# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l' Université Toulouse 1 Capitole
Discipline : Sciences Economiques

---

Présentée et soutenue par


**Anton Giulio MANGANELLI**


Le 11 septembre 2013

Titre :

# Essays on Cartels and Reverse Payments

---

JURY

Monsieur Bruno JULLIEN, D.R.CNRS, Université Toulouse 1 Capitole
Monsieur Gerard LLOBET, professeur, CEMFI Madrid
Monsieur Xavier MARTINEZ-GIRALT, professeur, UAB
Monsieur Patrick REY, professeur, Université Toulouse 1 Capitole

---

Ecole doctorale : Toulouse School of Economics
Unité de recherche : GREMAQ - TSE
Directeur de Thèse : Roberta DESSI'

# Table of Contents

# Abstract

This thesis consists of four essays on (real and alleged) cartels. The first one analyzes cartel pricing dynamics. It gives an unified explanation for the gradual rise of prices, their constant phase, price wars and cartel breakdowns. The driving forces are the consumers' reference-dependent preferences and the uncertainty over an external fringe's efficiency. Consumers are unwilling to pay much more than their (endogenous) expected price, so firms cannot set the monopoly price in the first period. As long as the fringe does not behave too efficiently, consumers become more pessimistic over the price, which allows firms to raise it gradually. This increasing price path is bounded from above by the presence of the fringe. If the fringe behaves efficiently during a sufficient number of periods, collusion is not sustainable, which triggers price wars and/or a breakdown of the cartel. Evidence from the Vitamin C cartel is consistent with the model. The second, third and fourth essays analyze the (neglected) pro-competitive effects of reverse payments in the pharmaceutical industry. Reverse payments are payments from a patent-holder to a generic manufacturer to settle a patent litigation over the validity or infringement of the patent. They typically delay generic entry and have often been treated as Antitrust violations *per se* by the European Commission and the FTC. In the second essay I show that this should not be the case when the parties' investment decisions are considered. Reverse payments increase the industry profits, which increases the generic manufacturer's incentive to invest. This increases the litigation rate, which increases consumer surplus. There also exists a tension in the originator's incentives to invest, absent from the patent literature, that can increase his investment too. The third essay analyzes the possibility of the generic entrant's bankruptcy. Banning reverse payments pushes the weak entrants out of the market, which reduces consumer surplus both before and after patent expiry. The overall impact of banning them can be negative and it is larger when the patent is strong, the economy is in a downturn and tacit collusion is sustainable among few players. The fourth essay analyzes the impact of reverse payments when the entrant's ability to compete is unknown and he may have to incur additional investments to be competitive. Reverse payments reduce the entrant's incentive to invest, which increases welfare when the investment cost is high enough. Under general conditions and several bargaining rules, reverse payments should be treated in a more lenient way when the technology is difficult. When the originator chooses whether to use a reverse payment, independently from how the surplus is split, and the technology is difficult, there is an irrelevance result: the cap on the generic entry date has no effect on welfare, as the negative effect of a later entry is exactly compensated by a lower probability of using a reverse payment.

# Acknowledgements

I want to thank Bruno Jullien and Patrick Rey for several interesting discussions and Roberta Dessì for being my supervisor. I am really grateful to Natalia Fabra for the semester spent at Universidad Carlos III and Joe Harrington for the time spent at Johns Hopkins University. Yassine Lefouili deserves a deep and special thank: without any obligation, he spent much time in reading my papers and helped me continuously in improving them. His time and comments have been invaluable to me and I am really indebted to him. I also thank Adrian Majumdar for the opportunity of working at RBB: the chapters on reverse payments come from there.

I also want to thank some friends that were close to me in these academic years: Elena, one of my best friends, who *inter alia* showed me the wonders of the London markets; Luis, my "master of life" in Barcelona; Luca and Gabriele, great friends and PhD-mates; my Italian friends in Toulouse Simone and Gian Piero (great mates also in the extra-academic life), Mattia, Lorenzo and Mario. I was very lucky in meeting them. I also thank André, with whom I always had great conversations, and Saniya. A special thank goes to my Spanish friends Carlos, Yiyi and Jose, who show that the good stereotypes over Spanish people are true.

I finally thank my parents for all their support.

# Introduction

This thesis focuses on real and alleged cartels. Cartels are considered the worst and most evident Antitrust crime. Their welfare effects are well known: a transfer of surplus from consumers to producers and a deadweight loss. Competition Authorities have been active in fighting them and they have have been searching for elements that signal a collusive industry. Some of these elements are: a gradual price rise, followed by a constant phase, price wars and cartel breakdowns. The theories addressing the first question have been developed by Harrington (2004, 2005) and Chen-Harrington (2006). The second question is addressed by Green and Porter (1984), who explain price wars through imperfect information between the firms in the cartel, Rotemberg and Saloner (1986), Kandori (1991) and Haltiwanger and Harrington (1991), who explain temporary price wars by assuming a difference between current and future demand. I try, in my first essay, to provide a theory that explains all these facts in a unique framework.

Afterwards, I consider an alleged anticompetitive behavior that is usually associated with cartels: reverse payments in the pharmaceutical industry. Reverse payments are payments from a patent-holder to a generic manufacturer to settle a patent litigation over the validity or infringement of the patent. They have often been treated as Antitrust violations *per se* by the European Commission and the FTC, as they normally delay generic entry until the patent expiration. I analyze the neglected pro-competitive effects of these payments, pointing out that they differ in substantial ways from the typical cartels.

The thesis is structured into four essays. The first one provides an unified explanation for the cartel facts. The second, third and fourth essays analyze reverse payments. The second essay considers the impact of reverse payments on the parties' incentives to invest and on consumer surplus. The third essay analyzes their impact on consumer surplus when the generic manufacturer can have liquidity problems. The fourth essay considers their impact on welfare when the generic manufacturer may have to incur additional investments to be competitive.

The first essay analyzes cartel pricing dynamics when a cartel is formed. It gives an unified explanation for the gradual rise of prices, their constant phase thereafter, price wars and cartel breakdowns. The driving forces are the reference-dependent preferences on the consumers' side and the uncertainty over an external fringe's efficiency. Differently from the previous literature about the gradual rise of prices, consumers are fully rational and understand which equilibrium is being played.[1] Consumers are unwilling to buy at a price much higher than their (endogenous) expected one, so firms cannot directly jump to the monopoly price.

[1] In Chen and Harrington (2006) the gradual rise of prices is due to the probability of being

The uncertainty over the fringe's efficiency shapes the consumers' price expectations: as long as the fringe does not behave too efficiently, consumers become more and more pessimistic over the price. This allows the firms to raise it gradually. This increasing price path is bounded from above by the fringe's presence. If the fringe behaves efficiently during a sufficient number of periods, collusion becomes not sustainable anymore, which triggers price wars and/or the breakdown of the cartel. Evidence from the Vitamin C cartel is consistent with the model: the world's largest producers kept maintaining cartel policies even after the entry of some small Chinese firms. These firms acquired more and more market shares over time (they behaved efficiently), until the cartel eventually broke down due to their competitive pressure.

The second essay moves on to the pharmaceutical industry and analyzes the relationship between reverse payments and incentives to invest. The Antitrust Authority can allow or ban them. If it allows them, it can also choose a latest entry date for the generic manufacturer. This is the same as setting a cap on reverse payments. Allowing reverse payments increases the industry profits, because the firms can agree on letting the originator be the monopolist until the latest entry date and share the additional profits through the reverse payment. This increases the generic manufacturer's incentive to invest. This increases the litigation rate when there is some ex ante uncertainty over the probability that the patent is invalid or not infringed. This, in turn, increases consumer surplus. Reverse payments delay entry but increase entry and litigation. There is also a tension in the originator's incentives to invest from allowing reverse payments, absent from the traditional patent literature. When the latest entry date is sufficiently late, allowing reverse payments makes the entrant offer a better settlement to the originator. The reason is that the higher industry profits make him less willing to risk litigation. This increases the originator's incentives to invest too. Under several parameter sets, allowing reverse payments increases consumer surplus.

The third essay shows another reason why reverse payments should not be banned *per se*. When the possibility of the generic manufacturer's bankruptcy is taken into account, banning reverse payments pushes the weak generic manufacturers out of the market. This reduces consumer surplus both before and after patent expiry through the reduction of the number of competitors. The negative effect of banning reverse payments is larger when the patent is strong, the economy is in a downturn and tacit collusion is sustainable among few players. The stronger is the patent, the better it is to allow reverse payments: a strong patent

fined, which depends on the difference between past and present prices. This creates an incentive for raising prices slowly, in order to avoid fines. Buyers are not rational players: they are just assumed to become suspicious when they observe anomalous pricing.

delays entry when the settlement does not involve a reverse payment, but the benefit of reverse payments (keeping the weak generic manufacturers on the market) is unchanged. The fact that economy is in a downturn is modeled through an increase in the probability that the entrant is weak. The larger the fraction of weak entrants, the better it is to allow reverse payments. Tacit collusion among few players also plays a role: if an additional manufacturer makes consumer surplus increases by much, it becomes more useful to allow reverse payments.

The fourth essay shows that reverse payments should not be banned *per se* when the entrant's ability to compete is unknown and he may have to incur additional investments to be competitive. The competitive ability of the entrant is his private information and the investment can occur only prior to the settlement stage. Allowing reverse payments reduces the entrant's incentive to invest and this saving can be higher than the welfare loss due to late entry. Allowing them with a later entry date further reduces his incentives to invest, as the equilibrium reverse payment will be higher. When the difficulty of the technology (the probability that the entrant is not competitive) is low enough, we have a pure strategy equilibrium where the originator offers a reverse payment and the entrant never invests. When the difficulty of the technology is high enough, we have a mixed strategy equilibrium where both the originator offers a reverse payment and the entrant invests with positive probability smaller than one. Reverse payments reduce the entrant's incentive to invest, which increases welfare for an investment cost high enough. Under general conditions and various bargaining rules, the threshold for the investment cost is decreasing in the difficulty of the technology, so reverse payments should be treated in a more lenient way when the technology is difficult. When the originator chooses whether to use a reverse payment and the technology is difficult, there is also an irrelevance result: the cap on the latest generic entry date has no effect on welfare, as the positive effect of an earlier entry is exactly compensated by a higher probability of settling through a reverse payment (independently from how the surplus is split).

To conclude, the first essay gives an unified interpretation for the most common cartel facts. The second, third and fourth ones deal with reverse payments in the pharmaceutical industry. They show that reverse payments should not be banned *per se* when a more complete picture is considered. Investments to be on the market, to be productive and liquidity problems give sufficient reasons not to ban reverse payments *per se*.

# 1 Chapter 1: Cartel Pricing Dynamics, Price Wars and Cartel Breakdown

I provide an unified explanation for three of the principal facts of the cartel literature: (i) prices gradually rise, then remain constant, (ii) some cartels eventually break down and (iii) there can be temporary price wars. Consumers have reference-dependent preferences (RDPs) and the efficiency of a competitive fringe is not publicly observable. In the best collusive equilibrium, RDPs may make consumers unwilling to buy at a price too high compared with their price expectations: firms' prices then increase over time together with consumers' price expectations. This increasing price path is bounded from above by the presence of the fringe. If the fringe sets a low price during a sufficient number of periods, collusion is not sustainable anymore, which triggers price wars and/or a breakdown of the cartel. Evidence from the Vitamin C cartel is consistent with the model.

## 1.1 Introduction

The analysis of discovered cartels in the last decades has shown that: (i) prices have a transitory phase during which they gradually rise and eventually remain constant; (ii) some cartels succeed in reaching stability, while others eventually break down; (iii) some cartels suffer temporary price wars. The first fact has been found in many of the largest discovered cartels, both for intermediate products[2], as well as for final products.[3] The second and the third fact are documented both from an empirical (Levenstein (1997)) and from a theoretical point of view (Green-Porter (1984), Rotemberg-Saloner (1986), Kandori (1991), Haltiwanger-Harrington (1991)).

[2] Like the Citric Acid and the Lysine cartel (see Connor (2001) and Levenstein-Suslow (2001)).

[3] Like the French mobile cartel (2000-2002, see http://www.autoritedelaconcurrence.fr/pdf/avis/05d65.pdf), the Italian pasta cartel and the German coffee cartel (see http://www.bundeskartellamt.de/wEnglisch/download/pdf/Fallberichte/B11-019-08-ENGLISH.pdf. This cartel served also bulk customers, like hotels and vending machine operators.

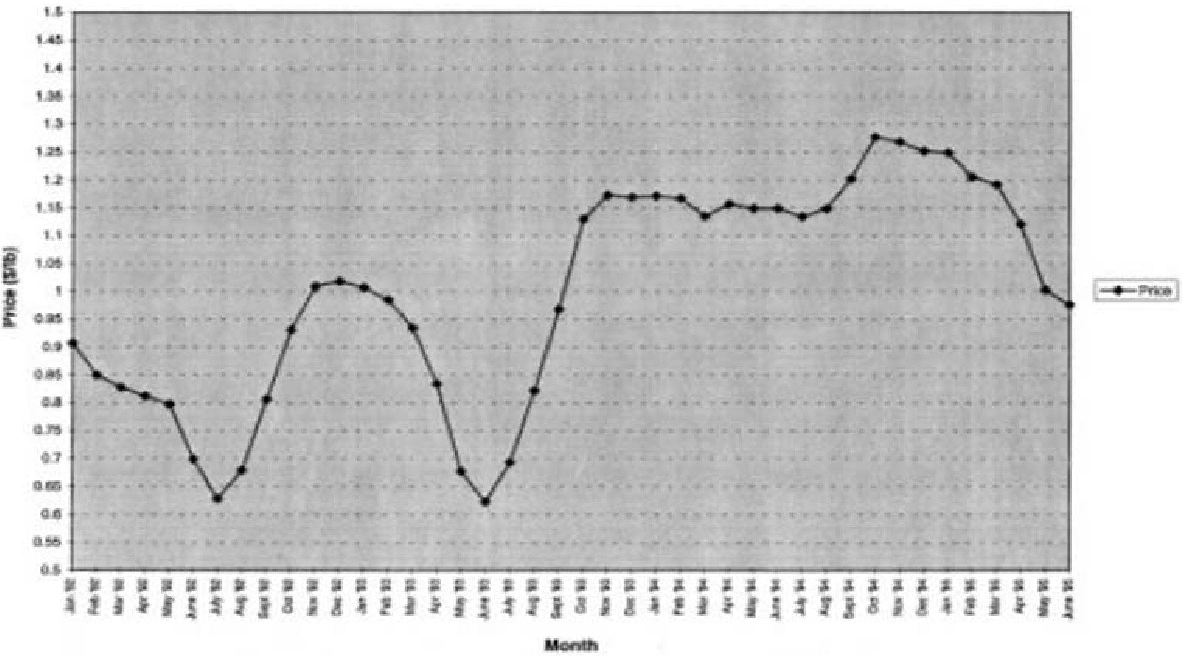



*Figure 1.* ADM U.S. lysine price, January 1992–June 1995.

Figure 1 shows the dynamics of lysine's price between 1992 and 1995. In June 1993 it reached the lowest level. In July the cartel was constituted and price began to rise until November 1993, when it remained stable during ten months. After another price rise, the cartel collapsed. The price dynamics of other cartels show similar patterns.[4]

The objective of this chapter is to answer to (i) why prices have not directly jumped to the maximal level, (ii) why cartels can suffer temporary price wars and (iii) why cartels can eventually break down.

The theories addressing the first question have been developed by Harrington (2004, 2005) and Chen-Harrington (2006). Harrington (2005) analyzes the impact of different Antitrust policies on cartel price dynamics. He shows that higher damage multiple and probability of detection lower the steady-state price, but also that the level of fines does not alter it. Furthermore, a more competitive benchmark to calculate damages can increase the steady-state price. All the results are derived by assuming that the incentive compatibility constraints (ICCs) are fulfilled and slack. Harrington (2004) generalizes the results above by allowing the ICCs to bind. First, when they bind, the cartel may first raise prices and then decrease them towards the steady-state level, in order to maintain the ICCs fulfilled. Second, Antitrust laws may have a perverse effect, as in some cases they

[4] The pasta cartel, for example, shows the same feature, though with a less clear dynamics, due to the high variance of durum wheat cost, the principal input in pasta production, over the cartelized period, .

allow the cartel to eventually price higher. This is due to the fact that the risk of being fined can stabilize a cartel and thereby allow it to set higher prices. Chen and Harrington (2006) motivate the increasing prie trend with the probability of being fined, which depends on the difference between past and present prices. This creates an incentive for raising prices slowly, in order to avoid fines. Buyers are not modeled as rational players and they are assumed to become suspicious when they observe anomalous pricing.

The second question is addressed by Green and Porter (1984), who explain price wars through imperfect information between the firms in the cartel, Rotemberg and Saloner (1986), Kandori (1991) and Haltiwanger and Harrington (1991), who explain temporary price wars by assuming a difference between current and future demand: this changes the incentive to deviate from one period to another, so firms may prefer to reduce prices during "booms" in order to keep collusion sustainable.

The present model gives an unified explanation for these facts, while assuming rational consumers. *Inter alia*, they understand which equilibrium is being played. Building on insights from the behavioral literature[5], I assume that consumers have reference-dependent preferences (RDPs). RDPs mean that consumers' utility depends on the comparison between the outcome and a reference point. Consumers have RDPs in the *price* dimension, i.e. price expectations directly enter their utility function: the higher the difference between the actual and the expected price, the higher the utility loss (and vice versa).

Evidence of RDPs is widespread both in the experimental and in the empirical literature. Thaler (1980) proposes this concept to explain why consumers often do not behave as consumer's theory predicts. Kahneman et al. (1991) enumerate a number of biases that are not explainable by traditional economic theory, among which loss aversion[6] (a particular case of RDPs), providing a series of experiments. Kahneman and Tversky (1991) discuss other experimental evidence and propose a model based on loss aversion that explains these biases by a deformation of the indifference curves about the reference point. Further evidence came along in the following years: with empirical data, Bowman et al. (1999) show the existence of loss aversion in saving decisions, Genesove and Mayer (2001) in the housing market and Haigh and List (2005) among professional traders. Novemsky and Kahneman (2006) and Gill and Prowse (2012) provide experimental evidence. Finally, Fox

[5] Kahneman and Tversky (1979), Erickson and Johansson (1985), Winer (1986), Kahneman, Knetsch and Thaler (1991), Chi-Kin-Jim and Kalwani (1992), Rotemberg (2004), Koszegi and Rabin (2006), Ellison (2006), Heidhues and Koszegi (2008).

[6] Loss aversion means that sensitivity to losses (w.r.t. a reference point) is greater than to gains. RDPs are a more general concept, as they just assume a positive sensitivity to losses and gains.

et al. (2007) investigate neural correlates of loss aversion and show that people typically exhibit greater sensitivity to losses than to equivalent gains.

Koszegi and Rabin (2006) and Koszegi and Heidhues (2008) build formal models on it. Koszegi and Rabin (2006) analyze consumer behavior with loss aversion and endogenous reference point and show that, when the outcome is uncertain, the willingness to pay increases in the expected price, conditional on purchase. Heidhues and Koszegi (2008) use loss aversion to explain the existence of focal prices[7] in a static game.[8] Koszegi and Rabin (2007), Macera (2009) and Gill and Stone (2010) consider dynamic games with loss aversion and endogenous reference point in consumption plans, labor contracts and tournaments, respectively.

I will instead develop an infinite-time horizon dynamic game with *endogenous reference point* that focuses on *collusion*. The basic ingredients are (i) RDPs and (ii) the existence of a competitive fringe. The reference point, i.e. the expected price for the present period, is updated using the information coming from the fringe. It can thus change in every period and consumers are fully rational.

The existence of small firms outside the cartel is common in cartel cases. The Vitamin C cartel is an example: the world's largest producers kept maintaining cartel policies even after the entry of some small Chinese firms. These firms acquired more and more market shares over time, until the cartel eventually broke down due to their competitive pressure.[9]

The reference point is based on an exogenous probability that the fringe, whose efficiency is unknown to both firms and consumers, sells the good at a low price. The fact that its efficiency is unknown is consistent with new firms that (i) have just entered the market, (ii) are more sensible than incumbents to exogenous shocks, (iii) employ the strategy "hit and run", etc. Consumers and firms have a common prior[10] about the two possible probabilities that the fringe draws a low marginal cost in the current period. If the fringe sets the high price, consumers and firms update their beliefs by giving more weight to the possibility that the fringe is inefficient. Thus, the larger the number of periods in which the fringe sets the high price, the more consumers expect a high price in the future - they become more and more pessimistic about the fringe's efficiency. Depending on the fringe's cost draws, the reference point can rise, which reduces the effect of RDPs, or fall, making the sustainability of collusion harder. Two key insights of the model are that: (i) the effect of RDPs can make firms just gradually raise prices,

[7] Equal prices across differentiated goods, even if their production costs differ.

[8] Heidhues and Koszegi interpret the reference point as the consumers' "lagged rational expectation", but, since the setting is static, the reference point is exogenous.

[9] For more information, see http://www.nd.edu/~mgrecon/datafiles/articles/vitamins.html.

[10] Nothing would qualitatively change if firms knew more than consumers about the fringe's efficiency, as long as they do not have a perfect knowledge. For the sake of exposition, I assume that their beliefs are the same.

so as to keep consumers willing to buy, and (ii) firms can sustain collusion if and only if the probability that the fringe is efficient is sufficiently low. As the fringe continues setting the high price, consumers update their price beliefs upwards, thus allowing firms to further increase prices. This process can continue up to the maximal collusive price, that is the high cost of the fringe. The transitory path is then over. On the other hand, if the fringe sets the low price during a sufficient number of periods, firms rationally anticipate that the fringe is too efficient to make continuation profits high enough to deter deviation. In this case, firms revert to Nash pricing. If, afterwards, the fringe sets the high price during a sufficient number of periods, firms can set the collusive price again. The Nash pricing periods can be interpreted as temporary price wars or a cartel breakdown; if the fringe does not set the high price during a sufficient number of periods again, firms keep setting the one-shot price - the breakdown of the cartel.

The chapter is organized as follows: Section 1.2 presents the mechanism underlying the cartel pricing dynamics. Section 1.3 explains temporary price wars and cartel breakdown. Section 1.4 shows some simulations. Section 1.5 analyzes the robustness of the model and discusses possible modifications. Section 1.6 concludes. All the proofs are in the Appendix.

## 1.2 Setup

Consider an infinite-time horizon Bertrand game. In each period, a unit mass of consumers has unit demand for a homogeneous good, produced by $n$ firms and by a competitive fringe. Consumers have the same willingness to pay, which depends on the intrinsic utility and RDPs. If they buy, they obtain:

$$u_t = \upsilon - p_t - \lambda(p_t - E[p_t]) \tag{1.1}$$

The parameter $\upsilon$ represents the maximal willingness to pay, absent RDPs; $p_t$ is the price paid in period $t$; $\lambda \geq 0$ is the RDPs coefficient (the higher $\lambda$, the higher the utility gain/loss in paying a price different than the expected one); $E_{t-1}[p_t]$ is the expected price in $t$, given the information available up to (the end of) period $t-1$. If they do not buy, they get $u_t = 0$.

Firms have a common and uniform marginal cost $\underline{c}$. The competitive fringe's cost can change in every period: with a probability $\mu$, it is $\underline{c}$, and with $(1-\mu)$ it is $\bar{c} > \underline{c}$.

Neither consumers nor firms know this probability: they know that the true probability $\mu$ is $\underline{\mu}$ with probability $\rho$ and $\bar{\mu}$ with probability $(1-\rho)$.

| | $\underline{c}_t$ | $\overline{c}_t$ |
|---|---|---|
| $\rho$ | $\underline{\mu}$ | $1-\underline{\mu}$ |
| $1-\rho$ | $\overline{\mu}$ | $1-\overline{\mu}$ |

Figure 2

The fringe sets a price equal to its cost. Clearly, the maximal price that collusive firms can set in equilibrium is $\bar{c}$, as this is the maximal price set by the competitive fringe.

The parameters ($\rho$, $\underline{\mu}$ , $\bar{\mu}$, $\underline{c}$, $\bar{c}$) are common knowledge. Firms' only strategic variable is price.[11] For firm $i = \{1, .., n\}$ and period $t$, each firm's strategy is $p^*_{it} : [\hat{\rho}_t] \rightarrow R_+$, where $\hat{\rho}_t$ is the common updated belief in $t$ that the true probability of a low price by the fringe is $\underline{\mu}$. Consumers' decision is from which supplier, if any, to buy. Their strategy can be described by $\sigma_t : [\hat{\rho}_t; p_{1t}, .., p_{nt}, p_{Ft}] \rightarrow [1, .., n, F, \varnothing]$ $\forall t$, i.e. they decide where to buy, if they do, given their belief $\hat{\rho}_t$ over the fringe's efficiency and the price the fringe and firms ask for. In each period, consumers buy the good with the lowest price, provided that their utility be at least equal to 0. We assume, for simplicity, that if $m \leq n$ firms and the fringe sets the same price, consumers just split their demand among the $m$ firms. So, denoting $u_{k,t}$ the consumers' utility from buying from firm $k$ in period $t$, firm $k$'s demand is $D_{k,t} = \frac{1}{m}$ if $p_{kt} = min\{p_{1t}, .., p_{nt}, p_{Ft}\}$ and $u_{k,t} \geq 0$, where $m$ is the multiplicity of firms with the lowest price, and $D_{k,t} = 0$ if $p_{k,t} > min\{p_{1t}, .., p_{nt}, p_{Ft}\}$ or $u_{k,t} < 0$. The stage game profits are $\pi_{k,t} = (p_{k,t} - \underline{c})/m$ if $p_{k,t} = min\{p_{1t}, .., p_{nt}, p_{Ft}\}$ and $u_{k,t} \geq 0$; $\pi_{i,t} = 0$ if $p_{k,t} \neq min\{p_{1t}, .., p_{nt}, p_{Ft}\}$ or $u_{k,t} < 0$.

