cap, under either exhaustion regime. The unconstrained price cap set in the South applies globally under parallel trade, which depresses investment (top left panel). The price cap is set at the unconstrained level \( u/4 \) both with and without parallel trade, but parallel trade lowers \( u \) and thus the cap as well (top right panel). Consumers surplus in the South is lower under parallel trade because of the negative impact on quality (bottom left panel). Price regulation in the South benefits consumers in the (unregulated) North under parallel trade. However, for low \( k \), the South

\(^{13}\)This threshold value comes from equating, in the specific case with quadratic investment \( C(u) = \frac{u^2}{4} \), \[ \frac{u_{NC}^4}{4} - C(u_{NC}) = \frac{1}{12} \] with \[ \frac{32k_{PC}^2}{16} + \frac{9(g_{PC}^2)^2}{12k_{PC}^2} - C(\tilde{u}^P) = \frac{9k_{PC}^2}{12k_{PC}^2} - \frac{18}{112}. \]
is relatively larger, and therefore global welfare still largely follows the same behavior as consumer surplus in the South (bottom right panel).

On the other hand, when \( k \) is large (\( k > k^{PC} \)), incentives must be given to induce the firm to supply the (relatively small) South, as it is quite costly to do so. As shown by Proposition 2, parallel trade further reduces investment down to the lowest level as without any commitment at all (the difference with and without parallel trade shrinks as \( k \) increases, since the South becomes less and less relevant; see the top left panel for large values of \( k \)). The price cap always benefits consumers in the North which would otherwise be charged the monopoly price \( p^* \). This explains why, when \( k \) is sufficiently large and the North has a bigger weight, global welfare increases (again, the difference fades away when \( k \to \infty \), which is not shown in the range considered in the bottom right panel of Figure 2).

**Full commitment (FC)** We now consider the possibility that the foreign government behaves differently. With the purpose of increasing its reputation, the government of the South is committed to set a price regulation that anticipates its full effects not only on the local market coverage, but also on the global investment in R&D. Such precommitment is the timing specifically considered in Grossman and Lai (2008), although our models differ in several other respects. Hence the game now has the following timing. The government of the South gets to move first and sets the price-cap in its own market. Then the monopolist observes the price-cap and chooses the amount of R&D investment. Finally, the monopolist sets the price in the unregulated market (in the absence of parallel trade), as well as the market coverage in the South. The left branch of Figure 1 displays the complete timing of the game.

Without parallel trade, solving by backward induction, in the last stage the firm sets \( p_N = u/2 \) in the North achieving a profit \( \pi_N = u/4 \), while the South market coverage is the same as in the case with partial commitment, that is (4).

In the second stage, the monopolist chooses the optimal level of R&D by maximizing its global profits

\[
\Pi = \frac{u}{4} + \frac{[p_S(1 - p_S/u)]^2}{2ku^2} - C(u),
\]

from which ensues

\[
\frac{\partial \Pi(p_S, u(p_S))}{\partial u} = \frac{1}{4} + \frac{p_S^3}{ku^4}(1 - \frac{p_S}{u}) - C'(u) = 0,
\]

which characterizes implicitly the optimal investment \( u(p_S) \) as function of \( p_S \). We now
establish an intermediary result.

Lemma 1. Imagine the South government can fully commit. If the South government can force a price above the monopoly price, it can induce the monopolist to choose a quality which is above the level set in the unregulated benchmark. This is not possible if instead it can set only a price ceiling.

Proof. From (3) and (8) we need to compare only \( \frac{u}{16k} \) with \( \frac{p_S^3}{ku^2}(1 - \frac{p_S}{u}) \). This last expression is first increasing in \( u \) and then decreasing, and reaches a maximum when \( p_S = 2u/3 > p^* = u/2 \). The value taken at this maximum is \( \frac{8u}{8u} > \frac{u}{16k} \), and therefore the case with full commitment can generate a higher investment than \( u^* \) in case the marginal revenue intersects \( C' \) in this range. If instead a forced price is impossible, then at most the price can be \( p^* \), in which case marginal revenue with full commitment can at most coincide with the benchmark case. QED

It is somewhat surprising that investment can be made higher than in the unregulated benchmark: however, for this to happen, the South government must be able to “force” very high prices of the drug (which may not be in its own interest in the first stage of the game). This comes from the fact that the quantity sold at each location takes the expression \( 1 - p_S/u \). Hence an increase in \( u \) implies a clockwise rotation around the horizontal intercept at 1 when \( p_S = 0 \). For very low prices the impact of \( u \) is negligible, while it becomes more and more relevant the higher the price is. However, the Lemma also finds that, if the government can only set a cap (and the firm is free to set prices below the cap), then an investment higher than in the benchmark cannot be obtained. Since it not easy to find cases in practice where governments can force prices above the monopoly level, we do not consider this possibility any further below.

At the first stage, the maximization problem of the South government amounts to setting a price cap \( p_S = p_C \) to maximize

\[
\max_{p_C} CS_S = \frac{[u(p_C) - p_C]^3 p_C}{2ku(p_C)^2},
\]

from which we derive our next result.

Proposition 3. Imagine the South government can fully commit and there is no parallel trade. It is always more costly (i.e., a less stringent cap is needed) to elicit the same investment than under partial commitment. (i) When \( k > 3/16C'' \), it is always \( u/4 < p_C < p^* \); it is also \( u^{NC} < u^{FC} < u^{PC} < u^* \) if \( k \) is high enough. (ii) When \( k \leq 3/16C'' \), it is always \( u^{FC} = u^* > u^{PC} > u^{NC} \) and \( p_C = p^* \).
Proof. The first-order condition for the South is
\[
\frac{dC_S}{dp_C} = \frac{\partial C_S}{\partial p_C} + \frac{\partial C_S}{\partial u} \frac{du}{dp_C} = \frac{(u - p_C)^2(u - 4p_C)}{2ku^2} + \frac{(u - p_C)^2p_C(u + 2p_C)}{2k^3} \frac{du}{dp_C} = 0. \tag{9}
\]

Suppose first that the South government wants to get a certain level \(u\) of quality. From (9), the first term is zero for \(p_C = u/4\), which is the same cap set under PC. The sign of the second term depends on the sign of \(\frac{du}{dp_C}\).

By implicit differentiation of (8) we obtain
\[
\frac{du}{dp_C} = -\frac{\partial^2 \Pi / \partial u \partial p_S}{\partial^2 \Pi / \partial u^2} = \frac{u p_C^2(3u - 4p_C)}{ku^4C'' + p_C^2(2u - 3p_C)} > 0,
\]

since from the previous Lemma the price will be at most \(p^* = u/2\). Hence we obtain that \(p_C > u/4\). Alternatively, imagine that the South sets the same price cap as under PC, \(p_C = u/4\). Then from (8) the marginal incentive to supply quality is
\[
\frac{4}{1 + 16kC''(u_{FC})}
\]

which is strictly lower than the incentive, for the same price cap level, under PC, which is obtained from (5) and amounts to
\[
\frac{9}{16kC''(u_{PC})}.
\]

Consider now the impact on \(C_S\) when the South chooses the unregulated price, \(p_C = p^*\). As \(\frac{du}{dp_C}\) simplifies to \(\frac{4}{1 + 16kC''(u_{FC})}\), we obtain
\[
\left. \frac{dC_S}{dp_C} \right|_{p_C = p^*} = \frac{u(3 - 16kC'')}{8k(1 + 16kC'')}. \]

If \(k > 3/[16C''(u_{FC})]\), then \(p_C < p^*\). In the limit, as \(k \to \infty\), it is \(\frac{du}{dp_C} \to 0\) and \(p_C = u/4\). Together with the previous finding, this proves that \(u_{FC} < u_{PC}\) for very high values of \(k\). If instead \(k\) is close to \(3/[16C''(u_{FC})]\), then \(p_C\) is close to \(p^*\) and investment will approach \(u_{FC} = u^* > u_{PC}\).