In the following subsections I explain the timing of the game.

### 1.2.1 Timing

Denote $\hat{\rho}_{t,\tau,\hat{t}}$ the belief that the true $\mu$ is $\underline{\mu}$ at the beginning of period $t$, after $\tau$ periods of low fringe's price and $\hat{t}$ periods of high price.

In $t = 0$:

1. The probability that the fringe draws the low cost is $\mu \in \{\underline{\mu}, \bar{\mu}\}$, where $\mu = \underline{\mu}$ has probability $\rho$. Consumers form price expectation $E_0[p_1]$. The cost for the competitive fringe is $c_t \in \{\underline{c}, \bar{c}\}$ where $\underline{c}$ has probability $\mu$.

2. The fringe draws $c_t$ and sets its price $p_{Ft} = c_t$.

3. The firms observe $c_t$ and set their price.

[11]To keep things simple, I assume that firms cannot invest, for example, in advertising to influence consumers' beliefs.

4. Consumers observe prices and do their purchase decision.

5. Stage game payoff are realized.

In $t \geq 1$ all the steps are the same, except step 1 that becomes 1*) $c_t$ is redrawn according to the probability $\mu$ drawn in $t = 1$. Consumers update their beliefs $\hat{\rho}_{t,\tau,\hat{t}}$ by observing the price of the fringe.[12]

### 1.2.2 Firms' Problem, Price Expectations and Prices

In each period, the cartel chooses a price to maximize the expected sum of discounted profits. Denote $p_{Ft}$ the price of the fringe in period $t$, $p_t(\hat{p}_{t,\tau,\hat{t}})$ the price of each firm in period $t$ for a belief $\hat{p}_{t,\tau,\hat{t}}$ and $\bar{\pi}$ the expected profits. The cartel's problem is:

$$Max_{p_t}\pi_0(\rho, p_{F1}) + \sum_{t=1}^{\infty} \delta^t \bar{\pi}_t(\hat{p}_{t,\tau,\hat{t}}, p_{Ft}) \tag{1.2}$$

$$s.t. \sum_{t=0}^{\infty} \delta^t \pi_t(\hat{p}_{t,\tau,\hat{t}}, p_{Ft} = \bar{c}) > n\pi_1(\hat{p}_{t,\tau,\hat{t}}, p_{Ft} = \bar{c}) \tag{1.2a}$$

$$p_t \leq \bar{c} \ \forall t \tag{1.2b}$$

$$u_t \geq 0 \ \forall t \tag{1.2c}$$

The ICC (1.2a) has a role in the cartel breakdown and price wars. If the ICC is not fulfilled, firms would undercut the collusive price, so they just set the competitive price $p_t = \underline{c}$. Note that this ICC varies over time, as it depends on the updated belief $\hat{p}_{t,\tau,\hat{t}}$ over the fringe's efficiency $\mu$. I assume in the following that the ICC is fulfilled and I analyze in Section 1.3 when this is not the case. The constraint (1.2b) is due to the fringe: if firms set a price higher than the fringe's high cost, they sell nothing. This is the competitive constraint the fringe exerts on the cartel, which limits its price increase. The constraint (1.2c) is the consumers' participation constraint. In order to make the problem interesting, I do the following two assumptions.

ASSUMPTION 1. $\bar{c} < \upsilon$;

[12] I assume that firms know the current cost draw of the fringe because they have a good knowledge of the industry. All the results would still hold if firms' costs varied too, provided that their range is sufficiently smaller than the one of the fringe. Assuming this consists in assuming that firms are less vulnerable to industry-wide shocks, say because of economies of scale, a better knowledge of the industry etc. For simplicity I assume that firms' cost variation is zero. Results would also hold if the firms did not know the fringe's cost realization before setting their own price.

ASSUMPTION 2. (i) Perfect collusion is sustainable if $\mu = \underline{\mu}$ is common knowledge ($\rho = 1$), but not if $\mu = \bar{\mu}$ is common knowledge ($\rho = 0$). (ii) Perfect collusion is sustainable for the initial belief $\rho$.

I consider equilibria where, if a firm deviates, firms revert to Nash equilibrium forever (which constitutes the maximal punishment). In the best collusive equilibrium, in $t = 1$ consumers know that market price will be $p_t = \underline{c}$ if and only if the fringe sets $p_{Ft} = \underline{c}$. From the consumers' point of view, this occurs with probability $\rho\underline{\mu} + (1-\rho)\bar{\mu}$. If, on the other hand, $p_{Ft} = \bar{c}$, then firms set $p_1 = p_1^*$, where $p_1^*$ is the price that solves (2); this simply consists in maximizing the industry's current profits. The current firms' price, indeed, has no impact on the consumers' beliefs over the fringe's efficiency - and so neither on future prices. In the beginning of every period $t \geq 2$, consumers update their belief about $\mu$ given the information received from the fringe. The updated belief $\hat{p}_{t,\tau,\hat{t}}$ has a crucial role in the formation of price expectation.

**Lemma 1** *If consumers observe the fringe setting $p_{Ft} = \underline{c}$ during $\tau$ periods and $p_{Ft} = \bar{c}$ during $\hat{t}$ periods, independently from their order, the updated $\hat{p}_{t,\tau,\hat{t}}$ is*

$$\hat{p}_{t,\tau,\hat{t}} = \frac{\rho\underline{\mu}^{\tau}(1-\underline{\mu})^{\hat{t}}}{\rho\underline{\mu}^{\tau}(1-\underline{\mu})^{\hat{t}} + (1-\rho)\bar{\mu}^{\tau}(1-\bar{\mu})^{\hat{t}}} \tag{1.3}$$

**Proof.** See Appendix 1.7.1.

One can easily check that $\frac{d\hat{p}_{t,\tau,\hat{t}}}{d\hat{t}} > 0$ and $\frac{d\hat{p}_{t,\tau,\hat{t}}}{d\tau} < 0$. This Lemma explains how the belief $\hat{p}_{t,\tau,\hat{t}}$ over $\mu$ evolves over time, after any sequence of fringe's prices. This will be one of the two driving forces of the results.[13] In the best collusive equilibrium 1) each firm maximizes its own discounted stream of profits, given the fringe and the other firms' behavior, 2) consumers maximize their own utility, given the fringe and firms' prices, 3) players' beliefs are consistent with the equilibrium played.

In particular, point 3 means that consumers know whether firms are colluding or competing. This is a basic difference with Chen and Harrington (2006). There, they assume that consumers do not have correct beliefs about the equilibrium being played - they infer whether firms collude or not by observing price differences over time: the higher the difference between the current and the past price, the higher the consumers' belief over collusion. Here, on the other hand, consumers already know what equilibrium is being played.[14]

[13]The other one being the sustainability of collusion, analyzed in Section 2.

[14]The presence of an AA to which consumers can report their (correct) belief that firms are colluding would not change the results if hard proofs are needed to convict the cartel members.

In the *competitive equilibrium*, RDPs and the uncertainty over the fringe's efficiency have no impact over the equilibrium outcomes, which remain the standard Bertrand ones:

1) $p_{it}^* = \underline{c}\ \forall i, t$ 2) $D_{it} = \frac{1}{n}\ \forall i, t$, 3) $E_{t-1}[p_t] = \underline{c}\ \forall t$.[15]

In the best collusive equilibrium, when collusion is sustainable:

1) $p_{it}^* = p_t^*\ \forall i, t$, where $p_{it}^* = \tilde{p}_t := max\{p|u_t = 0, \bar{c}\}$, 2) $D_{it} = \frac{1}{n}\ \forall i, t$. When collusion is not sustainable, firms just maximize stage game profits: $p_{it}^* = \underline{c}\ \forall i, t$.[16] The difference between "competition" and "not sustainable collusion" is that, under the second, if the belief $\hat{p}_{t,\tau,\hat{t}}$ eventually becomes sufficiently high, firms can set a supracompetitive price again. This is the idea of temporary price wars: firms set $p_t^* = \underline{c}$ during some periods, as long as the belief $\hat{p}_{t,\tau,\hat{t}}$ over the fringe's inefficiency is too low. If $\hat{p}_{t,\tau,\hat{t}}$ never becomes sufficiently high again, firms keep setting the competitive price. For an observer, this is a cartel breakdown - even if there has been no deviation and firms would be willing to collude.

Any deviation from this makes firms revert to the competitive equilibrium forever (grim trigger strategy). $E_{t-1}[p_t]$ is given by the following Lemma; this will determine also the optimal price $p_t^*$.

**Lemma 2** *In the best collusive equilibrium, the expected price in t is:*

$$E_{t-1}[p_t] = [\hat{p}_{t,\tau,\hat{t}}\underline{\mu} + (1 - \hat{p}_{t,\tau,\hat{t}})\bar{\mu}]\underline{c} + [1 - \hat{p}_{t,\tau,\hat{t}}\underline{\mu} - (1 - \hat{p}_{t,\tau,\hat{t}})\bar{\mu}]p_t^* \tag{1.4}$$

*where* $t = \tau + \hat{t} + 1$.

**Corollary 1** *Combining (3) and (4), the expected price increases (decreases) after a period with high (low) fringe's price, that is* $\frac{dE_{t-1}[p_t]}{d\hat{t}} > 0$ *and* $\frac{dE_{t-1}[p_t]}{d\tau} < 0$ *.*

Every time consumers observe $p_{Ft} = \bar{c}$, they expect to pay a higher price in the future, because they (correctly) become more pessimistic over the fringe's efficiency. In other words, they give more weight to the lower probability that the fringe sets the low price. Therefore, the firms can raise the price while keeping consumers' participation constraint binding. The higher the number of periods during which consumers see the high fringe's prices, the higher the expected future price and so the actual firms' price (and vice versa).

The following Lemma gives the expression for the first period collusive price.

[15] $D_{it} = 1/n$ because we have assumed that, if the firms and the fringe sets the same price, consumers just buy from the firms. Removing this assumption has no impact on the results.

[16] Firms *cannot* temporarily reduce price to a supracompetitive level to make (1.2a) bind - this is impossible, as will be explained later.

**Lemma 3** *In the best collusive equilibrium, firms set the first period price equal to:*

$$p_1^* = \tilde{p}_t \equiv min\{\frac{\upsilon + \lambda[\rho\underline{\mu} + (1-\rho)\bar{\mu}]\underline{c}}{1 + \lambda[\rho\underline{\mu} + (1-\rho)\bar{\mu}]}), \bar{c}\} \tag{1.5}$$

**Proof.** See Appendix 1.7.2.

Call $\hat{p}_1 = \frac{\upsilon+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]\underline{c}}{1+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]}$ (the first argument in the *min* operator) the "transitory price" for period 1. A collusive price has a transitory phase if $\frac{\upsilon+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]\underline{c}}{1+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]} < \bar{c}$. The following Proposition gives a condition under which collusive prices have a transitory phase.

**Proposition 1** *Collusive prices have a transitory phase if and only if customers' willingness to pay $\upsilon$ is not too large:*

$$\bar{c} < \upsilon < \bar{c}(1+\lambda) - \frac{\lambda[\rho\underline{\mu} + (1-\rho)\bar{\mu}](\nu + \lambda[\rho\underline{\mu} + (1-\rho)\bar{\mu}]\underline{c})}{(1 + \lambda[\rho\underline{\mu} + (1-\rho)\bar{\mu}]) + [1 - \rho\underline{\mu} - (1-\rho)\bar{\mu}]\underline{c}}$$

**Proof.** See Appendix 1.7.3.

The intuition is that if the willingness to pay $\upsilon$ is too large, firms can directly jump to the maximal collusive level - the high fringe's cost. When $\upsilon$ is not that big, firms can only raise prices smoothly, as long as the fringe sets the high price.

**Lemma 4** *The transitory phase price $\hat{p}_1$ depends negatively on (i) the probabilities $\underline{\mu}$ and $\bar{\mu}$ that the fringe is efficient, (ii) the size of RDPs $\lambda$ and positively on the belief $\rho$ that the fringe is inefficient.*

If consumers believe that the probabilities that the fringe sets the low price are high (via $\mu$ or $\rho$), then they expect a low price. This makes them suffer a high utility loss when they pay a high price: this, in turn, forces firms to lower their price in order to the consumers' participation constraint binding. A high RDP parameter $\lambda$ yields an intuitive result: $\hat{p}_1$ will be smaller for any price expectation, as a higher $\lambda$ means that consumers suffer a larger utility loss. This forces firms to lower prices too. When $t \geq 2$, through Bayesian updating we get the following proposition .

**Proposition 2** *In the best collusive equilibrium, after $\tau$ low and $\hat{t}$ high prices from the fringe, the collusive price firms set in period $t$ is*

$$p_t^* = \min\{\frac{\upsilon + \lambda[\frac{\rho\mu^\tau(1-\underline{\mu})^{\hat{t}}}{\rho\mu^\tau(1-\underline{\mu})^{\hat{t}} + (1-\rho)\mu^\tau(1-\bar{\mu})^{\hat{t}}}\underline{\mu} + (1 - \frac{\rho\mu^\tau(1-\underline{\mu})^{\hat{t}}}{\rho\mu^\tau(1-\underline{\mu})^{\hat{t}} + (1-\rho)\mu^\tau(1-\bar{\mu})^{\hat{t}}})\bar{\mu}]\underline{c}}{1 + \lambda[\frac{\rho\mu^\tau(1-\underline{\mu})^{\hat{t}}}{\rho\mu^\tau(1-\underline{\mu})^{\hat{t}} + (1-\rho)\mu^\tau(1-\bar{\mu})^{\hat{t}}}\underline{\mu} + (1 - \frac{\rho\mu^\tau(1-\underline{\mu})^{\hat{t}}}{\rho\mu^\tau(1-\underline{\mu})^{\hat{t}} + (1-\rho)\mu^\tau(1-\bar{\mu})^{\hat{t}}})\bar{\mu}]}, \bar{c}\} \tag{1.6}$$

The left term in the *min* operator is the transitory phase price $\hat{p}_t$ for a generic period $t$. When $p_t^* = \hat{p}_t$, the consumers' participation constraint (1.2c) is binding and the fringe's constraint (1.2b) is not. The price $\hat{p}_t$ rises with $\hat{t}$ because, after every $p_{Ft} = \bar{c}$, consumers update their beliefs $\hat{p}_{t,\tau,\hat{t}}$ upwards. This makes their utility loss smaller and allows firms to increase their price while keeping the consumers' participation constraint (1.2c) binding. When $p_t^* = \bar{c}$, instead, the price has reached its maximum level: now the binding constraint is the fringe's one (1.2b) and, typically, the consumers' participation constraint (1.2c) does not.[17]

The gradual rise of prices is a phenomenon that occurs when the fringe sets the high price and firms collude. If the fringe sets the low price, market price suddenly falls. Afterwards, if the fringe sets the high price again, market price gradually rises again - consumers buy from the cartel for any price smaller than $\bar{c}$ that fulfills $u_t \geq 0$. So, there is an asymmetry in the market price's behavior: under collusion, the price slowly rises when the fringe sets the high price, but it falls quickly when the fringe sets the low price. I discuss this in the next section.

## 1.3 Sustainability

Up to now this theory could have applied to a monopoly too, as the no-deviation ICC (1.2a) is assumed to be fulfilled. By the first part of Assumption 2, we know that the collusive price is sustainable if and only if $\hat{p}_{t,\tau,\hat{t}} \geq \rho^*$, i.e. if and only if the belief about the inefficiency of the fringe is sufficiently high. Finding a necessary and sufficient condition is very cumbersome[18] and we shall be content of having both a sufficient condition and a necessary condition that assure that the ICC is fulfilled. Here I analyze them and describe what occurs when they are violated.

**Lemma 5** *In a generic period t, a sufficient condition to fulfil the no-deviation ICC (2a) of the firms' maximization problem is:*

$$\hat{p}_{t,\tau,\hat{t}} \geq \rho_S^* = \frac{n(1-\delta) - 1 + n\delta\bar{\mu}}{n\delta(\bar{\mu} - \underline{\mu})} \tag{1.7}$$

**Proof.** See Appendix 1.7.5.

By the second part of Assumption 2, the initial $\hat{p}_{t,\tau,\hat{t}} = \rho$ is above $\rho^*$. Then collusive price is sustainable and firms set the price as explained up to now. Every

[17] The fact that $p_t^*$ reaches $\bar{c}$ depends on the parameters. If $\mu$ is sufficiently high, even an arbitrarily high $\hat{t}$ will not make consumers so pessimistic that they accept to pay $p_t = \bar{c}$. In this case, prices rise forever, at a slower and slower pace. For more details, see the Appendix.

[18] A necessary and sufficient condition is difficult to obtain because the cap $\bar{c}$ on prices makes expected future prices weakly decreasing over time. In other words, $\hat{p}_{t,\tau,\hat{t}}$ is the best estimate for the fact that $\mu$ is the true probability, but $p_t^*$ is not the best estimate for future prices, because of the cap $\bar{c}$.

time the fringe sets the low price, $\hat{p}_{t,\tau,\hat{t}}$ is updated downwards - see (1.3). Intuitively, after a certain number of low fringe's prices, $\hat{p}_{t,\tau,\hat{t}}$ can become so low that the collusive price is not sustainable anymore. In that case, as long as $\hat{p}_{t,\tau,\hat{t}}$ is smaller than $\rho^*$, firms revert to the static Nash equilibrium price, because they recognize that each one has an incentive to deviate.

**Lemma 6** *In a generic period $t$, a necessary condition to fulfil the ICC of the firms' maximization problem is:*

$$\hat{p}_{t,\tau,\hat{t}} \geq \rho_N^* = \frac{n(1-\delta) - 1 + \delta\bar{\mu}}{\delta(\bar{\mu} - \underline{\mu})} \tag{1.8}$$

**Proof.** See Appendix 1.7.5.

It is easy to check that $\frac{d\rho_N^*}{d\delta} = -\frac{(n-1)}{\delta^2(\bar{\mu}-\underline{\mu})} < 0$ and $\frac{d\rho_S^*}{d\delta} = -\frac{(n-1)}{\delta^2 n(\bar{\mu}-\underline{\mu})} < 0$, so a higher $\delta$ makes $\rho_N^*$ and $\rho_S^*$ smaller. The more patient the firms are, the more the fringe can be efficient without the cartel breaking down.

When collusion is not sustainable, firms' prices can eventually become higher than costs again only if the fringe sets the high price during a sufficient number of periods - a number of periods such that $\hat{p}_{t,\tau,\hat{t}}$ becomes bigger than $\rho^*$. However, differently from Rotemberg and Saloner (1986) and Haltiwanger and Harrington (1991),[19] when firms set supracompetitive prices, future prices are expected to be smaller or equal than present ones. The reason is that in expectations $\hat{p}_{t,\tau,\hat{t}}$ does not change over time - so prices are expected to be the same over time - and the fringe limits the possible price increases - making future expected prices possibly smaller. Reducing prices, therefore, does not make the ICC easier to fulfill, because future expected profits are smaller or equal than current ones for any $\hat{p}_{t,\tau,\hat{t}}$. So firms, in the best collusive equilibrium, will just switch from the collusive price to the competitive price and vice versa.[20] This consists in temporary price wars: they are not, here, due to imperfect monitoring or to strategic price reductions to keep collusion sustainable, but to the fact that firms recognize that each one would have an incentive to deviate.

[19] In Rotemberg and Saloner (1986) demand is determined in an i.i.d. fashion in every period, so future demand is independent from the present one. They show that the ICC is more difficult to fulfill when the present demand is high, because the incentive to deviate is higher. In Haltiwanger and Harrington (1991) the evolution of demand is deterministic and they show that collusion is more difficult to sustain when future demand is low, because the foregone profits are lower. In both these models future demand is expected to be different from the present one, which creates the different incentives to deviate depending on the present state. In my model, when firms set supracompetitive prices, future prices are expected to be not higher than current ones, so there is no room for temporary price cuts to reduce the temptation to deviate and "wait for better days".

[20] When collusive price is not sustainable, consumers just expect $E_{t-1}[p_t] = \underline{c}$; when a high fringe's cost makes collusive price sustainable, the price expectation is again the one in (1.4).

All this analysis leads to the following Proposition:

**Proposition 3** *There exist parameter sets for which, in the best collusive equilibrium:*

*1) prices have a transitory and then a constant phase,*

*2) firms can temporarily switch to competitive pricing,*

*3) the cartel can break down.*

The following section provides a numerical example for each of these facts.

## 1.4 Numerical examples

Denote $\delta_N^*$ the minimum discount factor such that the necessary condition (1.8) is fulfilled for $\hat{p}_{t,\tau,\hat{t}} = 1$. Assume $n = 2$, $v = 10$, $\lambda = 2$, $\bar{c} = 7.68$, $\underline{c} = 2$, $\rho = 1/4$, $\bar{\mu} = 1/2$, $\underline{\mu} = 1/8$ and that $\delta$ satisfies Assumption 2.

It is easy to verify that the conditions of Proposition 1 are fulfilled, so collusive prices cannot directly jump to $\bar{c}$. The following figures show how the price pattern looks like for different realizations of the fringe's cost.

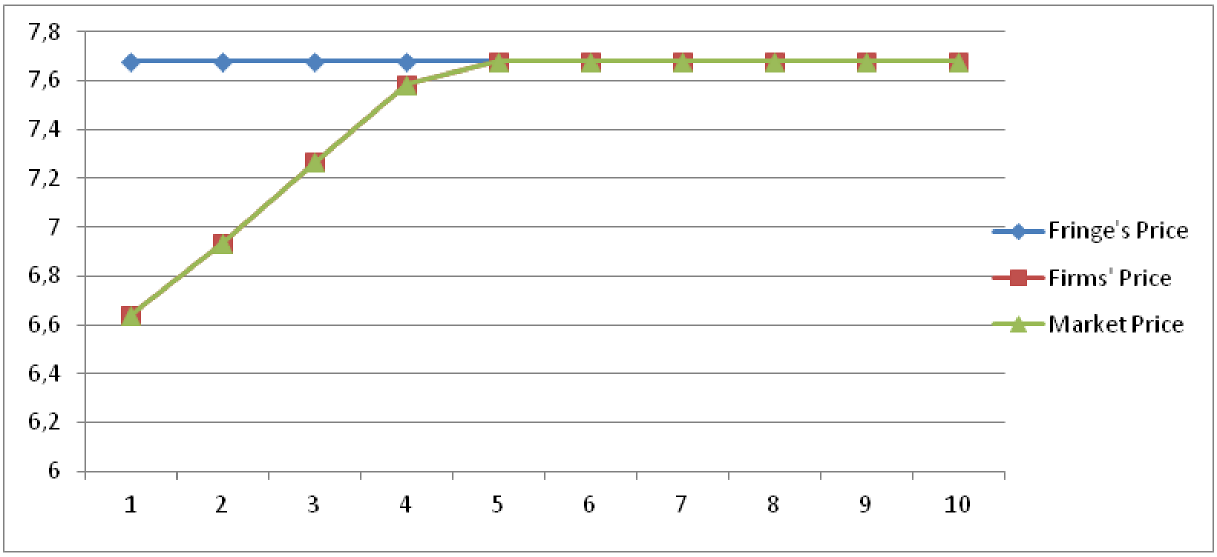



Figure 3. Simulation 1 (successful cartel): the fringe always sets the high cost $p_{Ft} = \bar{c}$. Price rises until period $t = 5$ and then remains constant and equal to $\bar{c}$.

Simulation 1 represents the case of a cartel that always sells the good from period 1 to 10. When the fringe always sets $p_{Ft} = \bar{c}$, price rises from period 1 to period 5 and then remains constant. Between periods 1 and 4 it grows more quickly than between periods 4 and 5: the reason is that between periods 1 and

4 the binding constraint is the consumers' participation constraint (1.2c), while from period 5 onwards the binding one is the fringe's one (1.2b). From period 1 to 4 firms make consumers pay the price that makes them indifferent between buying or not ($u_t = 0$ for $t = 1, 2, 3, 4$). Since period 5 onwards consumers are so pessimistic that they lose so little utility in paying a high price that they are willing to pay the maximal price $p_t = \bar{c}$ too. From that period on, the force that constrains firms' price is the competitive fringe.

Figure 3 resembles the actual rise of lysine's price of Figure 1: there, we have (i) some periods (from July to October 1993) in which price rose at an almost constant rate, like in periods 1-4 of this simulation; then (ii) a much smaller increase in November, like in period 5 of the simulation, and finally (iii) an almost constant price during other ten months, like in periods 6-10; we have see the same behavior, i.e. some periods of the transitory phase at an almost constant pace and thereafter a constant price.

The next figure shows a price pattern for the cartel breakdown.

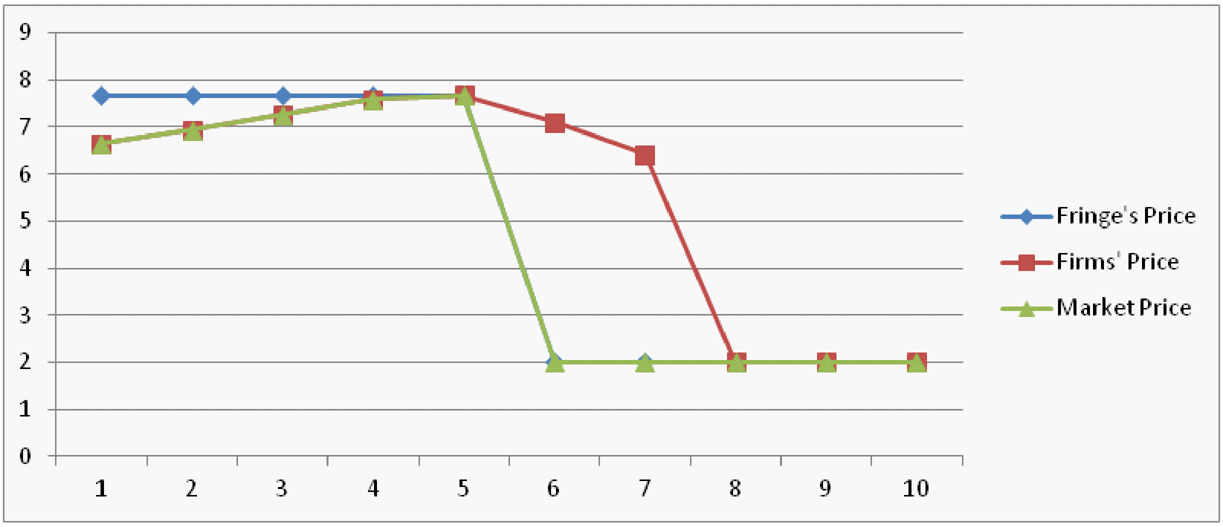



Figure 4. Simulation 2 (cartel breakdown): after some periods of $p_{FT} = \underline{c}$, the collusive price is not sustainable anymore. Firms initially just lower the price; afterwards, firms set the competitive price.