If \(k \leq 3/[16C''(u_{FC})]\), then (9) is always positive at \(p = p^*\). Hence the South will set \(p_C = p^*\) and achieve \(u_{FC} = u^* > u_{PC}\). As a curiosity, making use of the previous Lemma, if the South government could “force” a price higher than the unregulated price, it would do so and achieve \(u_{FC} > u^*\). QED

Essentially, when \(k\) is small, investment is very responsive to price regulation, which becomes more lenient. In fact, we established that when \(k\) is low enough the South government would even theoretically want to force prices above the unregulated monopoly level. Since these cannot be enforced, the best the South government can do is not to regulate at all, in which case we fall back into the unregulated benchmark. Hence,
despite having a full commitment ability, the South uses it by withdrawing regulation entirely. If instead investment is not too responsive to price regulation (which happens when the relative size of the South is small as the cost of supplying it is very large), then the South government will want to set a binding price cap. In the limit, when $k$ is made arbitrarily large, it would set the same price cap as under partial commitment: however this will achieve a strictly lower investment because, under full commitment, R&D costs are not yet sunk and it is now more difficult to elicit investment.

Now we imagine that parallel trade is permitted. The effect of the price cap policy chosen by the South government affects the profits of the innovative firm also in the market of the North. The game takes the same timing as in the no parallel trade regime.

As above, at the third stage the monopolist defines the South market coverage with the aim to maximize its global profit. In the second stage, the firm determines its R&D investment according to the following maximization problem:

$$\max_u \Pi = (1 + x)p_C(1 - \frac{p_C}{u}) - k \frac{x^2}{2} - C(u)$$

subject to

$$x = \frac{p_C - p_C}{ku}.$$  

This leads to the following first-order condition

$$\frac{\partial \Pi(p_C, u(p_C))}{\partial u} = \frac{p_C^2}{u^2} + \frac{p_C^3}{ku^2}(1 - \frac{p_C}{u}) - C'(u) = 0,$$  

which characterizes the optimal investment $u(p_C)$ as function of price set by the government of the South.

In the first stage, the South government defines the price-cap, anticipating the monopolist’s investment decision in R&D while still ensuring delivery into the South. If the firm refuses to supply the South, the firm can ensure a payoff equal to the case without commitment, whereby only the North is supplied at the monopoly price. Thus, the problem of the government of the South becomes

$$\max_{p_C} C_S = \frac{[u(p_C) - p_C]^3 p_C}{2k(u(p_C))^2}$$

subject to

$$\Pi \geq \frac{u^{NC}}{4} - C(u^{NC}).$$

**Proposition 4.** Imagine the South government can fully commit and there is parallel trade. When $C(u)$ is not too convex, the South government always withdraws regulation, so that $\tilde{p}_C = p^*$ and $\tilde{u}^{FC} = u^*$. Parallel trade weakly increases investment compared to
its absence, and strictly so when $k$ is high enough.

Proof. The (unconstrained) first-order condition still takes the form as in (9) where $\frac{du}{dp_C}$ is now derived from (10), leading to

$$\frac{du}{dp_C} = \frac{up_C(2ku + 3up_C - 4p_C^2)}{ku^4C'' + 2p_C^2u - 3p_C^4 + 2p_C^4ku}.$$  

(12)

We obtain that, at the unregulated price, it is

$$\frac{dCS_S}{dp_C} \bigg|_{p_C=p^*} = \frac{u(8k + 3u - 16kuC'')}{8k(8k + u + 16kuC'')}.$$  

At $p_C = p^*$ we can also simplify the monopolist’s FOC (10) that becomes the same as (3) which we re-write as

$$k = \frac{u}{4[4C''(u) - 1]},$$

where it must hold that $C'' > 1/4$. We substitute this expression and finally get

$$\frac{dCS_S}{dp_C} \bigg|_{p_C=p^*} = \frac{(4C'' - 1)(12C'' - 1 - 4uC'')}{8C'' + 2 + 8uC''}.$$  

The first bracket at the numerator is always positive. The second bracket can be written as $4C'' - 1 + 4(2C'' - uC'')$, which is always positive as long as $C''$ is not too convex.\textsuperscript{14} At this price the participation constraint does not bind, since the monopolist earns profits as in the unregulated benchmark, strictly more than with no commitment. Notice that, even if it were feasible and the South tried to force a price above $p^*$ (but in the South alone), then parallel trade would make this policy ineffective as the price $p^*$ set in the North would apply in the South too. Hence the South government can just withdraw regulation and achieve the benchmark.