In this simulation the fringe sets the high price during five periods and then, during the following five periods, it sets the low price. Firms' price rises in the first five periods in the way explained in simulation 1; then it gets lower, after every period of low fringe's price. At some point (depending on $\delta$), $\hat{p}_{t,\tau,\hat{t}}$ is so low that the ICC is not fulfilled, so collusive price is not sustainable and firms switch back to competitive pricing.

The following figure shows a price pattern for a temporary price war.

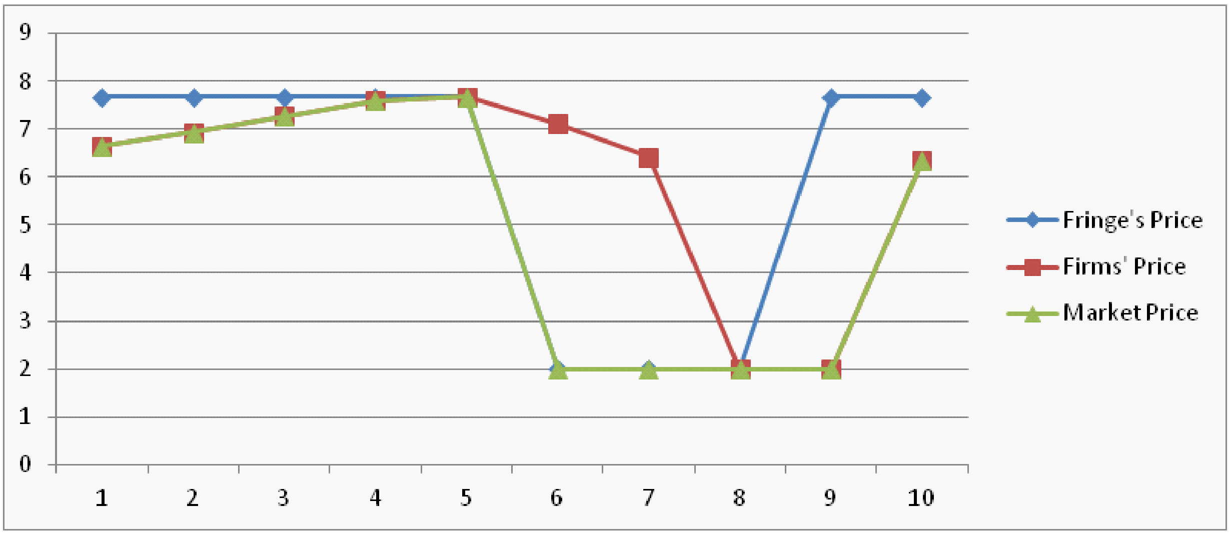



Figure 5. Simulation 3 (price war): after some periods of low fringe's price, collusion is not sustainable anymore, but then it becomes sustainable again.

Simulation 3 shows the price path of a temporary price war. After some periods of low fringe's price, collusive price is not sustainable anymore, like in simulation 2; here, in period 9 the fringe sets the high price again, which makes $\hat{p}_{t,\tau,\hat{t}}$ higher, but still not sufficient to make the collusive price sustainable. A further period of high price makes $\hat{p}_{t,\tau,\hat{t}}$ sufficient to make it sustainable again, so firms increase their price again according to the collusive price equation (1.6). As we can see from the graph, there is a lag between the fringe's price and the firms' prices. The firms' price follows the fringe's price: when the fringe's price falls, the firms' price still remains to relatively high levels (collusion remains sustainable, price falls slightly to keep consumers on their $u_t = 0$ level[21]) and only after some periods it falls to the competitive level. When the fringe raises its price, the firms' price may remain during some periods to the competitive level, until $\hat{p}_{t,\tau,\hat{t}}$ raises to a level sufficient to make collusion sustainable again. This resembles the price behavior of lysine in the beginning of the cartel period, i,.e. around and few months before June 1993. In that month the price fell down to the minimal level and then rose again.

[21] Firms set the price that would keep consumers to their zero-utility level, but consumers buy from the fringe. However, firms have no reason to deviate from this price, given that any deviation to $p > \underline{c}$ would yield no profit, and any deviation to $p < \underline{c}$ would yield negative profits. If goods were differentiated, firms' price reduction would be stronger, but the qualitative results would remain the same.

## 1.5 Robustness and discussion

In this section I discuss how results are qualitatively robust to a number of variations. The driving forces for the increasing prices are the RDPs and the uncertainty over the fringe's efficiency $\mu$ and, for the price wars and the cartel's breakdown, simply the uncertainty over $\mu$.

I have assumed, for simplicity, that consumers have the same willingness to pay and the same outside option. Heterogeneity in the consumers' maximal willingness to pay or heterogeneous outside option create similar results: firms would sell the good to a possibly smaller mass of consumers, but the transitory phase of prices, the temporary price wars and the cartel's breakdown would remain qualitatively unchanged. Heterogeneity in the RDPs $\lambda_i$ or assuming differentiated products makes the firms compute the new optimal equilibrium price, but the qualitative results remain the same. The only slight change would be in the speed of price fall when the fringe sets the low price, but not on the price increase. Price wars and the cartel's breakdown are not affected.

With adaptive (instead of rational) expectations on the consumers' side[22] prices would still gradually rise and then be constant at $\bar{c}$. This case is discussed in more detail in Appendix 1.7.6.

The fact that firms do not know $\mu$ has no importance for the transitory phase, but it has for the breakdown of the cartel. If firms know $\mu$ perfectly, then the cartel is sustainable - or not - since the beginning of the game. In this case, a low fringe's price has no effect on the sustainability of collusion, because firms would do no Bayesian updating at all. If firms know $\mu$ imperfectly but better than consumers, no qualitative change arises: the uncertainty on the consumers' side makes price increase possible, while the uncertainty the firms' side makes price wars (and cartel's breakdown) possible. The fact that firms see the fringe's price before they set their own prices just simplifies the computation for the conditions on the sustainability of collusion, but has no role on price dynamics.

If the fringe's costs can take more than two values, price drops would be less sudden, but still faster than the rises. The reason is that when the price increases, the RDPs limit it and makesit smoother, while when it decreases this force is not present. Moreover, the "zero-one" type of collusion remains: firms can perfectly collude as long as their belief over the fringe's inefficiency is sufficiently high, but when the threshold is passed firms can only switch to competitive pricing.

The fringe can interpreted literally, like in the Vitamin C cartel, as a less efficient firms unwilling to enter the cartel, or as a firm using the "hit and run" strategy: the fringe is taken as a "black box" that embeds all the external com-

[22]For example $E_{t-1}[p_t] = min\{p^*_{t-1}, p^*_{F,t-1}\}$, with consumers expecting to pay the price they paid in the previous period.

petitive pressure that the cartel faces. In all these cases, it is reasonable to assume that consumers and incumbent firms are not sure of its efficiency, as the fringe represent new firms on the market.

This model accounts for the fact that some cartels are successful in achieving higher prices and reaching stability, while others are not. These two possibilities are represented by the firms' long run behavior when the probability of the low cost $\mu$ is, respectively, $\underline{\mu}$ and $\bar{\mu}$. Although price wars are always possible, in the first (second) case, the fringe does not (does) exert sufficient competitive pressure to make the cartel unstable in the long run. The model also accounts for price wars followed by reversals to collusive pricing. If the fringe sets the low price during a sufficient number of periods, firms revert to Nash equilibrium, while, if the fringe sets the high price during a sufficient number of periods, collusive price can become sustainable again.

This model addresses the issue of the dynamic reference point. The reference point here is endogenous. Koszegi and Heidhues (2008) elegantly showed the effects of loss aversion in a variation of the Salop model (1979) to explain the rationale for focal prices in a static environment: their reference point is the "lagged rational expectation" and they do not investigate how this is formed. Koszegi and Rabin (2006) analyze how loss aversion impacts purchase and working decisions, taking the rational expectation over outcomes as the reference point, but still in a static environment. Koszegi and Rabin (2007) and Macera (2009) do consider a dynamic game with loss aversion, but in a different framework.

This analysis is the first, at my knowledge, that deals with collusion and RDPs. It is also the first one that explains the gradual rise of prices with rational consumers holding correct beliefs over the equilibrium being played.

This model also gives some testable predictions, which can be compared to the ones in Harrington (2004, 2005) and Chen and Harrington (2006). This model predicts that the gradual rise of prices is independent from the existence of an AA, while in the models above an AA is needed and the stronger it is (in terms of probability of detection and damage multiples), the slower is the price rise. One could test data from different countries and test whether different antitrust policies have a different impact on the price path. Assuming the existence of an AA, here, does not necessarily change the results: even if an AA is sure of the existence of a cartel by looking at the price path, in order to convict its firms it needs hard evidence.

This model also predicts that external competitive pressure can lead the cartel to temporary price wars and eventually to break down. This is the case of the Vitamin C cartel, where the Chinese manufacturers remained outside the cartel and increased constantly their market share, until the cartel eventually broke down.

## 1.6 Conclusion

This model explains, through RDPs and uncertainty over the fringe's efficiency, the dynamic pattern of cartel prices, temporary price wars and cartel breakdowns. The gradual rise of prices is well known in the cartel literature and up to now the main explanations were based on the fear of Antitrust fines. My explanation is, on the contrary, based on consumers' tastes. Consumers dislike paying a price higher than the expected one and this can force firms to raise prices smoothly. When the external competitive pressure seems weak, (rational) consumers become more pessimistic towards the price they will pay. This makes them willing to pay more, as their utility loss due to RDPs is smaller. This allows firms to charge higher prices. This process eventually ends when the price reaches the high fringe's cost. When the fringe, instead, sets the low price repeatedly, (i) consumers become more optimistic and are willing to pay less and (ii) firms become more pessimistic about the value of colluding, possibly leading to price wars and the cartel breakdown.

This model yields testable predictions, e.g. that an increased competitive pressure may bring the cartel to breakdown. Empirical evidence from the Vitamin C cartel seems to confirm this. Moreover, the Antitrust enforcement should have a limited effect on the speed of the price rise.

## 1.7 APPENDIX

### 1.7.1 Proof of Lemma 1

The possibilities that make the fringe set $p_{Ft} = \bar{c}$ are: 1) the true probability is the low one ($\mu = \underline{\mu}$) and the high cost is drawn; 2) the true probability is the high one ($\mu = \bar{\mu}$) and the high cost is drawn. On the other hand, the possibilities that make the fringe set $p_{Ft} = \underline{c}$ are: 1) the true probability is the low one ($\mu = \underline{\mu}$) and the low cost is drawn; 2) the true probability is the high one ($\mu = \bar{\mu}$) and the low cost is drawn.

By observing that in $t = 1$ the fringe sets, say, $p_{F1} = \bar{c}$, consumers update their belief about $\mu$: knowing that $Pr(\mu = \underline{\mu}) = \rho$, $Pr(\mu = \bar{\mu}) = 1 - \rho$, $Pr(c_t = \underline{c}) = \mu$, $Pr(c_t = \bar{c}) = 1 - \mu$, by applying the Bayes' rule the updated belief about $\mu$ is $\rho_{1,0,1}= \frac{\rho(1-\underline{\mu})}{\rho(1-\underline{\mu})+(1-\rho)(1-\bar{\mu})}$. On the contrary, if the fringe sets $p_{F1} = \underline{c}$, their updated belief is $\rho_{1,1,0} = \frac{\rho\underline{\mu}}{\rho\underline{\mu}+(1-\rho)\bar{\mu}}$.

We get the result by iterating this procedure.

### 1.7.2 First Period Price

**Proof of Lemma 3.** We have two different cases:

1) The consumers' participation constraint (1.2c) slacks: $u_1 > 0$ when $p_1 = \bar{c}$, in which case $p_1^* = \bar{c}$ (due to the constraint of the fringe);

2) The consumers' participation constraint (1.2c) binds: $u_1 \leq 0$ when $p_1 = \bar{c}$, in which case $p_1^* = \hat{p}_1$, where $\hat{p}_1$ is the price that makes $u_1 = 0$.[23] In this case, we just have to solve for $u_1 = 0$. Using (1) we get $\hat{p}_1 = \frac{\upsilon+\lambda E_0[p_1]}{1+\lambda}$.

Given the expression (1.4) and substituting it here above, we get $\hat{p}_1 = \frac{\upsilon+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]\underline{c}}{1+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]}$. In general, the actual price $p_1^*$ is the minimum between $\bar{c}$ and $\hat{p}_1$ for three reasons. First, a price higher than $\bar{c}$ would yield zero profits due to the fringe; second, a price higher than $\frac{\upsilon+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]\underline{c}}{1+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]}$ would yield zero profits because no consumer would buy, as $u_1 < 0$; third, a price lower than $\frac{\upsilon+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]\underline{c}}{1+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]}$ does not maximize profits.

### 1.7.3 Transitory Phase

**Proof of Proposition 1.** The first inequality is Assumption 1 and the second one impedes firms to directly jump to the maximal collusive price $\bar{c}$. This condition basically states that $u_1 < 0$ when $p_1 = \bar{c}$, so $p_1^*$ will be equal to $\hat{p}_1$. We are in this case if and only if $\bar{c} < \nu < \bar{c}(1+\lambda) - \lambda E_0[\hat{p}_1]$. Substituting $E_0[\hat{p}_1]$ with its expression (4), we get

$$\bar{c} < \nu < \bar{c}(1+\lambda) - \lambda[\rho_{1,\tau,\hat{t}}\underline{\mu} + (1-\rho_{1,\tau,\hat{t}})\bar{\mu}]\hat{p}_1 + [1-\rho_{1,\tau,\hat{t}}\underline{\mu} - (1-\rho_{1,\tau,\hat{t}})\bar{\mu}]\underline{c}$$

Using $\hat{p}_1 = \frac{\upsilon+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]\underline{c}}{1+\lambda[\rho\underline{\mu}+(1-\rho)\bar{\mu}]}$), we get the result.

### 1.7.4 Duration of the Transition

Here I discuss the minimal duration of the transitory phase. Define $T$ as the period in which firms reach the maximal collusive price when the fringe always sets the high price. Formally, $T = \{t \in N | \; p_t^* = \bar{c},\, \hat{p}_{t-1}^* = \hat{p}_{t-1} < \bar{c},\, p_{Ft} = \bar{c} \; \forall t\}$

So $T$ represents the minimal[24] number of periods after which the high price reaches $\bar{c}$. Using (6), $T$ is the solution of the following system of equations:

[23] If the fringe sets the low price, firms have no incentive to deviate to a lower price: they make $\pi = 0$ by sticking to this strategy, they make $\pi = 0$ by matching the fringe's price and they make negative profits by undercutting it.

[24] Minimal because we assume here that the fringe always sets the high price.

$$\frac{\nu+\lambda[\frac{\rho(1-\underline{\mu})^T}{\rho(1-\underline{\mu})^T+(1-\rho)(1-\bar{\mu})^T}\mu+(1-\frac{\rho(1-\underline{\mu})^T}{\rho(1-\underline{\mu})^T+(1-\rho)(1-\bar{\mu})^T})\bar{\mu}]\underline{c}}{1+\lambda[\frac{\rho(1-\underline{\mu})^T}{\rho(1-\underline{\mu})^T+(1-\rho)(1-\bar{\mu})^T}\mu+(1-\frac{\rho(1-\underline{\mu})^T}{\rho(1-\underline{\mu})^T+(1-\rho)(1-\bar{\mu})^T})\bar{\mu}]}\geq\bar{c} \tag{1.9a}$$

$$\frac{\nu+\lambda[\frac{\rho(1-\underline{\mu})^{T-1}}{\rho(1-\underline{\mu})^{T-1}+(1-\rho)(1-\bar{\mu})^{T-1}}\mu+(1-\frac{\rho(1-\underline{\mu})^{T-1}}{\rho(1-\underline{\mu})^{T-1}+(1-\rho)(1-\bar{\mu})^{T-1}})\bar{\mu}]\underline{c}}{1+\lambda[\frac{\rho(1-\underline{\mu})^{T-1}}{\rho(1-\underline{\mu})^{T-1}+(1-\rho)(1-\bar{\mu})^{T-1}}\mu+(1-\frac{\rho(1-\underline{\mu})^{T-1}}{\rho(1-\underline{\mu})^{T-1}+(1-\rho)(1-\bar{\mu})^{T-1}})\bar{\mu}]}<\bar{c} \tag{1.9b}$$

An explicit solution is quite cumbersome. Nevertheless, we can state some properties of $T$. If $T$ exists, it increases with (i) the efficiency of the fringe, (ii) the importance of the RDPs and (iii) the fringe's high cost; it decreases with (i) the belief that the fringe is inefficient, (ii) the fringe's (and firms') low cost and (iii) the willingness to pay $\nu$.

A higher $\rho$ and lower $\underline{\mu}$ and $\bar{\mu}$ make $T$ smaller because, *ceteris paribus*, consumers expect to pay a higher price because of the lower probability of the low fringe's price. Their utility loss due to RDPs is smaller and this makes the convergence to $\bar{c}$ faster. A higher $\nu$ makes $T$ smaller too because it makes the utility larger than zero in a shorter number of periods.

A higher RDPs parameter $\lambda$ makes $T$ larger, because it makes the utility loss due to RDPs higher for any price higher than price expectations. A higher $\bar{c}$ also makes $T$ larger. The reason why $\underline{c}$ and $\bar{c}$ have different effects on $T$ is that $\underline{c}$ directly enters the expectation expression (1.4), while $\bar{c}$ enters the system of disequations (1.9a-1.9b) on the right side. So, firstly, $\bar{c}$ does not impact the price expectation in the transitory phase and, secondly, it requires a higher $T$ to make the left hand expression in (9a) greater than $\bar{c}$.

### 1.7.5 Conditions for Sustainability

**Proof of Lemma 6: Sufficient condition.** Assume that, if the fringe sets the low price $p_{Ft}=\underline{c}$ in any period, continuation profits are zero. This clearly reduces the continuation payoff with respect to their real value. Assume further that profits are constant for any $t$ when the fringe sets the high cost $p_{Ft}=\bar{c}$. This reduces the continuation profits too, given that we ignore the increase of profits due to the fringe setting the high price. These two assumptions will yield, therefore, a sufficient condition for sustainability. Define

$\alpha_t := Prob(p_{Ft}=\bar{c}) = 1-\hat{p}_{t,\tau,\hat{t}}\underline{\mu}-(1-\hat{p}_{t,\tau,\hat{t}})\bar{\mu}$. So the sufficient ICC becomes $n\pi \leq \pi + \frac{\delta\alpha}{(1-\delta\alpha)\pi}$. Using the definition of $\alpha$ and after some algebra, we get the result.

**Proof of Lemma 7: Necessary condition.** The ex ante expected price is not increasing over time, so expected profits do not increase either. A necessary condition for sustainability is that deviation yields lower profits than continuing to set the collusive price when continuation profits are equal to the current profit in each period. This is just necessary for two reasons. First, we know that ex ante expected prices are non-increasing over time. Here we assume them (and so the expected profits) to be constant over time, so we are possibly overestimating the continuation profits from collusion. Indeed, we are not considering the cap on prices (and profits) that the fringe exerts on firms. Second, we assume the collusive price to be sustainable forever, if firms stick to the best collusive equilibrium. Both assumptions make continuation profits after sticking to the collusive price higher than they really are, so the condition is only necessary.[25]

Assuming that perfect collusion is always sustainable and that expected future profits are constant, the ICC is

$$n\pi \leq \pi + \delta\alpha\pi + \delta^2[\alpha(\alpha\pi) + (1-\alpha)(\alpha\pi)] + \dots$$

This yields $\pi \leq \pi + \frac{\delta\alpha}{(1-\delta)\pi}$. After substituting for $\alpha_t$, we get the result.

### 1.7.6 Adaptive Expectations

All the results are robust to assuming that consumers have adaptive expectations. Relax rational expectations and Bayesian updating on the consumers' side and assume that consumers now just expect to pay the price they paid in the previous period, i.e. the smallest price among the firms' and the fringe's ones. Formally we have $E_{t-1}[p_t] = min\{p^*_{t-1}, p^*_{Ft-1}\}$.

From $u_t = 0$, we have $\hat{p}_1 = \frac{\upsilon+\lambda E_0[p_1]}{1+\lambda}$. Given adaptive expectations and that there is no price prior to $p_1$, we must assume an expected price for $p_1$. Consumers know that they will never pay a price smaller than $\underline{c}$ or larger than $\bar{c}$, so, without loss of generality, assume $E_0[p_1] = k$, where $\underline{c} \leq k \leq \bar{c}$. The objective is to show that, for any $k$, results are qualitatively the same as under rational expectations. We get $\hat{p}_1 = \frac{\upsilon+\lambda k}{1+\lambda}$. In the best collusive equilibrium, when the fringe sets the high price, consumers buy from the cartel at a price that makes their participation constraint bind. So, using (1.1), we get $\hat{p}_2 = \frac{\upsilon(1+\lambda)+\lambda\upsilon+\lambda^2 k}{(1+\lambda)^2}$. By iterating this procedure, we get $\hat{p}_{\hat{t}} = \frac{\upsilon\sum_{l=0}^{\hat{t}-1}(1+\lambda)^l\lambda^{(\hat{t}-1-l)}+\lambda^{\hat{t}}k}{(1+\lambda)^{\hat{t}}}$. This is the price firms set after $\hat{t}-1$ periods of high price by the fringe. When there has been at least one low price by the fringe, $k$ is replaced by $\underline{c}$ (consumers expect to pay the price paid in the previous period, i.e. $\underline{c}$).

[25] This condition is closer to be sufficient the farther the price is from $\bar{c}$ and the farther $\rho$ is from $\rho^*$.

It is easy to check that $\frac{d\hat{p}\_\hat{t}}{d\hat{t}} > 0$, so after every period of high fringe's price, firms' prices rise. Still there is the cap $\bar{c}$, due to the fringe, so prices can rise up to $\bar{c}$ and afterwards remain constant.

Nothing changes for the cartel breakdown either, because what matters is the firms' belief over the fringe's efficiency. Given that we only changed consumers' expectations, firms are still able to collude as long as $\hat{p}_{t,\tau,\hat{t}} \geq \rho^*$ and revert to competition (temporarily or permanently, depending on the fringe's cost draws) otherwise.

# 2 Chapter 2: Reverse Payments and Incentives to Enter

In the pharmaceutical industry, a reverse payment is a payment from the originator to a generic entrant in exchange for a delay in his entry. In some recent cases, especially when the delay is up to the patent expiry, the US and the EU Antitrust Authorities have banned these agreements *per se*. This should not be the case when the parties' investment decisions are taken into account. Allowing reverse payments increases the industry profits, which increases the generic manufacturer's incentive to invest. This, together with ex ante uncertainty over the probability that the patent is invalid (or not infringed), increases the litigation rate, which increases consumer surplus. Reverse payments delay entry to increase entry. This positive effect is larger than the static negative effect of delaying entry *given the parties investment* under many parameter sets. There is also a tension in the originator's incentives to invest from allowing reverse payments, absent from the traditional patent literature, that can increase his incentives to invest. This analysis suggests that a rule of reason is more suited than a ban *per se*.

## 2.1 Introduction

In the pharmaceutical industry, patents are the main assets of firms. Originators have the right to enforce them by litigating in order to prevent other firms from exploiting their own inventive activities. An alternative to litigation is a settlement: patent settlements, however, could cover cartel-like agreements if not scrutinised by an Antitrust Authority (henceforth AA). A firm with a tiny patent could agree to share its monopoly profits with a rival and cover it behind a settlement over the (actual or potential) patent litigation. The simplest way to share monopoly profits is a payment from the originator to the entrant in exchange for a delay in its entry, i.e. a reverse payment.

In 2003, in the Bristol-Myers[26] case, the Cardizem[27] case and the Valley Drug-Geneva Pharmaceuticals[28] case, and in 2006 in the Tamoxifen[29] the incumbent paid a potential generic competitor to avoid litigating over the patent and stay out of the market until patent expiry. The FTC found these agreements anticompetitive. In 2008, the European Commission launched an inquiry in the pharmaceutical sector with "dawn raids" of originator and generic companies, with particular attention on settlement agreements involving reverse payments. In June 2013, in the

[26] Bristol-Myers Squibb Co., 135 F.T.C. 444 (2003).

[27] Cardizem CD Antitrust Litig., 332 F.3d 896 (6th Cir. 2003).

[28] Valley Drug Co. v Geneva Pharmaceuticals, Inc, 344 F.3d 1294 (11th Cir. 2003).

[29] Tamoxifen Citrate Antitrust Litig., 466 F.3d 187, 200 (2d Cir. 2006).