Contrasting these results with the previous Proposition 3, we derive that, when $k > 3/16C''$, parallel trade increases investment: $\tilde{u}^{FC} > u^{FC}$. Instead when $k \leq 3/16C''$, parallel trade achieves the same level of investment. \textbf{QED}

Our main result is that the effects of parallel trade change dramatically under partial and full commitment. Parallel trade can have the somewhat paradoxical effect of causing more rather than less investment \textit{only} under full commitment, and when $k$ is high enough. Indeed, when $k$ is low, the impact of the South on global investment is so large that it

\textsuperscript{14}For instance, if one uses a power function $C(u) = u^\alpha/\alpha$, a sufficient condition for the expression to be non-negative is $\alpha \leq 3$. 

19
is better for its government to withdraw any regulation, both with and without parallel trade. Parallel trade therefore has no impact as preferences are identical at every location in our model. It is when $k$ is high enough that parallel imports achieve more investment: when the South is insulated, its government would want to regulate the local price, reducing investment, compared to the case with parallel trade when it is optimal to leave the monopolist unregulated. For the same reason, parallel trade always increases the price cap when $k$ is high enough: only without parallel imports price cap regulation would in fact be binding.

In Figure 3 we plot again some key variables as a function of $k$, now for the full commitment case. The contrast with Figure 2 under partial commitment is quite stark. When the cost of supplying the South market is large ($k > 3/16 = 0.1875$), the government regulates its price only without parallel trade and elicits less investment. In this range, the North, which carries a big weight compared to a relatively small South, is
always left unregulated. Hence the North (and global welfare when $k$ is large) strictly benefits from parallel trade, not because the price regulation in the South is ‘exported’ into the North, but because of the investment effect.

Another interesting set of findings deals with the classical time inconsistency problem and the possible gains earned from commitment capabilities. While it is always convenient to forgo the fully discretionary policy with no commitment at all, it is immediate to show that partial commitment is better than full commitment for the South. This is particularly evident under parallel trade: under full commitment the best the South can do is to withdraw regulation, as otherwise too little investment would occur. Instead, under partial commitment, there is always a region of parameters such that the government prefers to set a binding cap, after investment has occurred. Since the unregulated price would also be always available under partial commitment (but is not chosen), a simple revealed preference argument implies that the South must be better off under partial commitment. More in general, under partial commitment, it is cheaper to elicit investments which allows the South government to regulate the price more strictly. The South government can guarantee to the monopolist ex post the same profits it would obtain serving the unregulated market only, without having to compensate for the R&D investment. Although we have not discussed the additional costs that the South government might have to incur in order to achieve commitment, it follows from our analysis that the South government should not unilaterally spend additional resources to achieve full commitment and move first.

Being able to commit before the firm’s choice on distribution is a point that has been highlighted in the policy literature. The ability of the government to be credibly committed in advance allows the monopolist to reduce the risk related to new R&D investments, promoting thus the disclosure of critical new treatments, e.g., vaccines (Berndt et al., 2007). In order to skip the government’s temptation to set a price which covers only the manufacturing costs and not the research costs, the foreign government needs to commit itself to a minimum price per dose of drugs, for a certain amount of individuals being immunized. Indeed this is what our treatment of market coverage is about, although we stress that the emphasis on ex ante global R&D incentives is only one part of the picture. Even ex post, local incentives for delivery have to be accounted for.