Lundbeck[30] case, the European Commission fined Lundbeck and other companies for delaying generic entry through the use of reverse payments. In 2006, however, in the Tamoxifen and Valley Drug-Geneva Pharmaceuticals cases, respectively the Sixth and the Eleventh Circuit reversed the initial judgement of the FTC and found the agreements not to be illegal, as they did not extend beyond the original patent terms. Originators, here, were given the *full* right of exclusion from their patents. The main objection to this kind of judgement is that patents should be considered only *probabilistic* property rights (Shapiro 2003, Lemley and Shapiro 2005). A patent may be later found invalid or the entrant's product may be found not to infringe it. In other words, the patent-holder does not have the right to exclude a party from using its own patent, but only to *try* to exclude it. This makes entry possible, if parties litigate. Shapiro (2003) suggests to allow for settlements that leave consumers, in expectations, at least as well of as under litigation. He shows, *inter alia*, that reverse payments are a clear sign of an anticompetitive settlement. Lemley and Shapiro (2005) state that reverse payments should be presumed anticompetitive, as they delay entry relative to continued litigation and settlements not involving reverse payments. Willig and Bigelow (2004), following Shapiro's approach, deal with the reasons why a settlement with reverse payment can be beneficial to consumers. Differences in (i) the information about the future states of the market, (ii) the expectation of success in the litigation and (iii) the impact that entry of another firm has on the incumbent and the entrant explain why reverse payments can be procompetitive. However, the results crucially hing on some *caveats*.[31] Gratz (2012) compares *per se* legality, illegality and rule of reason. She finds that *per se* legality induces maximal collusion, *per se* illegality entirely prevents it and the rule of reason induces limited collusion when antitrust enforcement is subject to error. This limited collusion can be welfare enhancing, as it increases the expected settlement profits, thus fostering generic entry. This result, however, crucially depends on Antitrust Authorities making errors.

The aim of this essay is to contribute to the literature on reverse payments. In particular, I consider the impact that the Antitrust policy on reverse payments has on the parties' incentives to invest and, finally, on consumer surplus. It is, therefore, a dynamic analysis of the long-term effects of allowing reverse payments. This essay shows that reverse payments should not be considered anticompetitive *per se*. Reverse payments allow the parties to make the originator's monopoly period longer and compensate the entrant through the reverse payment. Reverse

[30] http://europa.eu/rapid/press-release_IP-13-563_en.htm

[31] In case (i), for reverse payments to improve CS, very high costs of litigation are needed: in the example at footnote 6 of their paper, the incumbent's litigation cost is as high as monopoly profits. In case (iii), CS is higher only if the deadweight losses due to monopoly and duopoly are close to zero.

payments allow the parties to share the mnopoly profits in an easy way. This delays the generic manufacturer's entry and increases industry profits. This always increases the entrant's profits and, at the same time, creates a tension in the originator's incentives to invest. The entrant's incentives to invest always increase, which increases generic entry and, when the patent strength is not common knowledge, increases the litigation rate and consumer surplus. The originator's ones can increase or decrease. The originator's incentives to invest can become lower because the higher industry profits - due to reverse payments - can make the entrant enter, which reduces the originator's profits; but they can become higher too, because the higher industry profits make the entrant keener on settling on more favorable terms for the originator, in order to avoid litigation. On the other hand, allowing reverse payments typically delays the generic manufacturer's entry, which reduces consumer surplus for a *given* investment level. In other words, reverse payments delay generic entry to induce more entry. So, the impact of allowing reverse payments on consumer surplus is not trivial and deserves a careful analysis. For several parameter sets, the pro-investment effect offsets the negative static effect of delaying entry. This suggests that a rule of reason is more suited than a ban *per se*. The optimal policy is derived.

From a more general perspective, the present paper contributes to the literature of litigation and settlement. This issue has been studied by, among others, Salant and Rest (1982), P'Ng (1983), Bebchuk (1984), Salant (1984), Reinganum and Wilde (1986), Schweitzer (1989) and Daughety and Reinganum (1994). Almost all these models assume that the bargaining process occurs sequentially, where one part makes a take-it-or-leave-it offer and the other one accepts or rejects it. If the responder accepts, the terms of the offer are enforced; if he rejects, parties litigate. Except for Schweitzer (1989) and Daughety and Reinganum (1994), incomplete information is one-sided. Some models assume that the party making the offer is the informed one (P'Ng, Salant and Rest, Salant), in which case, due to the transmission of private information through the offer, equilibria are typically very numerous (a well known feature of signalling games), while Bebchuk assumes the opposite, in which case the equilibrium is unique. Other models of bargaining assume that the identity of the proposer is determined by a coin flip, like Rubinstein and Wolinsky (1985, 1990), Gale (1986a, 1986b, 1987) and Binmore and Herrero (1988a, 1988b), or that both parties make simultaneous announcements (Wolinsky 1990).

For the sake of simplicity, we stick to a *one-sided* incomplete information game, where the *uninformed* party (the entrant) makes a take-it-or-leave-it offer. This will let us avoid the multiplicity of equilibria of signalling games. Our results, as will become clear, hold also for more general bargaining rules..

## 2.2 The model

There are three players: an Antitrust Authority (AA), an originator and a generic manufacturer (the entrant). Normalize patent length to 1 and current date to 0.[32] In the first stage, the AA decides (i) whether reverse payments are banned *per se* or not and, if they are allowed and used, (ii) a latest entry date $\hat{D} \leq 1$ upon which generic entry must occur.[33] In the second stage, the originator decides whether to invest a sum $I_O$ to enter the market. In the third stage, if the originator has invested, the generic manufacturer can enter the market if he invests a sum $I_E$. In the fourth stage, the originator learns the true probability $\theta \in \{\underline{\theta}, \bar{\theta}\}$, where $0 \leq \underline{\theta} < \bar{\theta} < 1$, of winning the litigation. The probability of $\theta = \underline{\theta}$ is $\lambda$.[34] This signal can represent the information that arrives from the national patent office or from experts asked to evaluate the patent strength prior to the possible litigation. I assume that the originator has better information about the patent strength because he is the party that filed the patent application. Therefore, he is the party knowing better the possible problems related to his patent. In the fifth stage, if both the originator and the generic have invested, they can either litigate or settle. The bargaining process is sequential: the entrant makes a take-it-or-leave-it offer and the originator accepts or rejects it.[35] The offer consists of an entry date $D \leq \hat{D}$ and a payment $R$ from the originator to the entrant, if reverse payments are allowed and used, or just an entry date $D$ (possiby greater than $\hat{D}$) otherwise. If the originator accepts the offer, $D$ and $R$ are enforced. If he rejects it, parties litigate. $D$ represents the fraction of the patent period in which the entrant commits not to enter.

The timing is then the following:

1. **Policy choice.** The Antitrust Authority implements a policy $\rho \in \{N, (R, \hat{D})\}$.[36]

[32] Date 0 is the date when the entrant is ready to enter, which is the same as when the parties decide whether to litigate or to settle. This will be shown in the Appendix. The intuition is that the entrant always prefers to enter as soon as he can.

[33] I do not consider the imposition of a latest entry date if reverse payments are banned or not used. Doing so would complicate the analysis without adding relevant insights and, moreover, no AA has ever shown concerns about the entry date for settlements that do not involve reverse payments.

[34] Through the paper the patent's strength refers both to validity and non-infringement.

[35] The fact that the entrant makes the take-it-or-leave-it offer (or, better, a take-it-or-leave-it request) is not necessary for the result. Any form of bargaining that leaves the entrant with *some* additional surplus from the settlement, with respect to his outside option (the litigation payoff), gives our result. In other words, the only bargaining solution that is not compatible with the results is the originator making the take-it-of-leave-it offer.

[36] $N$ means that reverse payments are banned and $(R, \hat{D})$ that they are allowed with latest entry date $\hat{D}$.

2. **Originator's investment.** The originator decides whether to invest $I_O$ to enter the market or not.

3. **Entrant's investment.** The entrant decides whether to invest $I_E$ to enter the market or not.

4. **Originator's signal.** The originator receives the private signal $\theta \in \{\underline{\theta}, \bar{\theta}\}$.[37]

5. **Entrant's offer.** If the entrant has invested, the entrant makes a settlement offer.

6. **Originator's response.** The originator accepts or rejects it. Rejection implies litigation.

In case of litigation, the originator and the entrant bear, respectively, a cost of litigation $C_O$ and $C_E$.[38] Define $H$ the originator's profits if he is the sole supplier on the market for the entire patent period, $L$ the originator's profits if entry occurs immediately and $E$ the entrant's profits if it enters immediately. Hence, $L+E$ are the joint profits of the originator and entrant if entry occurs immediately. Assume that $H > L+E$, i.e. monopoly profits are larger than the industry duopoly profits.

The investment decisions of the originator and the entrant depend on their own expected profits upon entry. The originator first computes whether the entrant will enter; if he will, both players form expectations over the outcome in the litigation-settlement subgame: they weight the probability that settlement will take place times the settlement profits plus the probability that litigation will occur times the litigation profits.

When reverse payments are allowed, it can be easily shown that industry profits are higher than when they are banned (intuitively, because they are used only if they delay entry, so that the pledgeable profits become higher - see Lemma 7). Therefore (i) the originator can have more or less incentives to invest (more if the higher profits make the entrant offer more favorable terms; less if they make enter an entrant that would have otherwise stayed out), (ii) the entrant always has more incentives to invest, and (iii) $CS$ is lower for a *given* investment level. We will show that there exist parameter sets where CS increases under $\rho = R$, even for $\hat{D} = 1$, thanks to the pro-investment effects (i) and (ii).

In the following subsection we compute the litigation and settlement profits.

[37] The originator is assumed to learn the true probability of winning the trial. Results are robust to variations to this assumption (e.g. the originator only receiving a noisy signal over the true probability or the entrant getting a signal over the patent strength). The only necessary feature is some asymmetric information between the originator and the entrant over the patent strength.

[38] They can be seen as the incremental legal costs of litigation (i.e. those in excess of any legal costs associated with settlement).

### 2.2.1 Litigation-Settlement stage

The model will be solved backwards. Denote $\hat{\theta}$ the realization for which the originator is indifferent between litigating and settling and $E_\theta[\theta|\theta > \hat{\theta}]$ the Bayesian updating of $\theta$ given that the originator refuses the proposal based on $\hat{\theta}$.

If the parties *litigate*, they expect to obtain:
Originator: $\theta H + (1-\theta)L - C_O$
Entrant: $(1 - E_\theta[\theta])E - C_E$ before the settlement offer;
$(1 - E_\theta[\theta|\theta > \hat{\theta}])E - C_E$ after the settlement offer if refused.

By litigating, the originator knows he has a probability $\theta$ of winning the case, in which case he gets $H$; with probability $1-\theta$ he loses and gets only $L$. Whether he wins or loses, litigation costs are $C_O$. The entrant, instead, knows *ex ante* that he has a probability $(1 - E_\theta[\theta])$ of winning, in which case he gets $E$, otherwise he earns nothing. If the originator refuses the settlement offer, this probability becomes $(1 - E_\theta[\theta|\theta > \hat{\theta}])$. His litigation costs are $C_E$.[39]

If the parties *settle*, they obtain:
Originator: $DH + (1-D)L - R$
Entrant: $(1-D)E + R$

By settling, the originator earns $DH$ in the period before the agreed entry date and $(1-D)L$ in the period until patent expiry - in which he competes with the entrant. He also pays $R$ to him. The entrant earns $(1-D)E$ if he enters at date $D$ and receives the payment $R$.

Solving the model backwards, in stage 6 the originator settles if and only if this is more profitable than litigating, i.e. if $DH+(1-D)L-R > \theta_O H+(1-\theta_O)L-C_O$, which yields:

$$R \leq R^*(\theta) = (D-\theta)(H-L) + C_O \tag{2.1}$$

The maximal reverse payment that the originator is willing to pay is an increasing function of the difference $(D-\theta)(H-L)$: a higher $(D-\theta)$ means a longer expected duration of the monopoly period, while a higher $(H-L)$ makes it more profitable to keep the entrant out. A settlement also allows the originator to avoid paying the litigation costs $C_O$, so higher litigation costs makes him willing to pay a higher reverse payment.

[39] The fact that each party bears his own litigation costs is the so called American rule. The results of this paper are robust to changes in the allocation of litigation costs (like the British rule, for example, which makes the loser party pay the litigation costs of both parties).

When reverse payments are banned or not used, then $R = 0$ and firms only bargain over $D$. For the originator, after setting $R = 0$ and rearranging terms, settlement is preferred to litigation if:

$$D \geq D_O^*(\theta) = \theta - \frac{C_O}{(H-L)} \tag{2.2}$$

The minimal entry date that the originator is willing to accept depends positively on $\theta$: a higher patent strength makes the originator less willing to accept early entry, so he accepts the settlement offer only for a sufficiently late entry date. A settlement, on the other hand, makes the originator save litigation costs $C_O$. Their relative magnitude with respect to the profits $(H - L)$ the originator loses has a negative impact on the minimal entry date, making the originator keener on settling.

Note, first, that the entrant *cannot* write a menu of contracts and make the originator truthfully reveal his type (see Appendix 2.5.5). The intuition is that the originator's settlement profits do not depend on his type. Therefore the contract for the high type $\bar{\theta}$ must yield the same profits as the one for the low type $\underline{\theta}$ (in order to keep the incentive compatibility constraints fulfilled). The entrant can decide to make the participation constraint of the high type binding, in which cases both originator's types accept the settlement, or the one of the low type, in which case the high type will litigate. In other words, given that $\theta$ has only 2 possible realizations, $\underline{\theta}$ and $\bar{\theta}$, the entrant (provided that he invests) has only two potential optimal strategies: one that leaves the originator indifferent between litigating and settling when he draws $\theta = \underline{\theta}$ and one that leaves him indifferent when he draws $\theta = \bar{\theta}$. Call the realization $\hat{\theta} \in \{\underline{\theta}, \bar{\theta}\}$ that leaves the originator indifferent between litigating and settling under the entrant's offer the realization "targeted" by the entrant. Denote $\hat{\theta}^{\rho}$ the equilibrium targeted realization under the policy $\rho$. Upon entry, the entrant proposes an entry date $D(\hat{\theta})$ and, possibly, a reverse payment $R(\hat{\theta})$ such that the originator is indifferent between accepting the settlement or litigating when the realized $\theta$ is $\hat{\theta}$.[40] When reverse payments are allowed, the only problem is the choice of the reverse payment. Indeed, we have the following Lemma.

**Lemma 7** *Under $\rho = (R, \hat{D})$, if the entrant has invested and asks for a reverse payment, he proposes the latest possible entry date: $D^{R,\hat{D}} = \hat{D}$.*

**Proof.** See Appendix 2.5.1. The intuition is that a lower entry date implies a lower reverse payment (see eq. (2.1)): a marginally lower $D$ makes the entrant

[40] We assume that the originator, when indifferent between settling and litigating, decides to settle.

gain $E$ through $D$ and lose $(H - L)$ through $R$. Being the loss in the originator's profits $(H - L)$ higher than the entrant's profits $E$, the optimal $D$ is the latest entry date $D = \hat{D}$.

Note, moreover, the duality between the latest entry date $\hat{D}$ and an imposed maximal reverse payment $\hat{R}$.

**Lemma 8** *The choice of the latest entry date $\hat{D}$ is a perfect substitute for the choice of a maximal reverse payment $\hat{R}$.*

**Proof.** See equation (2.1). There is a biunivocal correspondence between $R^*(\theta)$ and $D^*_O(\theta)$, so it is sufficient to set a cap on only $R$ or $D$.

The intuition is that the entrant finds it optimal to keep the originator indifferent between settling and litigating for a particular realization of the patent strength $\hat{\theta}^\rho$. When the AA sets a latest entry date $\hat{D}$, the entrant chooses $D = \hat{D}$ and leaves the originator with only his expected litigation payoff through the reverse payment $R$. If the AA sets a cap on $\hat{R}$, the entrant asks for $\hat{R}$ if its "dual" $\hat{D}$ is smaller than 1 and then leave the originator with his expected litigation payoff through $\hat{D}$ (if the cap $\hat{R}$ is so high that its dual $\hat{D}$ is larger than 1, the cap will not be binding). We can therefore restrict our attention to a cap on $D$. This result is valid also for the two next essays.

From Lemma 7, we know that if the parties settle with a reverse payment, we can substitute $D$ with $\hat{D}$. Denote $\pi^\rho_{i,S}$ the settlement profits of party $i$, where $i = E$ is the entrant and $i = O$ is the originator, under policy $\rho$. Under $\rho = (R, \hat{D})$, if the parties use a reverse payment, the entrant obtains:

$$\pi^{R,\hat{D}}_{E,S} = (1 - \hat{D})E + R(\hat{\theta}^{R,\hat{D}}) \tag{2.3}$$

and the originator obtains

$$\pi^{R,\hat{D}}_{O,S} = \hat{D}H + (1 - \hat{D})L - R(\hat{\theta}^{R,\hat{D}}) \tag{2.3b}$$

If parties do not use a reverse payment, by settling the entrant obtains

$$\pi^{N}_{E,S} = [1 - D(\hat{\theta}^{N})]E \tag{2.4}$$

and the originator obtains

$$\pi^{N}_{O,S} = D(\hat{\theta}^{N})H + [1 - D^N(\hat{\theta}^{N})]L \tag{2.4b}$$

### 2.2.2 Profits and consumer surplus

Under $\rho = (R, \hat{D})$, when the entrant uses $R$, in equilibrium the originator will accept the request $D = \hat{D}$, $R = R(\hat{\theta}^{R,\hat{D}}) = (\hat{D} - \hat{\theta}^{R,\hat{D}})(H - L) + C_O$ if and only if the realized patent strength is not larger than the targeted one: $\theta \leq \hat{\theta}^{R,\hat{D}}$ (see equation (2.1)). Under $\rho = N$ or when the entrant does not use $R$, the originator will accept the request $D(\hat{\theta}^N) = \hat{\theta} - \frac{C_O}{(H-L)}$ if and only if $\theta \leq \hat{\theta}^N$ (see equation (2.2)).

Denote (i) $\pi_O^{R,\hat{D}}(\hat{\theta}^{R,\hat{D}})$ and $\pi_E^{R,\hat{D}}(\hat{\theta}^{R,\hat{D}})$ the ex ante expected profits that, respectively, the originator and the entrant get when the entrant proposes a settlement with $D = \hat{D}$ and $R = R(\hat{\theta}^{R,\hat{D}})$, and (ii) $\pi_E^N(\hat{\theta}^N)$ and $\pi_E^N(\hat{\theta}^N)$ their ex ante expected profits when the entrant proposes a settlement with $D(\hat{\theta}^N)$. Given that the originator is kept at his outside option level, when the entrant targets $\underline{\theta}$ the originator gets $\pi_O^{R,\hat{D}}(\underline{\theta}) = \underline{\theta}H + (1 - \underline{\theta})L - C_O$ and, when the entrant targets $\bar{\theta}$, he gets $\pi_O^{R,\hat{D}}(\bar{\theta}) = \bar{\theta}H + (1 - \bar{\theta})L - C_O$.

Under $\rho = (R, \hat{D})$, when the entrant targets $\underline{\theta}$, we have

$$\pi_E^{R,\hat{D}}(\underline{\theta}) = \lambda[(1 - \hat{D})E + (\hat{D} - \underline{\theta})(H - L) + C_O] + (1 - \lambda)[(1 - \bar{\theta})E - C_E] \quad (2.5)$$

When the entrant targets $\underline{\theta}$, the parties settle if and only if the realized $\theta$ is $\underline{\theta}$. This occurs with probability $\lambda$. With probability $(1 - \lambda)$, the realized $\theta$ is $\bar{\theta}$, which makes the originator prefer to litigate. Therefore, in case of litigation the entrant too knows that the patent strength is $\theta = \bar{\theta}$, so both the entrant and the originator compute their litigation payoffs accordingly.

When the entrant targets $\bar{\theta}$, which can occur only if $\hat{D} \geq \bar{\theta} - \frac{C_O}{H-L}$[41], we have

$$\pi_E^{R,\hat{D}}(\bar{\theta}) = (1 - \hat{D})E + (\hat{D} - \bar{\theta})(H - L) + C_O \quad (2.6)$$

In this case, litigation never occurs in equilibrium. When the entrant targets the high realization, he is paying for an insurance: he is leaving some information rent to the originator (in case the realized patent strength is weak) in exchange for the certainty not to litigate.

Under $\rho = N$, when the entrant targets $\underline{\theta}$, we have

$$\pi_E^N(\underline{\theta}) = \lambda(1 - \underline{\theta} + \frac{C_O}{H - L})E + (1 - \lambda)[(1 - \bar{\theta})E - C_E] \quad (2.7)$$

The probability $\lambda$ and the use of $\bar{\theta}$ in the litigation payoffs have the same explanation as above.

[41] If this was not the case, the entrant could get higher profits by just offering to enter at $D = \bar{\theta} - \frac{C_O}{H-L}$ without a reverse payment. Assume, for simplicity, that $\hat{D} \geq \bar{\theta} - \frac{C_O}{H-L}$. Assuming only $\hat{D} \geq \underline{\theta} - \frac{C_O}{H-L}$ makes computations more complex (the entrant can use a reverse payment when he targets $\underline{\theta}$ but not when he targets $\bar{\theta}$) and does not add any relevant insight.

When the entrant targets $\bar{\theta}$, we have

$$\pi_E^N(\bar{\theta}) = (1 - \bar{\theta} + \frac{C_O}{H-L})E \tag{2.8}$$

Before computing the subgame perfect equilibria, we establish the following lemmas.

**Lemma 9** *The policy $\rho = (R, \hat{D} \leq D(\underline{\theta}) = \underline{\theta} - \frac{C_O}{H-L})$ is equivalent to $\rho = N$.*

**Proof.** Allowing reverse payments with such an early entry date is equivalent to banning them, because the originator would accept to settle with entry date $\hat{D}$ only if the entrant made a payment to him, i.e. $R < 0$ (see (2.1)).

This means that the parties would be reducing industry profits with respect to a settlement without a reverse payment, as entry would occur earlier than without it. Therefore, the entrant does not use $R$ and makes an offer only based on $D$. Under $\rho = (R, \hat{D} > \bar{\theta} - \frac{C_O}{H-L})$, the entrant asks for a reverse payment when they are allowed, as this allows parties to increase industry profits and, therefore, the entrant to reap a higher profit. Settling under $\rho = (R, \hat{D} > \bar{\theta} - \frac{C_O}{H-L})$ is therefore more appealing than under $\rho = N$, both when he targets the low and the high realization. We restrict our attention to $\rho = (R, \hat{D} \geq \bar{\theta} - \frac{C_O}{H-L})$ and $\rho = N$.[42]

**Lemma 10** *The entrant targets $\hat{\theta} = \underline{\theta}$ under $\rho = (R, \hat{D})$ if and only if*

$$\lambda > \lambda^{R,\hat{D}}(\hat{D}) = \frac{(\hat{D} - \bar{\theta})(H-L) - (\hat{D} - \bar{\theta})E + C_O + C_E}{(\hat{D} - \underline{\theta})(H-L) - (\hat{D} - \bar{\theta})E + C_O + C_E}$$

*and, under $\rho = N$, if and only if*

$$\lambda > \lambda^N = \frac{\frac{C_O}{H-L}E + C_E}{(\bar{\theta} - \underline{\theta})E + \frac{C_O}{H-L}E + C_E}$$

**Proof.** Compare (2.5) with (2.6) and (2.7) with (2.8). $\lambda^{R,\hat{D}}$ and $\lambda^N$ are the minimal probabilities of drawing $\underline{\theta}$ such that the entrant prefers to target $\underline{\theta}$ under $\rho = (R, \hat{D})$ and $\rho = N$.

Importantly, note that the AA can affect the targeted realization of the entrant, as $\lambda^{R,\hat{D}}(\hat{D})$ is a function strictly increasing in $\hat{D}$.[43]

**Corollary 2** *A necessary and sufficient condition for $\lambda^{R,\hat{D}} > \lambda^N$ is $\hat{D} > \bar{\theta} + \frac{C_E}{E}$.*

[42] See footnote 41 for the case $\underline{\theta} - \frac{C_O}{H-L} < \hat{D} < \bar{\theta} - \frac{C_O}{H-L}$.

[43] $\frac{d\lambda^{R,\hat{D}}(\hat{D})}{d\hat{D}} = \frac{(\bar{\theta}-\underline{\theta})(H-L)(H-L-E)}{[(\hat{D}-\underline{\theta})(H-L)+(\bar{\theta}-\hat{D})E+C_O+C_E]^2} > 0.$

**Proof.** Just compare $\lambda^R$ and $\lambda^N$.

This means that, for a sufficiently late entry date, if the entrant targets $\underline{\theta}$ under $\rho = (R, \hat{D})$ he will also target $\underline{\theta}$ under $\rho = N$, but not necessarily vice versa. This result will be useful when the entrant must be incentivized to offer a more favorable settlement to the originator in order to make him invest.

If a settlement takes place, consumer surplus (CS) is simply the monopoly CS until the generic's entry date $D$ and the duopoly CS from that moment until patent expiry $(1-D)$. Denoting $\underline{S}$ the monopoly CS and $\bar{S} > \underline{S}$ the duopoly CS, we have $CS_S = D\underline{S} + (1-D)\bar{S}$. If litigation occurs, I follow Shapiro (2003) by assuming that CS is equal to the probability that the originator wins the case times the monopoly CS, plus the probability that the entrant wins times the duopoly CS. Therefore we get $CS_L = \theta\underline{S} + (1-\theta)\bar{S}$.

*Disregarding the investment decisions*, it is easy to see that CS is higher under $\rho = N$. A "laissez faire" policy $\rho = (R, 1)$ makes firms choose $D = 1$, while banning them makes the entrant propose $D(\hat{\theta}) = \hat{\theta}^N - \frac{C_O}{(H-L)}$, that is smaller than 1 for both realizations $\theta \in \{\underline{\theta}, \bar{\theta}\}$. Being $\bar{S} > \underline{S}$, it is clear that CS is higher when reverse payments are banned, *provided that both firms invest*. When they are banned, the originator must allow the entrant to enter prior to patent expiry to make him willing to accept the settlement. This supports the FTC and the Commission's opinion that reverse payments should be banned *per se*. However, when considering the impact of both parties' incentives to invest, banning reverse payments can *reduce* CS. A ban on reverse payments, indeed, can reduce both the originator and the entrant's expected profits, making both parties less willing to invest. The originator's profits can be reduced because the smaller industry profits can make the entrant more aggressive in his settlement proposal, while the entrant's profits are always reduced because of the smaller industry profits. When the originator does not invest the CS falls to zero and when the entrant does not invest litigation never occurs, which also reduces CS. For several parameter sets, as will be shown, these two effects dominate the negative effect of late entry.