The lack of commitment technology is a widely recognized problem, and this is particularly true for the discovery of new treatments, which are considered as an international

\[\text{On this regard different proposals have been advanced. For more details see http://www.gavialliance.org/funding/pneumococcal-amc/ and http://www.yale.edu/macmillan/igh/}.\]
public good, and for that reason governments tend to free ride on the research promoted by others. We additionally stress that, in the definition of the optimal policy under parallel trade, the best outcome might be obtained adopting an intermediate, as so more feasible, degree of commitment.\textsuperscript{16}

The quality of the goods supplied is one of the foreign government’s goals, as a better quality level increases the consumer surplus, but pursuing this target, anticipating the monopolist’s R&D investment decisions, requires a more costly intervention. Indeed, setting a price prior to discovery means that the South government would be forced to meet the innovation cost, which makes price regulation a blunt instrument (to the extent it may not be employed at all). On the contrary, an intermediate level of government commitment bears advantages for the foreign government. It still ensures delivery, without renouncing the ability to cap the drug price. This protects only partially the innovative firm from the government’s opportunistic behavior, but it allows the South to guarantee a secure coverage of the regulated market, along with a certain quality level.

\section{Summary and conclusions}

The exhaustion of intellectual property rights introduced by the TRIPs agreement represents one of the most controversial issues in the debate over the protection of IPRs, especially in the pharmaceutical sector. The pharmaceutical industry claims that the use of these exhaustions are detrimental for the pace of innovation, because incentives to invest in R&D shrink. Providing a complete welfare analysis, this paper offers insights into the role of these international exhaustions.

We studied a stylized game between a monopolist, based in the North, and a foreign government, based in the South, and we considered the interdependence between parallel trade and the price regulation policies available to the South government. Our model is deliberately simplified, assuming identical preferences in each country, invalidating any effect of parallel trade in the absence of government regulation. We focused instead on the interaction between international exhaustion and price regulation and we accounted for investments. In particular, we drew a distinction between “global” R&D investments, which determine the quality of a drug everywhere, as opposed to “local” investments, which relate to supply decisions in the South. Because of weak infrastructure and skills, access to drugs for people living in the rural areas of the South is limited by high costs.

\textsuperscript{16}A successful example of this level of commitment is possible to be identified in the regional negotiations adopted by eastern Caribbean countries (for more details see \url{http://www.unmillenniumproject.org/documents/TF5-medicines-Complete.pdf}).
that can discourage the monopolist.

The model that we used is simple and tractable, yet quite rich in the results that it achieves. We discussed the consequences of public intervention under the assumption that the South government can credibly commit to its announced regulated prices with respect to the two types of investments, global or local. We discussed different timings of the game, capturing the extent to which the government is able to influence local market coverage (partial commitment), or even anticipate its impact on the quality level of the drug (full commitment). Indeed, the standard hold-up problem can be overcome if the government of the South has some commitment ability. Specifically, when the foreign government introduces price regulation to reduce the price of patented goods, we have found that parallel trade can produce contrasting effects on investment. Under the partial commitment case, when local delivery investments are accounted for by the South, but global R&D investment is still considered as sunk, parallel trade unambiguously reduces R&D investment. Since parallel trade ‘exports’ the regulated price also into the North, from which consumers in the unregulated North benefit, the net welfare assessment of parallel trade can still be positive. When the South government defines its price policy before the investment in R&D is set (full commitment), thus anticipating its effect both locally and globally, we have found that parallel trade plays no role if the cost to supply the South market is low: the South government always prefers to renounce any regulation in any case. If instead supplying the South market is costly, the government of the South still withdraws regulation when its price applies globally, while it would set a more stringent regulation if its market was insulated: this clarifies the circumstances under which parallel trade, despite weakening IPRs, can actually create higher incentives to conduct R&D. This higher investment also translates into a higher global welfare.

We conclude by emphasizing once more that the welfare implications of parallel trade cannot be fully understood if one omits from the analysis its interaction with the governments’ commitment capabilities. In our model, the government in the South faces different incentives for regulating prices when parallel imports are allowed by its trade partner in the North from when they are not. Parallel trade makes government policies interdependent and forces every government to consider the consequences of its actions on global incentives to invest. Therefore, a balanced approach towards the evaluation of the costs and benefits of allowing parallel imports should fully incorporate these strategic effects of the exhaustion regime on the level of both price and quality of drugs.
References