### 2.2.3 Subgame perfect equilibria

In this subsection, I compute the subgame perfect equilibria of the game. The possible outcomes of the game depend on (i) the originator's investment, (ii) the entrant's investment and settlement offer.[44]

1) If $I_O > H$, the originator does not invest and the game ends. Consumer

[44] The originator's investment decision is a "yes/no" variable, while the entrant's decisions are: (i) invest and target $\underline{\theta}$, (ii) invest and target $\bar{\theta}$, (iii) do not invest; in (i) and (ii) the entrant chooses whether to use a reverse payment or not, if they are allowed.

surplus is $CS(O_{out}) = 0$. This is trivial: monopoly profits are smaller than the investment cost, so the originator prefers to stay out of the market.

2) If $\pi_O(\bar{\theta}) \leq I_O \leq H$, then the originator invests if and only if the entrant does not enter, i.e. $I_E > \pi_E^{\rho}(\hat{\theta}^{\rho})$. The originator's investment cost is so high that he can make non-negative profits if and only if the entrant does not enter. In this case consumer surplus is $CS(E_{out}) = \underline{S}$, the monopoly outcome. If $I_E \leq \pi_E^{\rho}(\hat{\theta}^{\rho})$, the entrant would invest, which deters the originator from investing in the first place. In this case consumers surplus is $CS(O_{out}) = 0$.

3) If $\pi_O(\underline{\theta}) < I_O \leq \pi_O(\bar{\theta})$, then the originator invests if the entrant does not enter *or* if he enters and targets $\bar{\theta}$. His investment cost allows him to enter only if the entrant does not make an aggressive settlement proposal. The entrant enters if and only if $I_E \leq \pi_E^{\rho}(\hat{\theta}^{\rho})$. If $I_E > \pi_E^{\rho}(\hat{\theta}^{\rho})$, the entrant stays out and consumer surplus is $CS(E_{out}) = \underline{S}$. If the entrant enters, under $\rho = N$ he targets $\underline{\theta}$ if $\lambda > \lambda^N$ and $\bar{\theta}$ if $0 < \lambda \leq \lambda^N$;[45] under $\rho = (R,\, \hat{D})$, he targets $\underline{\theta}$ if $\lambda > \lambda^{R,\hat{D}}$ and $\bar{\theta}$ if $0 < \lambda \leq \lambda^{R,\hat{D}}$.[46] Note that, under $\rho = (R,\, \hat{D})$, the targeted realization depends on $\hat{D}$, as $\hat{D}$ has a positive impact on $\lambda^{R,\hat{D}}$. Being this impact is positive, a higher latest entry date $\hat{D}$ can make the entrant target $\bar{\theta}$, while he would otherwise target $\underline{\theta}$. The idea is that allowing for a longer monopoly period makes the entrant less willing to risk litigation. He can, indeed, get a share of these higher profits through the reverse payment. This can make him offer a better settlement to the originator (i.e. target the high realization), giving the originator the incentives to invest in the first place. For any policy, if the entrant would enter and target $\underline{\theta}$, consumer surplus is $CS(O_{out}) = 0$, as the originator would not invest. Under $\rho = N$, if the entrant enters and targets $\bar{\theta}$, then

$$CS^N(\bar{\theta}) = (\bar{\theta} - \frac{C_O}{H-L})\underline{S} + (1 - \bar{\theta} + \frac{C_O}{H-L})\bar{S}$$

Under $\rho = (R,\, \hat{D})$, if a reverse payment is used and the entrant targets $\bar{\theta}$, then $CS^{R,\hat{D}}(\bar{\theta}) = \hat{D}\underline{S} + (1 - \hat{D})\bar{S} < CS^N$.[47] Note that the policy choice, and in particular the choice of $\hat{D}$, has an impact not only on CS, but also on the entrant's profits. This impacts the entrant's entry decision *and* the target decision, which in turn can impact the originator's investment decision.

4) If $0 \leq I_O \leq \pi_O(\underline{\theta})$, then the originator invests for any choice of the entrant and, therefore, for any policy. The entrant invests, as usual, if and only if $I_E \leq \pi_E^{\rho}(\hat{\theta}^{\rho})$. If $I_E > \pi_E^{\rho}(\hat{\theta}^{\rho})$, he does not enter and consumer surplus is $CS = \underline{S}$. If he enters, under $\rho = N$ he targets $\underline{\theta}$ if $\lambda > \lambda^N$ and $\bar{\theta}$ if $0 < \lambda \leq \lambda^N$; under $\rho = (R,$

[45] i.e. he offers $D(\underline{\theta}) = \underline{\theta} - \frac{C_O}{H-L}$ if $\lambda \geq \lambda^N$ and $D(\bar{\theta}) = \bar{\theta} - \frac{C_O}{H-L}$ otherwise.

[46] i.e. he offers $D = \hat{D}$ and $R = (\hat{D} - \underline{\theta})(H-L) + C_O$ if $\lambda \geq \lambda^{R,\hat{D}}$ and $R = (\hat{D} - \bar{\theta})(H-L) + C_O$ otherwise.

[47] This inequality is true because the entrant will use a reverse payment if and only if $\hat{D} > \bar{\theta} - \frac{C_O}{H-L}$ (this argument is very similar to Lemma 9).

$\hat{D}$), he targets $\underline{\theta}$ if $\lambda > \lambda^{R,\hat{D}}$ and $\bar{\theta}$ if $0 < \lambda \leq \lambda^{R,\hat{D}}$, as above. Under $\rho = N$, when he targets $\underline{\theta}$, consumer surplus is

$$CS^N(\underline{\theta}) = \lambda[(\underline{\theta} - \frac{C_O}{(H-L)})\underline{S} + (1 - \underline{\theta} + \frac{C_O}{(H-L)})\bar{S}] + (1-\lambda)[\bar{\theta}\underline{S} + (1-\bar{\theta})\underline{S}]$$

and, when he targets $\bar{\theta}$, consumer surplus is

$$CS^N(\bar{\theta}) = (\bar{\theta} - \frac{C_O}{(H-L)})\underline{S} + (1 - \bar{\theta} + \frac{C_O}{(H-L)})\bar{S}$$

Consumer surplus for the low realization $CS^N(\underline{\theta})$ has this expression because the originator draws $\underline{\theta}$ with probability $\lambda$, in which case he settles, and $\bar{\theta}$ with probability $(1-\lambda)$, in which case he litigates. In case of litigation, consumer surplus is computed with the true probability that the originator wins, i.e. $\bar{\theta}$.[48] Under $\rho = (R, \hat{D})$, if the entrant targets $\underline{\theta}$, consumer surplus is

$$CS^{R,\hat{D}}(\underline{\theta}) = \lambda[\hat{D}\underline{S} + (1-\hat{D})\bar{S}] + (1-\lambda)[\bar{\theta}\underline{S} + (1-\bar{\theta})\underline{S}]$$

and, when he targets $\bar{\theta}$, consumer surplus is $CS^{R,\hat{D}}(\bar{\theta}) = \hat{D}\underline{S} + (1-\hat{D})\bar{S}$.

From the analysis of these subgame perfect equilibria, we get to the main result.

**Proposition 4** *There exist parameter sets where banning reverse payments reduces consumer surplus.*

**Proof and Explanation.** Banning reverse payments reduces CS when (i) it impedes generic entry that would otherwise take place, provided that the originator invests, or (ii) it deters the originator's investment, because the lower industry profits make the entrant more aggressive in the settlement offer, which reduces the originator's profits. I show these two cases in detail. Recall that the originator's profits depend only on the realization $\theta$ targeted by the entrant: we have $\pi_O(\underline{\theta}) \equiv \pi_O^N(\underline{\theta}) = \pi_O^{R,\hat{D}}(\underline{\theta}) < \pi_O(\bar{\theta}) \equiv \pi_O^N(\bar{\theta}) = \pi_O^{R,\hat{D}}(\bar{\theta})$.[49]

*Case 1: more generic entry.* Consider now $\lambda > \lambda^R > \lambda^N$. This means that the probability that the patent is weak is so high that the entrant targets the low realization under any policy. We know that a necessary and sufficient condition for $\lambda^R > \lambda^N$ is $\hat{D} > \bar{\theta} + \frac{C_E}{E}$ (Corollary 2). Under $\rho = (R, \hat{D})$, being $\hat{D} >$

[48] No result would be qualitatively different if consumer surplus was computed with the expected patent strength $\lambda\underline{\theta} + (1-\lambda)\bar{\theta}$.

[49] When the entrant targets the realization $\hat{\theta}$, the originator is kept at the same profit level as under litigation when he draws $\hat{\theta}$. This feature is due to the fact that the entrant makes a take-it-or-leave-it offer. A more general bargaining rules without this feature does not change the qualitative results.

$\bar{\theta} + \frac{C_E}{E} > \bar{\theta} - \frac{C_O}{H-L}$, the entrant will ask for a reverse payment. Consider now the profit equations (2.5) and (2.7). Being the probability of settlement, the probability of litigation and the litigation profits the same under both policies, we can just compare the entrant's settlement profits. The entrant's settlement profits under $\rho = (R, \hat{D})$ are higher than under $\rho = N$ when the entrant asks for $R$: indeed the condition $\pi_E^{R,\hat{D}}(\underline{\theta}) > \pi_E^N(\underline{\theta})$ boils down to $\hat{D} > \underline{\theta} - \frac{C_O}{(H-L)}$, a necessary condition for parties to use a reverse payment, which is fulfilled as $\hat{D} > \bar{\theta} + \frac{C_E}{E} > \bar{\theta} - \frac{C_O}{H-L} > \underline{\theta} - \frac{C_O}{(H-L)}$.

Consider now the costs of investment $I_E$ and $I_O$. If $I_O > \pi_O^{R,\hat{D}}(\underline{\theta})$, the originator never invests and the game ends. If $0 < I_O \leq \pi_O^{R,\hat{D}}(\underline{\theta})$, we have three cases:

If (I) $0 \leq I_E \leq \pi_E^N(\underline{\theta})$, then the entrant invests under both policies. This makes CS higher under $\rho = N$, because $D(\underline{\theta}) = \underline{\theta} - \frac{C_O}{H-L} < \bar{\theta} + \frac{C_E}{E} < \hat{D}$, i.e. entry occurs earlier under $\rho = N$ when a settlement takes place.

If **(II)** $\pi_E^N(\underline{\theta}) < I_E \leq \pi_E^{R,\hat{D}}(\underline{\theta})$, the entrant invests only under $\rho = (R, \hat{D})$. CS is therefore higher under $\rho = (R, \hat{D})$, because it makes the entrant enter the market and, when the originator draws $\theta = \bar{\theta}$, litigation occurs, which increases CS.[50]

If (III) $\pi_E^{R,\hat{D}}(\underline{\theta}) < I_E$, the entrant does not invest under either policy. CS is then the same under both policies.

The existence of case (II) completes the proof.[51] The intuition is that if the originator's investment cost is not too high and the entrant's cost is intermediate, the originator invests but the entrant will not if reverse payments are banned (or if the latest entry date $\hat{D}$ is too small). The entrant will invest if and only if settlement profits are sufficiently high, which occurs when reverse payments are allowed (with a sufficiently high $\hat{D}$). This increases the size of the shareable profits, making the entrant more willing to invest. Some of these investments will end up in litigations, which increases CS.[52]

*Case 2*: *more originator's investment*. Consider now $\lambda^{R,\hat{D}} > \lambda > \lambda^N$. Here the probability of the low realization of the patent strength is such that the en-

[50] When the possible realizations for $\theta$ are $n > 2$ or $\theta$ is continuous, the fact that the entrant targets the lowest realization is *not* necessary for this case to exist. The only necessary condition is that the entrant does not target the highest one. The idea is that there must be *some* realization that triggers litigation.

[51] We can show that this case holds, with small modifications, also for $\lambda > \lambda^N > \lambda^R$. In this case, of course, $\hat{D} < \bar{\theta} + \frac{C_E}{E}$, so we just need that $\underline{\theta} - \frac{C_O}{H-L} < \hat{D} < \bar{\theta} + \frac{C_E}{E}$. The entrant still uses $R$ under $\rho = (R, \hat{D})$, as $\hat{D} > \underline{\theta} - \frac{C_O}{H-L}$, which makes his profits higher than under $\rho = N$. This creates a similar wedge between $\pi_E^N(\underline{\theta})$ and $\pi_E^{R,\hat{D}}(\underline{\theta})$, such that, if $I_E$ lies between them, CS is higher under $\rho = (R, \hat{D})$.

[52] The fact that some investments end up in litigating depends on the assumption $\lambda \geq \lambda^R$, i.e. that, under $\rho = (R, D)$, the entrant targets $\underline{\theta}$. In a more general framework with $N$ possible realizations or a continuous $\theta$, however, the only necessary condition is that the entrant does not target the highest realization of $\theta$.

trant targets $\underline{\theta}$ under $\rho = N$ and $\bar{\theta}$ under $\rho = (R, \hat{D})$. Therefore, in this case, the originator obtains higher profits when reverse payments are allowed (and $\hat{D}$ is sufficiently high). Like in the previous case, for $\lambda^{R,\hat{D}} > \lambda^N$ to be true we need $\hat{D} > \bar{\theta} + \frac{C_E}{E}$. When $\pi_O(\underline{\theta}) < I_O < \pi_O(\bar{\theta})$, allowing reverse payments makes the originator invest, which increases CS. The intuition is that allowing reverse payments increases industry profits, which makes the entrant keener on settling on more favourable terms for the originator, in order to increase the probabilities of settling. This makes the originator's profits higher, which in turn triggers the originator's investment that, otherwise, would not have taken place. This, of course, raises the consumer surplus, as it creates a market that would not exist otherwise. This is another reason why reverse payments can increase CS.

These two cases show that a ban *per se* is, therefore, suboptimal.[53] In the next subsection I rank the possible outcomes of the game depending on their CS and, in section 2.3, we provide one numerical example for each case.

### 2.2.4 Ranking of CS and optimal policies

Using the results of the previous subsection, we can rank these outcomes depending on their CS in the following way.

1) The best possible outcome is that both parties invest, the entrant targets $\underline{\theta}$ and the policy is $\rho = N$. In this case entry occurs as soon as possible if parties settle and litigation is possible. In this case, consumer surplus is

$$CS^N(\underline{\theta}) = \lambda[(\underline{\theta} - \frac{C_O}{H-L})\underline{S} + (1 - \underline{\theta} + \frac{C_O}{H-L})\bar{S}] + (1-\lambda)[\bar{\theta}\underline{S} + (1-\bar{\theta})\bar{S}]$$

2) If, under $\rho = N$, the entrant enters and targets $\bar{\theta}$, consumer surplus is

$$CS^N(\bar{\theta}) = (\bar{\theta} - \frac{C_O}{H-L})\underline{S} + (1 - \bar{\theta} + \frac{C_O}{H-L})\bar{S}$$ [54]

3) If, under $\rho = N$, the entrant does not enter because his profits are smaller than his investment cost, then consumer surplus is the one under monopoly: $CS^N(E_{out}) = \underline{S}$. If the AA allows reverse payments with an entry date $\hat{D}$ that at least equalizes the entrant's investment cost and profits, then consumer surplus is

$$CS^{R,\hat{D}}(\underline{\theta}) = \lambda[\hat{D}\underline{S} + (1-\hat{D})\bar{S}] + (1-\lambda)[\bar{\theta}\underline{S} + (1-\bar{\theta})\underline{S}]$$

when the entrant still targets $\underline{\theta}$, and $CS^{R,\hat{D}}(\bar{\theta}) = \hat{D}\underline{S} + (1-\hat{D})\bar{S}$ if the entry date $\hat{D}$ is such that the entrant now targets $\bar{\theta}$.[55] Note that consumer surplus is

[53] The two effects of (i) more generic entry and (ii) more originator's investment can coexist in a more general model with $N$ realizations for $\theta$ or a continuous $\theta$.

[54] $CS^N(\bar{\theta})$ is smaller than $CS^N(\underline{\theta})$ if and only if $\lambda \geq \frac{\frac{C_O}{H-L}}{(\bar{\theta}-\underline{\theta})+\frac{C_O}{H-L}}$. We assume that this is the case.

[55] We know that a higher $\hat{D}$ makes $\lambda^{R,\hat{D}}$ larger, which can make $\lambda^{R,\hat{D}}$ larger than $\lambda$.

decreasing in the entrant's investment cost, as a higher $I_E$ requires a later entry date to keep the entrant's profits $\pi_E^{R,\hat{D}}(\underline{\theta})$ non-negative. [56]

4) If, under $\rho = N$, the entrant enters and targets $\underline{\theta}$ and the originator's investment cost $I_O$ is between the profits he would get with the low and the high realization, the originator does not enter. Consumer surplus would, therefore, be $CS^N(O_{out}) = 0$. If the AA allows reverse payments with an entry date such that the entrant targets $\bar{\theta}$, then the originator invests. The entrant targets $\bar{\theta}$ if and only if the probability $\lambda$ that the patent is weak is smaller than $\lambda^{R,\hat{D}}$. Under such a policy, consumer surplus is $CS^{R,\hat{D}}(\bar{\theta}) = \hat{D}\underline{S} + (1-\hat{D})\bar{S}$.

5) If the originator's investment cost $I_O$ is higher than the profits $\pi_O(\bar{\theta})$ he makes when the entrant targets $\bar{\theta}$, the originator never invests if the entrant eventually enters the market. If he eventually enters, therefore, the originator does not invest in the first place and consumer surplus is $CS^N(O_{out}) = 0$. If the AA implements a policy such that the entrant's entry cost $I_E$ is larger than the profits he makes by entering, then the originator invests and the entrant stays out. Therefore $CS^N(E_{out}) = \underline{S}$. In this case, therefore, the AA has an incentive to reduce the entrant's profits, to make the originator invest - a monopoly is better than nothing.

6) If the originator's investment cost $I_O$ is higher than the monopoly profits $H$, the originator never invests. Therefore, $CS^N(O_{out}) = 0$.

The objective of the AA is to maximise CS. The full characterisation of the optimal policy is long and cumbersome (see the Appendix). Here I just focus on the cases where allowing reverse payments is optimal.

**Proposition 5** *Allowing reverse payments is optimal when:*

*(1) $\pi_O(\underline{\theta}) < I_O \leq \pi_O(\bar{\theta})$, (1.a) $0 < \lambda \leq \lambda^N$ and (1.a.i) $\pi_E^N(\bar{\theta}) < I_E \leq \pi_E^{R,1}(\bar{\theta})$: the optimal entry date is $\hat{D} = \min\{D \text{ s.t. } I_E = \pi_E^{R,\hat{D}}(\bar{\theta}) \text{ \& } \lambda = \lambda^{R,1}\}$; (1.b) $\lambda^N < \lambda \leq \lambda^{R,1}$: the optimal entry date is $\hat{D} = \max\{D \text{ s.t. } I_E = \pi_E^{R,\hat{D}}(\bar{\theta}),\ D \text{ s.t. } \lambda = \lambda^{R,D}\}$. This policy makes the originator invest.*

*(2) $0 \leq I_O \leq \pi_O(\underline{\theta})$ and (2.a) $\lambda^{R,(\bar{\theta}-\frac{C_O}{\lambda(H-L)})} < \lambda < \lambda^N$, then the optimal entry date is $\hat{D} = \min\{\underline{\theta} - \frac{C_O}{H-L}, D \text{ s.t. } I_E = \pi_E^{R,\hat{D}}(\underline{\theta}) \text{ \& } \lambda \geq \lambda^{R,\hat{D}}\})$; (2.b) $\lambda^N < \lambda \leq \lambda^{R,1}$ and $I_E > \pi_E^N(\underline{\theta})$, then the optimal entry date is $\hat{D} = \min\{D \text{ s.t. } I_E = \pi_E^{R,\hat{D}}(\underline{\theta}) \& \lambda \geq \lambda^{R,D},\ D \text{ s.t. } I_E = \pi_E^{R,\hat{D}}(\bar{\theta})\}$; if (2.c) $\lambda > \lambda^{R,1}$ and $I_E > \pi_E^N(\underline{\theta})$, then $\hat{D} = \min\{D \text{ s.t. } I_E = \pi_E^{R,\hat{D}}(\underline{\theta}))$. This policy makes the entrant enter.*

**Proof and Explanation.** In case (1), the originator's investment cost is such that he invests if and only if the entrant has not invested or, having invested, targets the high realization $\bar{\theta}$. Therefore, the objective of the AA is to make the

[56] The ranking $CS^N(\bar{\theta}) > CS^{R,\hat{D}}(\underline{\theta})$ is true as long as $\hat{D} > \bar{\theta} - \frac{C_O}{\lambda(H-L)}$, otherwise it is reversed.

entrant enter and target $\bar{\theta}$. In subcase (1.a), the probability of the low realization is so small that the entrant targets $\bar{\theta}$ upon entry, which occurs if his investment cost is sufficiently small. When $I_E < \pi_E^N(\bar{\theta})$, the optimal policy is simply $\rho^* = N$, as the entrant's investment cost is so small that he would enter anyway. But if his investment cost is intermediate (1.a.i), then the AA must set a policy that allows the entrant to recoup it: therefore the optimal policy is $\rho^* = (R,\ \hat{D} = \min\{D$ s.t. $I_E \leq \pi_E^{R,\hat{D}}(\bar{\theta})$ & $\lambda = \lambda^{R,1})$. This policy means that reverse payments are allowed and the latest entry date is such that the entrant recoups his investment *and* keeps on targetting the high realization $\bar{\theta}$.

In case (2), the originator's investment cost is so small that he invest for any realization the entrant eventually targets. Therefore, when the probability of the low realization is sufficiently small $(0 < \lambda < \lambda^{R,(\bar{\theta}-\frac{C_O}{\lambda(H-L)})})$[57], an optimal policy is $\rho^* = N$: indeed, the entrant would target the high realization under both policies, so it is better to make him target it when reverse payments are banned. But when the probability of the low realization is slightly higher (2.a), it is better to allow reverse payments combined with an early entry date. The entry date will be equal to $\underline{\theta} - \frac{C_O}{H-L}$ if the investment cost is sufficiently low, and to the $\hat{D}$ that makes the entrant just recoup it if it is higher, provided that it is not higher than the $\hat{D}$ that makes $\lambda = \lambda^{R,D}$ (otherwise the entrant would target the high realization, making $CS$ fall). For a higher probability of low realization (2.b), if the investment cost is sufficiently low the optimal policy is $\rho^* = N$, as it makes the entrant target the low realization, and $\rho^* = (R,\ \hat{D} = \min\{D$ s.t. $I_E \leq \pi_E^{R,\hat{D}}(\underline{\theta})$ & $\lambda \geq \lambda^{R,D}$, $D$ s.t. $I_E \leq \pi_E^{R,\hat{D}}(\bar{\theta})\&\lambda < \lambda^{R,D}\})$ otherwise. This policy makes the entrant recoup his investment cost with the smallest entry date and keeps him willing to target the low realization or, if it is impossible to make the entrant recoup his investment and target the low realization, just sets the smallest possible entry date. For the case (2.c), where the probability of the low realization is so high that the entrant always targets it, an optimal policy is $\rho^* = N$, provided that the entrant's cost is sufficiently low. For a higher investment cost, the optimal policy is $\rho^* = (R,\ \hat{D} = D$ s.t. $I_E = \pi_E^{R,\hat{D}}(\underline{\theta}))$, so that the entrant can recoup it and compete with the originator in the period $(1 - \hat{D})$.

Note that the patent strength has no clear impact on the optimal policy (see Appendix 2.5.4). The reason is that there are two opposite forces. On one hand, allowing reverse payments is good when a patent is weak, because the additional entry ending up in litigation has a higher probability of increasing consumer sur-

[57] When $\hat{D} < (\bar{\theta} - \frac{C_O}{\lambda(H-L)})$, one can check that $CS^N(\bar{\theta}) < CS^R(\underline{\theta})$. The intuition is that, for such a low latest entry date, it is better to allow the entrant to use a reverse payment, that will make him target the low realization, rather than banning reverse payments which will make the entrant target the high realization. With such a policy, CS is higher thanks to the possibility of litigation.

plus: in other words, the benefit of reverse payments is high. On the other hand, allowing them when the patent is strong induces just a small delay compared to the expected entry date, because it would occur lately anyhow - the cost is small. These two forces make the impact of patent strength on the policy implications ambiguous.

## 2.3 Numerical examples

In this section I provide a numerical example for the two cases where reverse payments increase CS.

**Case 1: more generic entry.** Assume $I_O = 25$, $E = 10$, $(H - L) = 20$, $C_E = C_O = 2$, $\underline{\theta} = \frac{1}{3}$, $\bar{\theta} = \frac{2}{3}$, $\lambda = \frac{2}{3}$. These values imply that $\lambda^R = \frac{10\hat{D}-2.66}{10\hat{D}+4}$ and $\lambda^N = 0.47$. For any $\frac{13}{15} < \hat{D} \leq 1$ we have $\lambda > \lambda^R > \lambda^N$.

Under $\rho = (R, \hat{D})$, the originator will (i) accept $D = \hat{D}$, $R(\bar{\theta}) = (\hat{D} - \bar{\theta})(H - L) + C_O$ for any realization of $\theta$, and (ii) accept $D = \hat{D}$, $R(\underline{\theta}) = (\hat{D} - \underline{\theta})(H - L) + C_O$ if and only if $\theta = \underline{\theta}$.

Therefore we have
(i) $\pi_E^R(\bar{\theta}) = (\hat{D} - \bar{\theta})(H - L) + C_O$
(ii) $\pi_E^R(\underline{\theta}) = \lambda[(\hat{D} - \underline{\theta})(H - L) + C_O] + (1 - \lambda)[(1 - \bar{\theta})E - C_E]$

Under $\rho = N$, the originator will (a) always accept $D^N(\bar{\theta}) = \bar{\theta} - \frac{C_O}{(H-L)}$ and (b) accept $D^N(\underline{\theta}) = \underline{\theta} - \frac{C_O}{(H-L)}$ if and only if $\theta = \underline{\theta}$.

So, we have
(a) $\pi_E^N(\bar{\theta}) = (1 - \bar{\theta} + \frac{C_O}{(H-L)})E$
(b) $\pi_E^N(\underline{\theta}) = \lambda(1 - \underline{\theta} + \frac{C_O}{(H-L)})E + (1 - \lambda)[(1 - \bar{\theta})E - C_E]$

Substituting for the parameters assumed above, under $\rho = (R, \hat{D})$ we get $\pi_E^R(\bar{\theta}) = 20\hat{D} - 11.33$ and $\pi_E^R(\underline{\theta}) = 13.33\hat{D} - 2.67$, therefore the entrant will target the low realization[58], as $\pi_E^R(\underline{\theta}) > \pi_E^R(\bar{\theta})$ for any $\hat{D} \leq 1$, while under $\rho = N$ we get $\pi_E^N(\bar{\theta}) = 4.33$, $\pi_E^N(\underline{\theta}) = 5.56$, so the entrant will target the low realization too.[59]

The profits of the originator are $\pi_O(\bar{\theta}) = 31.33$ and $\pi_O(\underline{\theta}) = 26.89$. Given that the entrant targets $\underline{\theta}$ under both policies, the relevant comparison is between $I_O$ and $\pi_O(\underline{\theta})$. We have $I_O < \pi_O(\underline{\theta})$, therefore the originator invests under both policies.

So we have three cases, depending on $I_E$:

[58] i.e. he asks for $D = \hat{D}$ and $R = R(\underline{\theta}) = (D - \underline{\theta})(H - L) + C_O$.
[59] i.e. he will require $D = D(\underline{\theta}) = \underline{\theta} - \frac{C_O}{(H-L)}$.

(I) $0 < I_E < 5.56$ : the entrant invests both under $\rho = (R, \hat{D})$ and under $\rho = N$; the CS is highest under $\rho = N$, as when a settlement takes place entry occurs earlier ($D(\underline{\theta}) = 0.23 < \hat{D} = 1$);

**(II)** $5.56 < I_E < 10.67$: the entrant invests only under $\rho = (R, \hat{D})$ : the CS is highest under $\rho = (R, \hat{D})$ with $\hat{D} = \frac{3I_E+8}{40}$, as this is the earliest entry compatible with the generic's investment and litigation occurs with positive probability (the probability $(1-\lambda) = \frac{1}{3}$ of drawing $\bar{\theta}$);

(III) $I_E > 10.67$: the entrant does not invest in either case.

Note that even a "laissez-faire" policy $\rho = (R, 1)$ yields a consumer surplus strictly higher than $\rho = N$, because of the positive probability of litigation.

**Case 2: more originator's investment.** Assume the same parameters as above, but with $\pi_O(\underline{\theta}) < I_O \leq \pi_O(\bar{\theta})$ and $\lambda = \frac{1}{2}$.

Under $\rho = (R, \hat{D})$ we have $\pi_E^R(\bar{\theta}) = 20\hat{D} - 11.33$, $\pi_E^R(\underline{\theta}) = 10\hat{D} - 1.67$. The entrant, therefore, targets $\underline{\theta}$ if and only if $\hat{D} \leq \frac{29}{30}$. Under $\rho = N$ we get $\pi_E^N(\bar{\theta}) = 4.33$, $\pi_E^N(\underline{\theta}) = 4.5$, so the entrant will always target $\underline{\theta}$, which consists in requiring $D = D(\underline{\theta}) = \underline{\theta} - \frac{C_O}{(H-L)}$. The profits of the originator are $\pi_O(\bar{\theta}) = 31.33$ and $\pi_O(\underline{\theta}) = 28$. Therefore, being $\pi_O(\underline{\theta}) < I_O < \pi_O(\bar{\theta})$, allowing reverse payments with $\hat{D} \geq \frac{29}{30}$ makes the originator invest, which increases CS. The optimal policy is exactly $\rho = (R, \hat{D} = \frac{29}{30})$, so the originator invests and the monopoly period is minimized. Also in this case, of course, a policy $\rho = (R, \hat{D} = 1)$ yields a strictly higher consumer surplus than $\rho = N$, because it makes the originator invest.

## 2.4 Discussion and conclusions

When the investment decisions of the originator and the generic manufacturer are taken into account and there is uncertainty over the outcome of the litigation, banning reverse payments can *reduce* consumer surplus. The reason is two-fold. First, banning reverse payments reduces the industry profits, which reduces the entrant's expected profits. This reduces the entrant's incentives to invest and implies less entry. This, together with ex ante uncertainty over the probability that the patent is invalid (or not infringed), reduces the litigation rate and consumer surplus. This result is robust to any bargaining rule between the originator and the entrant[60] (except when the originator makes a take-it-or-leave-it offer) and to other types of asymmetric information - there is no need that the originator

[60] This result just requires the entrant to get some, however small, additional surplus from the settlement, compared to litigation. Any bargaining rule that makes the entrant enjoy some surplus makes the entrant keener on entering. A bargaining rule that makes the originator enjoy some settlement surplus, moreover, would also increase the originator's incentives to invest when reverse payments are allowed.

receives a private signal, this result holds also when it is the entrant who gets a private signal. The only necessary feature is some asymmetric information, otherwise litigation would never occur in equilibrium and the further entry due to reverse payments would not increase consumer surplus. This positive effect is larger than the static negative effect of delaying entry *given the parties investment* under many parameter sets. Second, banning reverse payments creates a tension in the originator's incentives to invest. It makes industry profits smaller, which can make (i) the entrant not willing to enter, which increases the originator's incentives to invest, but also, provided that the generic invests, it can make (ii) the entrant less keen on settling on favorable terms for the originator, which reduces the originator's profits and his incentives to invest. This result is robust to any bargaining rule between the originator and the entrant.[61] If the second effect is stronger than the first one, the originator may be deterred from investing, which makes CS fall to zero.

This model shows that allowing reverse payments *delays* generic entry in order to *increase* it: when reverse payments are actually used, entry is delayed, but the very possibility of using them increases it. This suggests that a rule of reason is more suited than a ban *per se*. Note that in many cases (like in the two numerical examples discussed above) even a *laissez-faire* policy that allows reverse payments and does not specify a latest entry date is superior to a ban *per se*. Finally, note that the patent strength has an ambiguous impact on the optimal policy. Two forces are present: on one hand, a reverse payment on a weak patent makes it likely that a litigation ends up in generic entry; on the other one, a reverse payment on a strong patent involves a small cost, given that the additional delay is small. Results hold for both strong and weak patents, though no clear policy implications can be derived over the patent strength (see Appendix 2.5.4).

For a practical enforcement, in order to know which agreements should be banned, the AA should have a rough estimate of $I_E$, $I_O$, $\pi_E^N(\hat{\theta})$, $\pi_E^{R,\hat{D}}(\hat{\theta})$ and $\pi_O(\hat{\theta})$. These estimates can be recovered from the expenses in R&D (for $I_O$), bioequivalence studies and marketing authorizations (for $I_E$) and from the profits in the market object of analysis and from the settlement agreements for similar products (for $\pi_O(\hat{\theta})$, $\pi_E^N(\hat{\theta})$, $\pi_E^{R,\hat{D}}(\hat{\theta})$ and $\pi_O(\hat{\theta})$). The AA can also estimate the parties' profits by analyzing the expected price decline due to competition. This can be done by analyzing the price pattern of similar products when generic entry

[61] Any other bargaining rule that gives some settlement surplus to the originator still yields the tension explained above. This rule would reduce the negative impact on the originator's incentives to invest when the entrant enters (because the originator is now able to extract some additional surplus from the settlement), but it would also reduce the willingness of the entrant to settle on favorable terms to the originator (exactly because the entrant, now, enjoys less profits). Qualitatively, therefore, the tension for the originator still exists.

has occurred. With this estimate of the competitive price, the AA can estimate the profits $\pi_E^N(\hat{\theta})$ and $\pi_O(\hat{\theta})$ and compare them to the estimated investment costs $I_E$ and $I_O$. The closer the estimated investment costs are to the estimated parties' profits, the more the AA should be lenient towards settlements with reverse payments.

## 2.5 Appendix

### 2.5.1 Reverse payment and late entry

**Proof of Lemma 7.** Compare the profits the entrant obtains from offering the latest entry date $D^{R,\hat{D}} = \hat{D}$ with its reverse payment $R(\hat{\theta}^{R,\hat{D}}) = (\hat{D} - \hat{\theta}^{R,\hat{D}})(H - L) + C_O$ (eq. (2.1)) and from offering an earlier entry date $\tilde{D} < \hat{D}$ with its reverse payment $R(\hat{\theta}^{R,\tilde{D}}) = (\tilde{D} - \hat{\theta}^{R,\tilde{D}})(H - L) + C_O$. Note that $R(\hat{\theta}^{R,\tilde{D}})$ is the optimal $R$ for the earlier entry date, as it keeps the originator indifferent between accepting and refusing the offer for the (possibly new) optimal targeted realization $\hat{\theta}^{R,\tilde{D}}$. Consider, first, the case where the entrant targets the same realization: $\hat{\theta}^{R,\tilde{D}} = \hat{\theta}^{R,\hat{D}}$. The probability of litigation is, therefore, the same, so we can just compare the entrant's settlement profits from the offer $\{\hat{D}, R(\hat{\theta}^{R,\hat{D}})\}$ with the ones from the alternative offer $\{\tilde{D} < \hat{D}, R(\hat{\theta}^{R,\tilde{D}})\}$. Note that a lower entry date implies a lower reverse payment (see eq. (2.1)): a marginally lower $D$ makes the entrant gain $E$ through $D$ and lose $(H - L)$ through $R$. Being the loss in the originator's profits $(H - L)$ higher than the entrant's profits $E$, the optimal $D$ is the latest entry date $D = \hat{D}$. Consider now the case where the entrant, as a consequence of lowering $D$ from $\hat{D}$ to $\tilde{D}$, targets a different realization $\hat{\theta}^{R,\tilde{D}} \neq \hat{\theta}^{R,\hat{D}}$. Even under this different target, $\tilde{D} < \hat{D}$ is not optimal, because the entrant could now raise $D$ to $\hat{D}$ and extract the additional originator's surplus through $R$. Again, therefore, the optimal $D$ is $\hat{D}$. It is easy to show that under any bargaining rule this Lemma holds.[62]

### 2.5.2 No Entry Delay

Here I show that the entrant will enter as soon as he can (footnote 32). Change the notation in the following way: 0 is now the date when the entrant is ready to enter and $T$ the entry date actually chosen by the generic. $\hat{D}$ and 1 remain, respectively, the latest entry date and the patent expiry date. The entrant now can also choose the moment $T$ when he will decide, with the originator, whether

[62] The larger is $\hat{D}$, the longer the monopoly period - it is in both parties' interest to make it as long as possible.

to litigate or to settle. In order to simplify the notation, just assume that the patent strength is common knowledge.[63]

Now, if the parties *litigate*, they expect to obtain:
Originator: $TH + (1-T)[\theta H + (1-\theta)L] - C_O$
Entrant: $(1-T)(1-\theta)E - C_E$

The only additional element, here, is $T$: the larger $T$, the higher the monopoly profits the originator earns before the settlement-litigation decision.

If the parties *settle*, they obtain:
Originator: $TH + (1-T)[DH + (1-D)L] - R$
Entrant: $(1-T)(1-D)E + R$

The originator settles if and only if this is more profitable than litigating, i.e. if

$TH + (1-T)[DH + (1-D)L] - R > TH + (1-T)[\theta H + (1-\theta)L] - C_O$.
Substituting $D$ with $\hat{D}$, we have:

$$R \leq R^*(T) = (1-T)(\hat{D} - \theta)(H - L) + C_O \tag{2.9}$$

The maximal reverse payment the originator is ready to pay is a decreasing function of $T$. The larger the time elapsed between the moment when the entrant is ready to enter and the moment when he actually enters, the lower the amount the originator is ready to pay. This means that waiting is counterproductive for the entrant, as it only reduces the amount he can get. When reverse payments are banned or not used, then $R = 0$ and firms only bargain over $D$. For the originator, after setting $R = 0$ and rearranging terms, settlement is preferred to litigation if:

$$D \geq D^*(T) = \theta - \frac{C_O}{(1-T)(H-L)} \tag{2.10}$$

The minimal entry date that the originator is willing to accept depends negatively on $T$. The higher the time between the moment when the entrant is ready and the moment when he discusses the settlement with the originator, the earlier he can actually enter (in case of settlement) *relatively to the remaining (lower) patent validity*. This positive effect for the entrant must be weighted with the later moment $T$ when the settlement is discussed. Substituting $D^*(T)$ in the entrant's profits, we get $\pi_E^N = (1-T)(1 - \theta + \frac{C_O}{(1-T)(H-L)})E$. Its derivative with respect to $T$ is $\frac{d\pi_E^N}{dT} = -1 + \theta < 0$, therefore the entrant prefers to enter as soon as possible.

Both when reverse payments are allowed and when they are not, the entrant will set $T = 0$ and enter as soon as possible.

[63] It will be clear that results do not depend on this.

### 2.5.3 Optimal Policy

Here I give the full characterisation of the optimal policy.

**Proposition 6** *An optimal Antitrust rule*[64] *is the following:*

*if (1) $I_O > H$, any policy is ineffective;*

*if (2) $\pi_O(\bar{\theta}) < I_O \leq H$, then $\rho^* = N$;*

*if (3) $\pi_O(\underline{\theta}) < I_O \leq \pi_O(\bar{\theta})$, then if (3.a) $0 < \lambda \leq \lambda^N$ and (3.a.i) $0 \leq I_E \leq \pi_E^N(\bar{\theta})$ then $\rho^* = N$; if (3.a.ii) $\pi_E^N(\bar{\theta}) < I_E \leq \pi_E^{R,1}(\bar{\theta})$, then $\rho^* = (R,\ \hat{D} = \min\{D$ s.t. $I_E = \pi_E^{R,\hat{D}}(\bar{\theta})$, $D$ s.t. $\lambda = \lambda^{R,1}\})$; if (3.a.iii) $I_E > \pi_E^{R,1}(\bar{\theta})$ then any policy is ineffective; if (3.b) $\lambda^N < \lambda \leq \lambda^{R,1}$, then $\rho^* = (R,\ \hat{D} = \max\{D$ s.t. $I_E = \pi_E^{R,\hat{D}}(\bar{\theta})$, $D$ s.t. $\lambda = \lambda^{R,D})$; if (3.c) $\lambda > \lambda^{R,1}$, then any policy is ineffective;*

*if (4) $0 \leq I_O \leq \pi_O(\underline{\theta})$, then if (4.a) $0 < \lambda < \lambda^{R,(\bar{\theta}-\frac{C_O}{\lambda(H-L)})}$ then $\rho^* = N$; if (4.b) $\lambda^{R,(\bar{\theta}-\frac{C_O}{\lambda(H-L)})} < \lambda < \lambda^N$, then $\rho^* = (R,\ \hat{D} = \min\{\underline{\theta} - \frac{C_O}{H-L}, D$ s.t. $I_E = \pi_E^{R,\hat{D}}(\underline{\theta}) \& \lambda \geq \lambda^{R,D}\})$; if (4.c) $\lambda^N < \lambda \leq \lambda^{R,1}$, then $\rho^* = N$ if $I_E \leq \pi_E^N(\underline{\theta})$ and $\rho^* = (R,\ \hat{D} = \min\{D$ s.t. $I_E = \pi_E^{R,\hat{D}}(\underline{\theta}) \& \lambda \geq \lambda^{R,D}$, $D$ s.t. $I_E = \pi_E^{R,\hat{D}}(\bar{\theta})\})$ if $I_E > \pi_E^N(\underline{\theta})$; if (4.d) $\lambda > \lambda^{R,1}$, then $\rho^* = N$ if $I_E \leq \pi_E^N(\underline{\theta})$ and $\rho^* = (R,$ $\hat{D} = \min\{D$ s.t. $I_E = \pi_E^{R,\hat{D}}(\underline{\theta}))$ if $I_E > \pi_E^N(\underline{\theta})$.*

**Proof and Explanation.** Case (1) is the case where the originator's investment cost is higher than monopoly profits: trivially, nothing can be done to push him to invest. In case (2), where the originator's investment cost is smaller than monopoly profits but higher than the profits he makes when the entrant is present, the AA needs to deter the generic's entry. This makes the originator reap as much profits as possible, in order to make him willing to invest. The originator's investment cost is, indeed, higher than the profits he would get if the entrant entered - even if he targeted the high realization $\bar{\theta}$. By reducing the entrant's profits as much as possible, through $\rho^* = N$, some marginal entrants will not invest, making the originator invest.[65] This case highlights the *negative* impact of reverse payments on investment, which is absent from the traditional patent literature. In case (3), the originator invests if and only if the entrant has not invested or, having invested, targets the high realization $\bar{\theta}$. Therefore, the objective of the AA is to make the entrant enter and target $\bar{\theta}$. In subcase (3.a), the entrant targets $\bar{\theta}$ upon entry, which occurs if his investment cost is sufficiently small (3.a.i): in this case the optimal policy is $\rho^* = N$. If the entrant's investment cost is intermediate

[64] I write "an" optimal Antitrust rule, instead of "the" optimal, because there can be other policies that yield the same CS. For example, in case (2), any policy other than $\rho = N$ yields the same CS if $I_E > \max\{\pi_E^{R,1}(\bar{\theta}), \pi_E^{R,1}(\underline{\theta})\}$, because the generic would not enter in any case.

[65] No policy can be better than $\rho^* = N$ in deterring the generic's entry. With any other policy $\rho = (R, \hat{D})$, the generic manufacturer can choose between using a reverse payment or not. He will choose to use it only if it makes him get higher profits, while he will not choose it otherwise. It is therefore not possible to reduce the entrant's profits through a policy $\rho = (R, \hat{D})$.

(3.a.ii), then the AA must set a policy that allows the entrant to recoup it. In this case the optimal policy is $\rho^* = (R,\ \hat{D} = \min\{\hat{D} = \min\{D \text{ s.t. } I_E = \pi_E^{R,\hat{D}}(\bar{\theta}),\ D \text{ s.t. } \lambda = \lambda^{R,1}\})$. This policy means that reverse payments are allowed and the latest entry date makes the entrant recoup his investment *and* target the high realization $\bar{\theta}$. If the entrant's investment cost is high (3.a.iii), then any policy is indifferent, because the generic manufacturer would never enter and the originator would invest anyway (getting $H > I_O$). In case (4), the originator's investment cost is so small that he would invest for any realization the entrant may target. Therefore, when the probability of the low realization is sufficiently small (4.a)[66], an optimal policy is $\rho^* = N$. The entrant would target the high realization under both policies, so it is better to make him target it when reverse payments are banned. When the probability of the low realization is slightly higher (4.b), as explained in the main text, it is better to allow reverse payments combined with an early entry date. This will be equal to $\underline{\theta} - \frac{C_O}{H-L}$ if the investment cost is sufficiently low and to the $\hat{D}$ that makes the entrant just recoup it if it is higher, provided that it is not higher than the $\hat{D}$ that makes $\lambda = \lambda^{R,D}$ (otherwise the entrant would target the high realization, making $CS$ fall). For a higher probability of low realization (4.c), an optimal policy is $\rho^* = N$. It makes the entrant target the small realization, if his investment cost is sufficiently low, and $\rho^* = (R,\ \hat{D} = \min\{D \text{ s.t. } I_E \leq \pi_E^{R,\hat{D}}(\underline{\theta}) \& \lambda \geq \lambda^{R,D},\ D \text{ s.t. } I_E \leq \pi_E^{R,\hat{D}}(\bar{\theta}) \& \lambda < \lambda^{R,D})$ otherwise. This policy makes the entrant recoup his investment cost with the smallest entry date and keeps him willing to target the low realization, or just sets the smallest possible entry date if it is impossible to make the entrant recoup his investment and target the low realization. For the case (4.d), where the probability of the low realization is so high that the entrant always targets it, an optimal policy is, like for case (4.c), $\rho^* = N$, provided that the entrant's cost is sufficiently low. For a higher investment cost, an optimal policy is $\rho^* = (R,\ \hat{D} = D \text{ s.t. } I_E = \pi_E^{R,\hat{D}}(\underline{\theta}))$, so that the entrant can recoup it and compete with the originator in the period $(1 - \hat{D})$.

### 2.5.4 Patent Strength and Optimal Policy

This paragraph shows the ambiguous impact of the patent strength $\theta$ on the optimal policy. Recall that there are two opposite forces: 1) a weak patent makes the additional entry ending up in litigation have a higher probability of increasing consumer surplus; 2) a strong patent makes reverse payments induce just a small

[66]When $\hat{D} < (\bar{\theta} - \frac{C_O}{\lambda(H-L)})$, one can check that $CS^N(\bar{\theta}) < CS^R(\underline{\theta})$. The intuition is that, for such an early latest entry date, it is better to allow the entrant to use a reverse payment (which will make him target the low realization) rather than banning reverse payments (which will make the entrant target the high realization). With such a policy, CS is higher thanks to the possibility of litigation.

delay, because the expected entry under litigation is already late. I will show how each area of the optimal policy changes after an increase in the patent strength. Assume that an increase in the patent strength consists of an increase of $\underline{\theta}$ and $\bar{\theta}$ of the same size: for simplicity, call it an "increase in $\theta$".

In Area (1) of Proposition 6 nothing can be done to induce the originator to invest, so the patent strength has no impact on this area.

In Area (2) it is optimal to ban reverse payments: the condition $\pi_O(\bar{\theta}) < I_O \leq H$ characterizing this area can be rewritten as $\bar{\theta}H + (1-\bar{\theta})L - C_O < I_O \leq H$, so it becomes smaller.

In Area (3) the optimal policy depends on the size of $\lambda$ and $I_E$: it can be rewritten as $\underline{\theta}H+(1-\underline{\theta})L-C_O < I_O \leq \bar{\theta}H+(1-\bar{\theta})L-C_O$. Note, first, that an increase in $\theta$ has no impact on its size -both sides increase by the same amount. When (3.a) $0 < \lambda \leq \lambda^N$, which can be rewritten as $0 < \lambda \leq \frac{EC_O+(H-L)C_E}{(\bar{\theta}-\underline{\theta})E(H-L)+EC_O+(H-L)C_E}$, an increase in $\theta$ has no impact either. When (3.a.i) $0 \leq I_E \leq \pi_E^N(\bar{\theta})$, that can be rewritten as $0 \leq I_E \leq (1-\bar{\theta}+\frac{C_O}{H-L})E$. An increase in $\theta$ reduces the area and in this area it is optimal to ban reverse payments. So, like in Area (2), an increase in $\theta$ reduces the parameter region where a ban on reverse payments is optimal. This, however, is not the case when (3.a.ii) $\pi_E^N(\bar{\theta}) \leq I_E \leq \pi_E^{R,1}(\bar{\theta})$, that can be rewritten $(1-\bar{\theta}+\frac{C_O}{H-L})E \leq I_E \leq (1-\bar{\theta}+\frac{C_O}{H-L})(H-L)$. Here an increase in $\theta$ makes this area smaller, like above, but here it is optimal to allow reverse payments. This means that the parameter region where allowing reverse payment is optimal is reduced. The area that, therefore, grows is the one where (3.a.iii) $I_E > \pi_E^{R,1}(\bar{\theta})$, i.e. $I_E > (1-\bar{\theta}+\frac{C_O}{H-L})(H-L)$: in this area any policy is ineffective, as it is impossible to induce the generic manufacturer to enter. When (3.b) $\lambda^N < \lambda \leq \lambda^{R,1}$, which can be rewritten

$$\frac{EC_O+(H-L)C_E}{(\bar{\theta}-\underline{\theta})E(H-L)+EC_O+(H-L)C_E} < \lambda \leq \frac{(1-\bar{\theta})(H-L)-(1-\bar{\theta})E+C_O+C_E}{(1-\underline{\theta})(H-L)-(1-\bar{\theta})E+C_O+C_E}$$

the optimal policy is to allow reverse payments and an increase in $\theta$ reduces the area. When (3.c) $\lambda^{R,1} < \lambda \leq 1$, i.e. $\frac{(1-\bar{\theta})(H-L)-(1-\bar{\theta})E+C_O+C_E}{(1-\underline{\theta})(H-L)-(1-\bar{\theta})E+C_O+C_E} < \lambda \leq 1$, no policy can make the entrant target the high type $\bar{\theta}$, which means that the originator will never enter at the first place. An increase in $\theta$ makes this area larger.

In Area (4), i.e. when the originator's investment cost is very low, when (4.a) $0 < \lambda \leq \lambda^{R,\theta-\frac{C_O}{\lambda(H-L)}}$ the optimal policy is to ban reverse payments and an increase in $\theta$ makes the area larger. When (4.b) $\lambda^{R,\theta-\frac{C_O}{\lambda(H-L)}} < \lambda \leq \lambda^N$, it is optimal to allow reverse payments and an increase in $\theta$ makes the area larger too. But when (4.c) $\lambda^N < \lambda \leq 1$ an increase in $\theta$ has an ambiguous effect: when ((4.c.i)&(4.d.i)) $0 \leq I_E \leq \pi_E^N(\underline{\theta})$, the optimal policy is to ban them and an increase in $\theta$ makes the area of the originator's investment cost larger, but the areas of $\lambda$ and the entrant's investment cost smaller; and when ((4.c.ii)&(4.d.ii)) $I_E > \pi_E^N(\underline{\theta})$, the optimal policy is to allow reverse payments and an increase in $\theta$ makes the originator and the entrant's investment costs areas larger, but the area of $\lambda$ smaller. The overall effect is therefore ambiguous and no strong policy implications on the patent strength can be derived.

The following table resumes the results. The arrows (and the equality signs) in the last column represent the impact of the patent strength on the areas of, respectively, the originator's investment cost $I_O$, the probability of the low patent strength realization $\lambda$ and the entrant's investment cost $I_E$.

| Area | $I_O$ | $\lambda$ | $I_E$ | $\rho^*$ | $\frac{d(Area)}{d\theta}$ |
|---|---|---|---|---|---|
| 1 | $[H, +\infty)$ | any | any | ind. | $=$ |
| 2 | $[\pi_O(\bar{\theta}), H]$ | any | any | N | $\downarrow$ |
| 3.a.i | $[\pi_O(\underline{\theta}), \pi_O(\bar{\theta})]$ | $\lambda \leq \lambda^N$ | $[0, \pi_E^N(\bar{\theta})]$ | N | $\downarrow$ (=,=,$\downarrow$) |
| 3.a.ii | $[\pi_O(\underline{\theta}), \pi_O(\bar{\theta})]$ | $\lambda \leq \lambda^N$ | $[\pi_E^N(\bar{\theta}), \pi_E^{R,1}(\bar{\theta})]$ | R | $\downarrow$ (=,=,$\downarrow$) |
| 3.a.iii | $[\pi_O(\underline{\theta}), \pi_O(\bar{\theta})]$ | $\lambda \leq \lambda^N$ | $[\pi_E^{R,1}(\bar{\theta}), +\infty)$ | ind. | $\uparrow$ (=,=,$\uparrow$) |
| 3.b | $[\pi_O(\underline{\theta}), \pi_O(\bar{\theta})]$ | $\lambda^N < \lambda \leq \lambda^{R,1}$ | any | R | $\downarrow$ (=,$\downarrow$) |
| 3.c | $[\pi_O(\underline{\theta}), \pi_O(\bar{\theta})]$ | $\lambda^{R,1} < \lambda$ | any | ind. | $\uparrow$ (=,$\uparrow$) |
| 4.a | $[0, \pi_O(\underline{\theta})]$ | $\lambda \leq \lambda^{R,\theta-\frac{C_O}{\lambda(H-L)}}$ | any | N | $\uparrow$ ($\uparrow$,=) |
| 4.b | $[0, \pi_O(\underline{\theta})]$ | $\lambda^{R,\theta-\frac{C_O}{\lambda(H-L)}} < \lambda \leq \lambda^N$ | any | R | $\uparrow$ ($\uparrow$,=) |
| 4.c.i | $[0, \pi_O(\underline{\theta})]$ | $\lambda^N < \lambda \leq \lambda^{R,1}$ | $[0, \pi_E^N(\underline{\theta})]$ | N | ? ($\uparrow$,$\downarrow$,$\downarrow$) |
| 4.c.ii | $[0, \pi_O(\underline{\theta})]$ | $\lambda^N < \lambda \leq \lambda^{R,1}$ | $[\pi_E^N(\underline{\theta}), +\infty)$ | R | ? ($\uparrow$,$\downarrow$,$\uparrow$) |
| 4.d.i | $[0, \pi_O(\underline{\theta})]$ | $\lambda^{R,1} < \lambda$ | $[0, \pi_E^N(\underline{\theta})]$ | N | ? ($\uparrow$,$\downarrow$,$\downarrow$) |
| 4.d.ii | $[0, \pi_O(\underline{\theta})]$ | $\lambda^{R,1} < \lambda$ | $[\pi_E^N(\underline{\theta}), +\infty)$ | R | ? ($\uparrow$,$\downarrow$,$\uparrow$) |

### 2.5.5 (No) Menu of Contracts

This Appendix shows that the entrant cannot write a menu of contracts to extract a higher rent from the originator and avoid litigation. Consider a candidate menu of contracts $\{(\underline{D}, \underline{R}), (\bar{D}, \bar{R})\}$, where $(\underline{D}, \underline{R})$ is designed for the low type and $(\bar{D}, \bar{R})$ for the high type. The constraints to fulfill are:

$$\bar{D}H + (1 - \bar{D})L - \bar{R} \geq \underline{D}H + (1 - \underline{D})L - \underline{R} \qquad (IC_{\bar{\theta}})$$

$$\underline{D}H + (1 - \underline{D})L - \underline{R} \geq \bar{D}H + (1 - \bar{D})L - \bar{R} \qquad (IC_{\underline{\theta}})$$

$$\bar{D}H + (1 - \bar{D})L - \bar{R} \geq \bar{\theta}H + (1 - \bar{\theta})L - C_O \qquad (PC_{\bar{\theta}})$$

$$\underline{D}H + (1 - \underline{D})L - \underline{R} \geq \underline{\theta}H + (1 - \underline{\theta})L - C_O \qquad (PC_{\underline{\theta}})$$

The first two inequalities are the incentive compatibility constraints to make each type prefer not to pretend to be the other type. Note that their true type does *not* enter these equations - their type only enters their litigation payoff (their outside option). So, the only way to fulfill them is to make them have the same value. We have, therefore, $\bar{D}H + (1-\bar{D})L - \bar{R} = \underline{D}H + (1-\underline{D})L - \underline{R}$, that yields

$$(\bar{D} - \underline{D})(H - L) \geq \bar{R} - \underline{R} \qquad (IC_{\bar{\theta}} = IC_{\underline{\theta}})$$

The third and the fourth inequalities are the participation constraints to make each type prefer not to litigate. Given that their LHS must the same ($IC_{\bar{\theta}} = IC_{\underline{\theta}}$), only the inequality with the larger RHS can bind. This is $PC_{\bar{\theta}}$, as $\bar{\theta}H + (1-\bar{\theta})L - C_O$ is larger than $\underline{\theta}H + (1-\underline{\theta})L - C_O$ because $(\bar{\theta} - \underline{\theta})(H - L) > 0$. Therefore, type $\bar{\theta}$ is left with no rent and type $\underline{\theta}$ enjoys an information rent. Consider now the entrant's profits. Recall that $\lambda$ is the probability that the type is $\underline{\theta}$. The entrant's problem is:

$$\max_{(\underline{D},\underline{R}),(\bar{D},\bar{R})} \pi_E = \lambda[(1-\underline{D})E+(\underline{D}-\underline{\theta})(H-L)+C_O]+(1-\lambda)[(1-\bar{D})E+(\bar{D}-\bar{\theta})(H-L)+C_O]$$

The derivatives of $\pi_E$ w.r.t. $\underline{D}$ and $\bar{D}$ are $\frac{d\pi_E}{d\underline{D}} = \frac{d\pi_E}{d\bar{D}} = H-L-E > 0$, therefore the entrant asks for the maximal possible entry dates. This makes $\underline{D} = \bar{D} = \hat{D}$. The reverse payments $\underline{R}$ and $\bar{R}$ associated with these entry dates are equal to the one that makes $PC_{\bar{\theta}}$ binding: $\underline{R} = \bar{R} = (\hat{D} - \bar{\theta})(H - L) + C_O$. Given that $\underline{D} = \bar{D}$ and $\underline{R} = \bar{R}$, the candidate menu of contracts reduces to a single contract that leaves some information rent to the $\underline{\theta}$-type. This contract is exactly the same as the one that "targets" $\bar{\theta}$ (see the main text). When the probability that the originator has type $\bar{\theta}$ is too small, it is optimal to "shutdown" this type and only offer a contract based on $\underline{\theta}$ that extracts all his rent. This is exactly the contract that "targets" $\underline{\theta}$ in the main text.

# 3 Chapter 3: Reverse Payments and Liquidity Constraints

This chapter shows another reason why reverse payments should not be banned *per se*. When the possibility that the generic manufacturer goes bankrupt is taken into account, banning reverse payments can reduce consumer surplus. Banning them reduces the generic entrant's profits, not allowing him to stay on the market in case of liquidity problems. This reduces consumer surplus both before and after patent expiry through the reduction of the number of competitors. The negative effect of banning reverse payments is larger when the patent is strong, the economy is in a downturn and tacit collusion is sustainable among few players.

## 3.1 Introduction

This essay focuses on the liquidity problems of (generic) pharmaceutical firms. From an empirical point of view, Hall (2002) shows that small and new pharmaceutical firms, as generic producers often are, experience high costs of capital, which can lead to - and worsen - liquidity problems. Pisano (2006) shows that new biotechnology firms are financially constrained and that some drug development failures can lead to bankruptcy. From a policy oriented perspective, Arve (2012), in the context of a public procurement auction, provides a rationale for policies that help financially weak players.

The present essay analyzes the impact that the treatment of reverse payments has on consumer surplus when the generic producer may have liquidity problems. Liquidity problems can be due, for example, to drug development failures or to unexpectedly high investment costs.

The framework and the timing are similar to the previous chapter, with the difference that the entrant, who still makes the take-it-or-leave-it offer, is now the *informed* party.[67] In particular, he is informed over his own financial strength. Competition in the post-expiry period is considered too (in a reduced form).

I show that reverse payments should not be considered anticompetitive *per se*. I assume that a generic manufacturer with liquidity problems needs to earn a minimal amount of profits to remain on the market. Banning reverse payments does not allow the parties to delay the generic manufacturer's entry, which reduces

[67] I assume the entrant to make the offer because we need the entrant to grasp some additional surplus from the higher industry profits due to reverse payments. Results would still hold for any bargaining rule that gives the entrant *some* additional surplus when reverse payments are allowed. The only bargaining rule incompatible with the results is the originator making the take-it-or-leave-it offer. This is exactly the same as in the previous chapter.

the industry profits. This reduces the generic manufacturer's expected profits, making the weak entrants go bankrupt. This reduces consumer surplus both before and after patent expiry through the reduction of the number of competitors. The trade-off of allowing reverse payments is (i) having more competitors on the market, at the cost of (ii) a later entry date. This trade-off is similar to the one of chapter 2. There, reverse payments increase generic entry and possibly the originator's investment too; here, they make weak players able to stay on the market and compete. In both cases, reverse payments delay entry, but the very possibility of using them increases the competition on the market.

For several parameter sets, the positive effect of allowing reverse payments offsets the negative effect of delaying entry. The positive effect of allowing reverse payments is stronger when the patent is strong, the economy is in a downturn and few players are able to tacitly collude. This result suggests that a rule of reason is more suited than a ban *per se*.

## 3.2 The model

There are three players: an Antitrust Authority (AA), a originator and a generic entrant. In the first stage, the AA decides (i) whether reverse payments are banned *per se* or not and, if they are allowed and used, (ii) a latest entry date $\hat{D} \in [0,1]$ upon which generic entry must occur. We normalize patent length to 1 and current date to 0. $D$, therefore, represents the fraction of the patent period in which the entrant commits not to enter, like in the previous chapter. In the second stage, the entrant learns his type $\tau \in \{\underline{\tau}, \bar{\tau}\}$, where $\tau = \underline{\tau}$ with probability $\mu \in [0,1]$. The type $\tau = \underline{\tau}$ means that the entrant is *weak*: if the originator and the entrant do not settle with a reverse payment above a treshold $k$, the is not able to compete and exits the market immediately.[68] If the entrant is *strong*, he can remain on the market in any case. In the third and fourth stage, the originator and the entrant can litigate or settle their dispute. The bargaining process is sequential: in the third stage the entrant makes a take-it-or-leave-it offer and, in the fourth stage, the originator accepts or rejects it.[69] The offer consists of an entry date $D$ and, possibly, a reverse payment $R$ from the originator to the entrant. If a reverse

[68] This can represent a situation where the entrant goes bankrupt or simply prefers to abandon that market.

[69] The fact that the entrant makes the take-it-or-leave-it offer (or, better, a take-it-or-leave-it request) is not necessary for the result. Any form of bargaining that can leave the entrant with *some* additional surplus from the settlement with respect to his threat point (the litigation payoff) gives our result - for the framework in the main text, any form of bargaining that makes the equilibrium reverse payment equal to or bigger than $k$ gives our result. In other words, the only bargaining solution that is not compatible with our claim is the originator making the take-it-of-leave-it offer, like in the previous chapter.

payment is used, the entry date $D$ must be smaller or equal to $\hat{D}$. If the originator accepts the offer, $D$ and $R$ are enforced. If he rejects it, parties litigate. In the fifth stage, if parties litigate *or* if the reverse payment is below $k$, the weak entrant exits the market.

The timing is then the following:

1. **Policy choice.** The Antitrust Authority implements a policy $\rho \in \{N, (R, \hat{D})\}$.
2. **Entrant's type.** The entrant learns his type.
3. **Entrant's offer.** The entrant makes a settlement offer.
4. **Originator's response.** The originator accept or rejects it.
5. **Entrant's exit decision.** The entrant decides whether to exit the market.

The main differences from this timing and the one in the previous chapter is that here, before the settlement offer, (i) the entrant draws a private signal over his financial strength (in the previous chapter, the originator draws his private signal over the patent strength) and (ii) investment decisions are not taken into account.[70] The notation is like in the previous chapter. The originator's patent has strength $\theta$. Differently from the previous chapter, patent strength is common knowledge here: parties agree over the likelihood that the originator wins in a litigation.[71] In case of litigation, the originator and the entrant bear, respectively, litigation costs $C_O$ and $C_E$. In the pre-expiry period, denote $H$ the originator's profits if he is the sole supplier on the market for the entire patent period, $L$ if entry occurs immediately and $E$ the entrant's profits if he enters immediately. Hence, $L + E$ are the joint profits of the originator and entrant if entry occurs immediately. Like in the previous chapter, we assume $H > L + E$: monopoly profits are larger than industry duopoly profits. In the post-expiry period, denote $h$ the originator's profits if the entrant has exited the market, $l$ the originator's profits if the entrant is still on the market and $e$ the entrant's profits if he is still on the market. Denote $\underline{S}$ the consumer surplus (CS) in the pre-expiry period when only the originator is active, $\bar{S}$ the pre-expiry CS when also the entrant is on the market, $\underline{s}$ the post-expiry CS when the entrant has gone bankrupt and $\bar{s}$ the post-expiry CS when the entrant is still on the market. We assume $\bar{S} \geq \underline{S}$ and

[70] If we had kept the investment decisions, taking into account the possibility of the entrant's liquidity constraints lowers the threshold for the originator's investment, but does not add any relevant insight.

[71] This simplifies the analysis, but results also hold when parties can have different beliefs over the patent strength.

$\bar{s} \geq \underline{s}$. The weight of the post-expiry period, both in the parties' profits and in CS, is $\delta$. Note that, as the post-expiry period can be much longer than the pre-expiry one, $\delta$ does *not* need to be smaller than 1.[72]

The basic tradeoff for CS is that under $\rho = (R, \hat{D})$, when the latest entry date $\hat{D}$ is sufficiently high, the higher industry profits can allow the entrant to remain on the market. This increases CS both in the pre-expiry and in the post-expiry period. On the other hand, allowing reverse payment typically makes the parties agree on a later entry date in order to make the originator's monopoly period longer. This lowers the CS in the pre-expiry period if the entrant is strong. I will show that there are several parameter sets where the positive effect of allowing them offsets the negative one, even for $\hat{D} = 1$.

In the following subsection I compute the litigation and settlement profits in the *pooling* equilibrium where both entrant's types use the same strategy at the settlement offer and response stages.[73] In subsection 3.2.4 I explain why separating equilibria are not relevant in this setting.

### 3.2.1 Litigation-Settlement stage

If the parties *litigate*, they expect to obtain:

Originator: $\mu(H + \delta h) + (1 - \mu)[\theta H + (1 - \theta)L - C_O + \delta l]$

Entrant: $[(1 - \theta)E - C_E] + \delta e$ if he is strong, 0 otherwise

By litigating, the originator knows that with a probability $\mu$ the entrant is weak, in which case refusing to settle drives him out of the market. In this case, the originator enjoys the full monopoly profits in the pre-expiry period, $H$, plus the post-expiry profits with one competitor less, $h$. If the entrant is strong, he remains on the market: the originator has a probability $\theta$ of winning the case, in which case it gets $H$, and probability $1 - \theta$ of losing it and get $L$; in both cases, he pays litigation costs $C_O$ and gets $\delta l$ from the post-expiry period. The strong entrant knows that he has a probability $(1 - \theta)$ of winning, in which case he gets $E$, otherwise he earns nothing. His costs of litigation are $C_E$. The weak entrant, on the other hand, knows that he cannot stay on the market if they litigate, so he exits. His litigation payoff is therefore 0.

If the parties *settle* with a sufficiently high ($R \geq k$) *reverse payment*, they obtain:

[72] For example, the patent could expire in two years from the settlement offer, while the drug is not expected to be replaced by better drugs in the following ten years. Future, i.e. the post expiry period, can therefore have a much higher value than the present.

[73] This is not, therefore, a pooling equilibrium in a strict sense: when the originator litigates and the entrant is weak, in stage 5 the entrant exits, while when he is strong he remains. It is pooling only in the offer and response stages, i.e. stages 3 and 4.

Originator: $DH + (1 - D)L - R + \delta l$
Entrant: $(1 - D)E + R + \delta e$ independently from his type

By settling, the originator enjoys $DH$ in the period before the agreed entry date and $(1 - D)L$ in the period after it until patent expiry. He pays $R$ to the entrant and, finally, obtains $\delta l$ from the post-expiry period. The entrant earns $(1 - D)E$, receives the payment $R$ and obtains $\delta e$ after patent expiry.

If the parties *settle* with a null or small reverse payment ($0 \leq R < k$), their expected payoff is:

Originator: $\mu(H + \delta h) + (1 - \mu)[DH + (1 - D)L + \delta l] - R$
Entrant: $(1 - D)E + R + \delta e$ if he is strong, $R$ otherwise

When reverse payments are allowed, the originator will accept the entrant's settlement offer if and only if settling is more profitable than litigating. When $R \geq k$, the originator accepts if and only if $DH + (1 - D)L - R + \delta l > \mu(H + \delta h) + (1 - \mu)[\theta H + (1 - \theta)L - C_O + \delta l]$, which yields

$$R \leq R^*(D) = (D - \theta - \mu + \mu\theta)(H - L) + (1 - \mu)C_O - \mu\delta(h - l), \qquad (3.1)$$

The maximal reverse payment acceptable for the originator $R^*(D)$ is increasing in $D$: a later entry date allows him to earn higher profits, thus making him willing to pay a higher reverse payment. It is decreasing in the patent strength $\theta$: a stronger patent makes it more appealing for the originator to litigate, reducing his willingness to pay. It is decreasing in the probability $\mu$ that the entrant is weak: the higher the probability that the entrant is weak, i.e. that he would exit the market in case of litigation, the less the originator is willing to pay a reverse payment that keeps him on the market. In order to make the problem interesting, we assume that $R^*(1) \geq k$. This means that the maximal reverse payment acceptable for the originator $R^*$ associated with entry date upon patent expiry ($D = 1$) is higher than the amount $k$ the weak entrant must get to remain on the market.

If the AA sets $\rho = N$, then $R$ is constrained to be equal to 0 and firms can only bargain over $D$. For the originator, after setting $R = 0$ and rearranging terms, settlement is preferred to litigation if

$$D \geq D^* = \theta - \frac{C_O}{(H - L)} \qquad (3.2)$$

The minimal acceptable entry date is increasing in the patent strength $\theta$ (for the same reasons as above) and decreasing in the ratio between the originator's litigation costs and the profit loss in case of entry: a settlement makes the originator save $C_O$, whose importance must be weighted over the incremental profits

$(H - L)$ due to keeping the monopoly. The probability $\mu$ that the entrant is weak plays no role, because a settlement without reverse payments does not make a weak entrant survive.

The following two Lemmas will simplify the analysis.[74]

**Lemma 11** *Under $\rho = (R, \hat{D})$, if the entrant asks for a reverse payment, he asks for the latest entry date $D = \hat{D}$.*

**Proof.** Compare the profits the entrant obtains from the offer $D = \hat{D}$, $R = R^*(\hat{D}) = (\hat{D} - \theta - \mu + \mu\theta)(H - L) + (1 - \mu)C_O - \mu\delta(h - l)$ and from an alternative offer $D = \tilde{D} < \hat{D}$, $R = R^*(\tilde{D}) = (\tilde{D} - \theta - \mu + \mu\theta)(H - L) + (1 - \mu)C_O - \mu\delta(h - l)$. Note that $R^*(\tilde{D})$ is the optimal reverse payment given the entry date $\tilde{D}$, as it keeps the originator indifferent between accepting and refusing the settlement offer. A lower $D$ implies a lower $R$ (see eq. (3.1)): a marginally lower $D$ makes the entrant gain $E$ through $D$ and lose $(H - L)$ through $R$. Being $(H - L) > E$, the optimal $D$ is $D = \hat{D}$. Therefore, if the parties settle by using a reverse payment, $D = \hat{D}$.

From Lemma 11, we know that if the parties settle with a reverse payment, we can substitute $D$ with $\hat{D}$. Denote $\pi^{\rho}_{i,S}$ the settlement profits of party $i$, where $i = E$ is the entrant (independently from his type), $i = sE$ is the strong entrant and $i = O$ is the originator, under policy $\rho$. Under $\rho = (R, \hat{D})$, if the parties use a reverse payment $R \geq k$, the entrant, independently from his type, obtains:

$$\pi^{R \geq k, \hat{D}}_{E,S} = (1 - \hat{D})E + (\hat{D} - \theta - \mu + \mu\theta)(H - L) + (1 - \mu)C_O - \mu\delta(h - l) + \delta e \quad (3.3)$$

Under $\rho = (R, \hat{D})$, when $0 < R < k$, the strong entrant's profits $\pi^{R<k,\hat{D}}_{sE,S}$ are as in eq. (3.3) above, while the weak's ones are 0.

Under $\rho = N$ (and under $\rho = (R, \hat{D})$ if firms do not use a reverse payment), if firms settle the strong entrant obtains

$$\pi^{N}_{sE,S} = (1 - D^*_O)E + \delta e \quad (3.4)$$

and the weak entrant gets 0. The originator is always kept at his outside option level, his litigation payoff.

### 3.2.2 Minimal entry dates

By setting $R^*(D)$ of eq. (3.1) equal to $k$, the minimal entry date for which the originator accepts a settlement with $R \geq k$ is equal to:

[74] Lemma 11 is the simplified version of Lemma 7 of the previous chapter (simplified because here the patent strength is common knowledge).

$$D^{R\geq k} = \theta - \frac{C_O}{H-L} + \frac{k}{H-L} + \mu(1-\theta+\frac{C_O+\delta(h-l)}{H-L}) \qquad (3.5)$$

Note that $D^{R\geq k}$ is strictly higher than $D^*_O$ (compare this expression with (3.2)) and it is an increasing function of $\mu$. The higher the probability that the entrant is weak, the later must be the entry date to convince the originator to pay $R \geq k$, as this will keep the weak entrant on the market.

If $\hat{D} \geq D^{R\geq k}$, then the entrant can offer a sufficiently late entry date $\hat{D}$ such that the originator is willing to pay $R \geq k$, allowing therefore the weak entrant to survive.

Now consider the *strong* entrant's incentives. By comparing his settlement profits with and without reverse payments, the strong entrant will propose a settlement with a reverse payment if and only if the latest entry date is at least equal to:

$$D^{R\succeq N} = \theta - \frac{C_O}{H-L} + \frac{\mu[(1-\theta)(H-L)+C_O+\delta(h-l)]}{H-L-E} \qquad (3.6)$$

Note that $D^{R\succeq N}$ is strictly higher than $D^*_O$ and increasing with $\mu$ too. If $\hat{D} \geq D^{R\succeq N}$ , then the strong entrant prefers a settlement with a reverse payment. In order to make the problem interesting, we assume that $D^{R\succeq N} \leq 1$.

Consider now the weak entrant's offer. He needs a reverse payment of at least $k$ to remain on the market[75], so his incentive to ask for a reverse payment is stronger than for the strong entrant. But the originator knows that, if $\hat{D} < D^{R\succeq N}$, the strong entrant prefers not to ask for a reverse payment, so the originator understands that the entrant is weak if he asks for $R \geq k$. The originator knows that, by litigating against a weak type, he gets the full monopoly profits $H+\delta h$, so no settlement would ever take place. In other words, if the weak entrant asks for a reverse payment when the strong type would not, he is revealing that he is weak. Therefore the originator will always litigate, driving him out of the market. The weak entrant has, then, no better option than just mimicking the strong entrant's strategy. That is the reason why we just refer to the "entrant", regardless of his type, in the following analysis.

Note that $D^{R\geq k} \geq D^{R\succeq N}$ if and only if

$$k \geq k^* = \frac{\mu E}{H-L-E}[(1-\theta)(H-L)+C_O+\delta(h-l)]$$

[75] We assume that the weak entrant needs a sufficiently high reverse payment to avoid bankruptcy - an early entry date is not sufficient. This assumption just simplifies the analysis: allowing the entrant to remain on the market also for a sufficiently early entry date does not change the main conclusions. The only feature needed for the results is that a settlement with reverse payments yields higher profits than a settlement without it.

When $k \geq k^*$, we have three possible outcomes depending on the latest entry date $\hat{D}$ chosen by the AA: (1.1) $\hat{D} \geq D^{R \geq k} \geq D^{R \succeq N}$ : the entrant asks for $R \geq k$ and the weak type survives; (1.2) $D^{R \geq k} > \hat{D} \geq D^{R \succeq N}$ : the entrant asks for $R < k$ and the weak type exits the market; (1.3) $D^{R \geq k} \geq D^{R \succeq N} > \hat{D}$ : the entrant asks for $D^*$ and the weak type exits the market.

When $k < k^*$, the three cases are: (2.1) $\hat{D} \geq D^{R \succeq N} \geq D^{R \geq k}$ : the entrant asks for $R \geq k$ and the weak type survives; (2.2) $D^{R \succeq N} > \hat{D} \geq D^{R \geq k}$: the entrant asks for $D^*$ and the weak type exits the market;[76] (2.3) $D^{R \geq k} \geq D^{R \succeq N} > \hat{D}$: the entrant asks for $D^*$ and the weak type exits the market.

This leads to the following Lemma.

**Lemma 12** *The policy $\rho = (R, \hat{D} \leq \max\{D^{R \geq k}, D^{R \succeq N})$ is, for the weak entrant, equivalent to $\rho = N$.*

In other words, if $\rho = (R, \hat{D} \leq \max\{D^{R \geq k}, D^{R \succeq N}\})$, the weak entrant exits the market, like under $\rho = N$. The intuition is that the policies $\rho = (R, \hat{D} \leq D^{R \geq k})$ do not make the weak entrant able to remain on the market even when he receives a reverse payment and the policies $\rho = (R, \hat{D} \leq D^{R \succeq N})$ do not make the entrant ask for a reverse payment. This Lemma will turn out to be useful in the analysis of the optimal policy.

### 3.2.3 Consumer surplus and optimal policies

Recall that $\underline{S}$ is the CS in the pre-expiry period when only the originator supplies the product, $\bar{S}$ the pre-expiry CS when also the entrant is on the market, $\underline{s}$ the post-expiry CS when the entrant has gone bankrupt and $\bar{s}$ when the entrant is on the market. We have $\bar{S} \geq \underline{S}$ and $\bar{s} \geq \underline{s}$.

If a settlement with $R \geq k$ takes place, total CS is $CS_S^{R \geq k, \hat{D}} = \hat{D}\underline{S} + (1 - \hat{D})\bar{S} + \delta\bar{s}$. This is the sum of the monopoly CS $\underline{S}$ until $\hat{D}$, the duopoly CS $\bar{S}$ after $\hat{D}$ until patent expiry and the competitive CS with the entrant still active, $\bar{s}$, after patent expiry. When parties settle without a reverse payment, we have

$$CS_S^N = \mu(\underline{S} + \delta\underline{s}) + (1 - \mu)[(\theta - \frac{C_O}{(H - L)})\underline{S} + (1 - \theta + \frac{C_O}{(H - L)})\bar{S} + \delta\bar{s}]$$

With a probability $\mu$ the generic is weak and exits the market, leaving the monopoly to the originator until patent expiry; after it, competition with one competitor less takes place. With a probability $(1 - \mu)$ the generic is strong and, therefore, remains on the market, so CS is equal to the monopoly CS until

[76] The weak type exits the market because, even if $\hat{D} \geq D^{R \geq k}$ (i.e. the originator would accept to pay $R \geq k$, if he was unsure whether the entrant is weak), by asking for $R^*$ the entrant reveals that he is weak, as $D^{R \succeq N} > \hat{D}$. The originator is therefore not willing to settle, as he knows that by litigating he makes the entrant exit the market, enjoying the full monopoly profits.

$D^* = \theta - \frac{C_O}{(H-L)}$, to the duopoly CS after $D^*$ before patent expiry and to the competitive CS after it. Finally, the CS from a settlement with $0 < R < k$ is

$$CS_S^{R<k,\hat{D}} = \mu(\underline{S} + \delta \underline{s}) + (1-\mu)[\hat{D}\underline{S} + (1-\hat{D})\bar{S} + \delta \bar{s}]$$

If the entrant is weak, the originator keeps the monopoly until patent expiry and, afterwards, competition witn one competitor less takes place; if the entrant is strong, the originator keeps the monopoly until the latest entry date $\hat{D}$ and, afterwards, there is a duopoly until patent expiry and competition afterwards. Moreover, by comparing $CS_S^{R<k,\hat{D}}$, $CS_S^N$ and $CS_S^{R\geq k,\hat{D}}$ we have the following Lemma.

**Lemma 13** $CS_S^{R<k,\hat{D}}$ *is smaller than both* $CS_S^N$ *and* $CS_S^{R\geq k,\hat{D}}$.

**Proof.** From eq. (3.1) we know that $R > 0$ if and only if

$$(\hat{D} - \theta - \mu + \mu\theta)(H - L) + (1-\mu)C_O - \mu\delta(h - l) > 0$$

which can be rewritten as

$$\hat{D} > D^{R>0} = \theta - \frac{C_O}{H-L} + \mu(1 - \theta + \frac{C_O + \delta(h-l)}{H-L})$$

Note that $D^{R>0}$ is strictly greater than $D^* = \theta - \frac{C_O}{(H-L)}$ : therefore, for any settlement with $0 < R < k$, entry is delayed compared to a settlement without a reverse payment. Moreover, given that $R < k$, if the entrant is weak he will exit the market anyway, so the advantage of getting him on board is lost. Therefore, a policy that makes the parties settle with $0 < R < k$ is never optimal. The basic trade-off for the AA is between (i) making the weak entrant not go bankrupt, and (ii) early entry. It is, therefore, useless to set a policy such that reverse payments are allowed with a latest entry date that makes the corresponding reverse payment smaller than $k$. The reason is that entry date is delayed compared to $\rho = N$ and the objective of keeping the weak entrant on the market is not achieved. If the liquidity the entrant needs is higher than the reverse payment the originator is willing to give, the optimal entry date will never be such that the entrant asks for a reverse payment.

*Disregarding the entrant's bankruptcy problem*, it is easy to see that CS is higher under $\rho = N$. A "laissez faire" policy $\rho = (R, 1)$ makes firms choose $D = 1$, while banning them makes the entrant propose $D^* = \theta - \frac{C_O}{(H-L)}$, that is strictly smaller than 1. Being CS decreasing in the generic manufacturer's entry date, it is clear that CS is higher when reverse payments are banned. When they are banned, the originator must allow the entrant to enter prior to patent expiry in order to make him willing to accept the settlement. This supports the Commission's opinion that reverse payments should be banned *per se*.

However, banning reverse payments can reduce CS when we consider the entrant's bankruptcy problem. A ban on reverse payments, indeed, can force the weak entrant to exit the market, which makes the originator enjoy the full monopoly profits until patent expiry and competition with one competitor less afterwards. The CS both before and after patent expiry are reduced. For several parameter sets, this effect offsets the negative effect of late entry, making CS lower than if reverse payments had been allowed.

This leads to the main result.

**Proposition 7** *There exist parameter sets where banning reverse payments reduces consumer surplus.*

**Proof.** First of all, given Lemmas 12 and 13, we can simply focus on (i) a policy $\rho = N$ and (ii) the policies $\rho = (R, \hat{D} > \max\{D^{R\geq k}, D^{R\succeq N}\})$, disregarding the policies $\rho = (R, \hat{D} \leq \max\{D^{R\geq k}, D^{R\succeq N})$. In particular, (i) we can disregard the policies $\rho = (R, \hat{D} \leq D^{R\geq k})$ because, if the entrant asks for a reverse payment[77], $R$ is not sufficient to keep the weak entrant on the market, and (ii) we can disregard the policies $\rho = (R, \hat{D} \leq D^{R\succeq N})$ because the entrant would simpy not ask for a reverse payment.

Recall that

$$CS_S^{R\geq k,\hat{D}} = \hat{D}\underline{S} + (1-\hat{D})\bar{S} + \delta\bar{s}$$

and

$$CS_S^N = \mu(\underline{S} + \delta\underline{s}) + (1-\mu)[(\theta - \frac{C_O}{(H-L)})\underline{S} + (1-\theta+\frac{C_O}{(H-L)})\bar{S} + \delta\bar{s}]$$

Banning reverse payments reduces consumer surplus when $CS_S^N < CS_S^{R\geq k,\hat{D}}$, i.e. when

$$\delta\mu(\bar{s}-\underline{s}) > (\hat{D} - \mu - (1-\mu)(\theta - \frac{C_O}{H-L}))(\bar{S}-\underline{S}) \tag{3.7}$$

This inequality yields a number of interesting policy implications.

**Proposition 8** *The higher the probability $\mu$ that the entrant is weak, the larger the parameter set where it is optimal to allow for reverse payments.*

**Proof.** $\mu$ has a positive impact on the left hand side (LHS) of the inequality and a negative impact on the right hand side (RHS). Indeed, the derivative of the LHS on $\mu$ is $\delta(\bar{s}-\underline{s}) > 0$ and the derivative of the RHS is $(-1+\theta-\frac{C_O}{H-L})(\bar{S}-\underline{S}) < 0$.

[77] This occurs when $D^{R\geq k} > \hat{D} \geq D^{R\succeq N}$.

A consequence is that, if we interpret an economic downturn as a higher probability that the entrant is weak, *during an economic downturn* it is better to be *more lenient* towards reverse payments. An economic downturn means, *inter alia*, lower profitability and a worse credit crunch - conditions that impact negatively the survival rate of small firms. It becomes, therefore, more important that the AA be more lenient in this case, as a strict ban on reverse payments can force a number of generic firms out of the market, reducing competition both before and after patent expiry.

**Proposition 9** *(i) The higher the difference in future CS* $(\bar{s} - \underline{s})$ *from having one additional competitor and (ii) the lower the difference in present CS* $(\bar{S} - \underline{S})$ *from having one additional competitor, the larger the parameter set where it is optimal to allow for reverse payments.*

These two apparently contrasting points can be linked if some form of collusion is possible as long as the number of firms is sufficiently low. The idea is that one additional competitor, further than the traditional pro-competitive effect, can also make it impossible for the firms to (tacitly) collude. If that additional competitor makes firms switch to a more competitive equilibrium, $(\bar{s} - \underline{s})$ can be substantial. This makes $(\bar{s} - \underline{s})$ higher when passing from, say, three competitors to four, rather than from two to three - if some form of tacit collusion is sustainable among up to three firms. For the same reason, the difference $(\bar{S} - \underline{S})$ can be small if the originator and the entrant can achieve some form of collusion. This underlines the importance of analyzing the way competition takes place among firms. If some form of collusion is likely to occur among few players, allowing weak generic manufacturers to stay on the market can increase CS, as it may reduce the sustainability of collusion.

**Proposition 10** *The higher the patent's strength* $\theta$*, the larger the parameter set where it is optimal to allow for reverse payments.*

This comes from the fact that, when the patent is strong, the probability that the entrant wins the litigation is small: this makes the benefit of prohibiting reverse payments smaller too, because the parties already agree on a late entry date. The stronger the patent, the later the agreed entry date if reverse payments are banned: the AA should therefore be more lenient towards reverse payments if the patent is strong, as the advantage of banning them (obliging the originator and the strong entrant to an early entry date) shrinks, while the disadvantage (having less competition after patent expiry) is unchanged.[78]

[78] There is another disadvantage, that consists in less competition before patent expiry. This disadvantage, however, shrinks with the patent strength too.

Inequality (3.7) also shows that it is more likely that reverse payments increase CS when the post-expiry period has a high value compared to the present. This means that the drug is not expected to be replaced in the next future: the longer its expected life, the higher the post-expiry CS, therefore the better it is to have more competitors after patent expiry.

These Propositions show that a ban *per se* is, therefore, not optimal. We can now derive the optimal policy.

**Proposition 11** *When the AA maximizes consumer surplus, the optimal Antitrust rule is:*

*if (i)* $k \geq k^*$ *and* $\frac{k}{H-L} + \frac{\mu\delta(h-l)}{H-L} \leq \frac{\delta\mu(\bar{s}-\underline{s})}{(\bar{S}-\underline{S})}$, *then* $\rho^* = (R, D^{R \geq k})$;

*if (ii)* $k \geq k^*$ *and* $\frac{k}{H-L} + \frac{\mu\delta(h-l)}{H-L} > \frac{\delta\mu(\bar{s}-\underline{s})}{(\bar{S}-\underline{S})}$, *then* $\rho^* = N$;

*if (iii)* $k < k^*$ *and* $\frac{\mu[(1-\theta)(H-L)+C_O+\delta(h-l)]}{H-L-E} - \frac{C_O}{H-L} \leq \mu(1-\theta) - (1-\mu)\frac{C_O}{H-L} + \frac{\delta\mu(\bar{s}-\underline{s})}{(\bar{S}-\underline{S})}$, *then* $\rho^* = (R, D^{R \succeq N})$;

*if (iv)* $k < k^*$ *and* $\frac{\mu[(1-\theta)(H-L)+C_O+\delta(h-l)]}{H-L-E} - \frac{C_O}{H-L} > \mu(1-\theta) - (1-\mu)\frac{C_O}{H-L} + \frac{\delta\mu(\bar{s}-\underline{s})}{(\bar{S}-\underline{S})}$, *then* $\rho^* = N$.

**Proof.** For cases (i) and (ii), the fact that $k \geq k^*$ means that $D^{R \geq k} \geq D^{R \succeq N}$. The condition $\frac{k}{H-L} + \frac{\mu\delta(h-l)}{H-L} \leq \frac{\delta\mu(\bar{s}-\underline{s})}{(\bar{S}-\underline{S})}$ is the necessary and sufficient condition to have both (a) $R \geq k$ and (b) $CS^{R,\hat{D}} > CS^N$. Therefore, the minimal entry date that makes the entrant (1) ask for a reverse payment ($\hat{D} \geq D^{R \succeq N}$) which is (2) at least equal to $k$ ($\hat{D} \geq D^{R \geq k}$) is $D^{R \geq k}$. This entry date maximizes CS, because the weak entrant remains on the market and the monopoly period is minimized. In case (ii), $CS^{R,\hat{D}} < CS^N$, therefore the optimal policy is simply $\rho^* = N$. For cases (iii) and (iv), $k < k^*$ means that $D^{R \geq k} < D^{R \succeq N}$. The condition $\frac{\mu[(1-\theta)(H-L)+C_O+\delta(h-l)]}{H-L-E} - \frac{C_O}{H-L} \leq \mu(1-\theta) - (1-\mu)\frac{C_O}{H-L} + \frac{\delta\mu(\bar{s}-\underline{s})}{(\bar{S}-\underline{S})}$ is the necessary and sufficient condition to have both (a) the entrant preferring to ask for a reverse payment and (b) $CS^{R,\hat{D}} > CS^N$. Therefore, when $CS^{R,\hat{D}} > CS^N$ (case (iii)), the minimal entry date that makes the entrant (1) ask for a reverse payment which is (2) at least equal to $k$ is $D^{R \succeq N}$. In case (iv), like in case (ii), the optimal policy is simply $\rho^* = N$.

**Proposition 12** *A "laissez-faire" policy* $\rho = (R, 1)$ *yields higher CS than* $\rho = N$ *if and only if* $\delta\mu(\bar{s} - \underline{s}) \leq (1 - \mu - (1 - \mu)(\theta - \frac{C_O}{H-L}))(\bar{S} - \underline{S})$.

**Proof.** We know that $\hat{D} = 1$ makes the entrant (1) ask for a reverse payment which is (2) larger than $k$.[79] So, a necessary and sufficient condition to have a

[79] We have assumed that the entry date $D^{R \succeq N}$ such that the entrant prefers to ask for a reverse payment is not greater than 1 and the reverse payment when entry occurs only at the patent expiry $R^*(1)$ is larger than $k$.

higher CS under $\rho = (R, 1)$ than under $\rho = N$ is equation (3.7) with $\hat{D} = 1$, which yields $\delta\mu(\bar{s} - \underline{s}) \leq (1 - \mu - (1 - \mu)(\theta - \frac{C_O}{H-L}))(\bar{S} - \underline{S})$.

This Proposition shows that, by considering the possible liquidity constraint of the generic producer, even a "laissez-faire" policy can yield a higher consumer surplus (and, *a fortiori*, total welfare) than a ban on reverse payments.

### 3.2.4 Separating equilibria

Up to now, we have discussed the pooling equilibrium where the strong and the weak entrant use the same strategies at the settlement offer and response stages. In a separating equilibrium, by definition, the strong entrant signals himself as strong, so the originator perfectly knows what type of entrant he is facing. As a consequence, this makes him accept a settlement if and only if the entrant is strong, because refusing to settle with a weak entrant gives him the full monopoly profits. Therefore, in a separating equilibrium allowing reverse payments never increases CS: only settlements with a strong entrant would take place and the benefit of allowing reverse payments (keeping the weak generic manufacturers on the market) would be lost. Note, however, that these equilibria always yield smaller profits for *both* the entrant's types. This is explained in the next Proposition.

**Proposition 13** *No separating equilibrium yields higher profits for the entrant than the pooling equilibrium.*

**Proof.** A strong entrant can signal himself as strong if and only if he requires a reverse payment smaller than $k$ when a higher reverse payment would be accepted too. The best signalling proposal is therefore ($D = D^{R \geq k}$, $R = k - \epsilon$), provided that it is acceptable for the originator. If $R^*(\hat{D}) < k$, the proposal ($D = D^{R \geq k}$, $R = k - \epsilon$) is not acceptable for the originator. In this case, the best acceptable proposal is ($D = \hat{D}$, $R = R^*(\hat{D}) < k$), which is *not* part of a separating equilibrium. Indeed both entrant's types use this strategy: the strong entrant has no way to signal himself as strong. Consider now the case $R^*(\hat{D}) \geq k$. In the best separating equilibrium for the entrant, offering ($D = D^{R \geq k}$, $R = k - \epsilon$) is sufficient to make the originator sure that the entrant is strong. But the originator would have also accepted ($D = \hat{D}$, $R = R^*(\hat{D})$), the pooling offer, which yields higher profits for the entrant.

In other words, the existence of the weak type creates a strong negative externality on the strong entrant. The strong entrant can signal himself as strong, but this is very costly: when the originator would have accepted to pay $R^*(\hat{D}) > k$ (the only situation where signalling is ever possible), signalling means that the entrant can only requires at most $R \leq k - \epsilon$ (otherwise also the weak entrant would have an incentive to require it). In order to keep the originator indifferent between accepting and refusing the offer, the entrant gets an earlier entry date $D^{R \geq k} < D$, but this, nevertheless, is not sufficient to recover the profits lost because of the lower reverse payment. The reason is very close to Lemma 11: an earlier entry does not compensate for the foregone profits from a reverse payments, because earliy entry means more competition, which destroys some profits. Signalling, therefore, yields lower profits to the strong entrant than the pooling equilibrium. Both entrant's types are therefore worse off. The weak entrant is clearly worse off because the originator never settles and forces him out of the market, while the strong entrant is worse off because signalling is too costly.

## 3.3 Discussion and conclusions

When the possibility of bankruptcy of the generic entrant is taken into account, banning reverse payments can reduce consumer surplus. The reason is that banning reverse payments reduces the pledgeable profits, which makes the entrant get less than he would when reverse payments are allowed. This exacerbates his possible liquidity problems, possibly making him exit the market. This reduces consumer surplus both before and after patent expiry. This effect goes in the opposite direction of the static effect of banning reverse payments, which makes the parties agree on an earlier entry date (provided that the entrant remains on the market). The negative effect on consumer surplus, due to the possibility that the generic exits the market, can be greater than the consumer surplus loss due to entry upon the latest entry date $\hat{D}$, including when this is the patent expiry date. This suggests that a rule of reason is more suited than a ban *per se.* The negative effect of banning reverse payment is stronger where the patent is strong, the economy is in a downturn and collusion among few players is possible.

For a practical enforcement, in order to know which agreements should be banned, the AA should have a rough estimate of the parameters. These estimates can be recovered from the past ratio of generic firms that went bankrupt (for $\mu$ and $k$), the type of competition in the industry (for the profits and the surplus parameters), the expected life of the drug after the patent expiry (for $\delta$), the assessed patent strength (for $\theta$: this value can be inferred from the experts' advices, from past sentences of the local Patent Office, etc) and the litigation costs of the parties. The closer the estimated parameters are to the inequalities of cases (i) and (iii) of Proposition 11, the more the AA should be lenient towards settlements with reverse payments. A period of economic downturn makes the optimal policy more lenient too.

# 4 Chapter 4: Reverse Payments and Productive Investment

Reverse payments should not be banned *per se* also when the entrant's ability to compete is unknown and he may have to incur additional investments to be competitive. Allowing reverse payments reduces the entrant's incentive to invest and this saving can be higher than the welfare loss due to the later entry due to reverse payments. Under general conditions and various bargaining rules, the threshold for the investment cost is decreasing in the difficulty of the technology, so reverse payments should be treated in a more lenient way when the technology is difficult. When the originator chooses whether to use a reverse payment and the technology is difficult, there is also an irrelevance result: the cap on the generic manufacturer's entry date has no effect on total welfare, as the positive effect of an earlier entry is exactly compensated by a higher probability of settling through a reverse payment (independently from how the settlement surplus is split).

## 4.1 Introduction

This essay analyzes whether a ban *per se* on reverse payments is justified when the entrant may have to make an additional investment to become competitive prior to the settlement negotiations. This investment can be thought as a replication of the originator's investment costs (a social waste) that requires time to be developed. I consider the welfare effects of (i) allowing reverse payments and (ii) having different latest entry dates.[80] Allowing reverse payments is optimal when the social cost of the investment is higher than its social benefit, that consists in more competition on the market in case of successful litigation for the entrant. In this case, a lenient policy towards reverse payments can induce the entrant to avoid replicating these costs, in the view of obtaining a reverse payment that is independent from his ability to compete.

When the technology needed to be competitive is easy, allowing reverse payments brings the equilibrium investment to zero. The entrant is likely competitive, so the originator prefers to settle with a reverse payment in any case. This makes the entrant not invest even in the few cases where he actually cannot compete, because he knows that the originator will always offer a reverse payment. Allowing reverse payments, here, can be beneficial when the investment cost is sufficiently high, which occurs more likely when the technology is harder. The intuition is that, when the technology is easy, the entrant is almost surely competitive, so

[80] Here, differently from the two previous chapters, the focus is on total welfare, not on consumer surplus.

allowing reverse payments yields a lower social benefit (only few entrants need to invest in order to be competitive, so reducing investment to zero is less attractive) but at the same social cost (a later generic manufacturer's entry date for all the settlements). When the technology is sufficiently easy, the irrelevance result does not hold: when reverse payments are optimal, the optimal latest entry date is the minimal date such that (i) the originator offers a reverse payment and (ii) the entrant does not invest. Any later entry date only reduces welfare without changing the parties' behavior, while any shorter entry date is useless, if (i) the originator prefers not to offer a reverse payment, or counterproductive, if (ii) the entrant invests.

When the technology needed to be competitive is difficult, we have a mixed strategy equilibrium, where the originator randomizes between settling with and without a reverse payment, while the entrant randomizes between investing and not. The more lenient the policy towards reverse payments (i.e. the later the maximal allowed entry date), the higher the profits the entrant gets when he has not invested and, therefore, the lower his incentives to invest. This makes the originator offer a reverse payment less often, as it is more likely that the entrant is not able to effectively compete, which in turn increases the entrant's incentive to invest. When the originator chooses whether to use a reverse payment or not, independently from how the surplus is split, the entrant's incentives to invest are restored to the initial level.[81] Therefore, the latest entry date has no effect on the entrant's equilibrium investment.[82] A higher latest entry date, therefore, only *reduces* the originator's willingness to offer a reverse payment. Allowing reverse payments increases welfare as long as the investment cost is sufficiently high - when the savings on the investment costs are higher than the welfare loss due to the later entry. Furthermore, there is an irrelevance result: when the technology is difficult, the cap on the entry date has no effect on welfare, as the negative effect of a later entry is exactly compensated by the lower probability of settling through a reverse payment (independently from how the settlement surplus is split).

This analysis shows that allowing reverse payments can increase welfare, even when the entry date is upon the patent expiry. This makes a *rule of reason* approach more suitable than a ban *per se*.

This paper builds on the literature of litigation and settlement. For a summary,

[81] The fact that the originator has all the bargaining power is a sufficient but not necessary condition for this result. As shown in the Appendix, this result is robust to different sharing rules of the surplus, provided that the originator chooses whether to use a reverse payment or not.

[82] This result is robust not only to different sharing rules of the surplus, but also to a settlement rule that makes the entrant make a take-it-or-leave-it offer.

see the previous chapters. It also builds on the literature about opportunism in getting a favorable settlement, which is not large. Bessen and Meurer (2006) consider the investment decision of an entrant who can invest around a patent - but not in order to become competitive, as here. Their article is not very close to this one, as my aim is to analyze the relationship between the entrant's opportunism in investing to become competitive and the settlement policy.

For simplicity, I stick to a *one-sided* incomplete information game, where the *uninformed* party (the originator) makes a take-it-or-leave-it offer. This allows us to avoid the well known problem of multiplicity of equilibria of signalling games. In the Appendix I show that many of the results are robust to changes in the bargaining rule.

## 4.2 The model

There are three players: an Antitrust Authority (AA), an originator and an entrant. In the first stage, the AA decides which settlement policy to enforce. The AA can allow reverse payments, $\rho = (R, \hat{D})$, with a latest entry date $\hat{D}$, or ban them, $\rho = N$. Normalize the patent length to one, where 0 is when the parties decide on litigating or settling and 1 is the patent expiration. Under $\rho = (R, \hat{D})$, the parties can use a reverse payment, in which case they must choose an entry date for the generic manufacturer not later than $\hat{D} < 1$, or just settle without it; under $\rho = N$, the parties can only agree on an entry date (no reverse payment is allowed). The difference between the two policies, as shown in the previous chapters, lies in that $\rho = (R, \hat{D})$ allows the parties, in equilibrium, to make the originator's monopoly period longer. In the second stage, the entrant learns his type: it can be *weak*, with probability $\mu \in [0, 1]$, or *strong*. When the entrant is weak, he cannot actually compete (say, because of high production costs, an inefficient technology, etc) and prefers to exit the market, with or without the settlement. If the entrant is strong, he is competitive - he can remain on the market, compete and make positive profits. In the third stage, the weak entrant can invest $I$ to become competitive. Both the second and the third stage outcomes are the entrant's private information. In the fourth and fifth stage, the originator and the entrant litigate or settle their dispute. The bargaining process is sequential: in the fourth stage the originator makes a take-it-or-leave-it offer and, in the fifth stage, the entrant accepts or rejects it. In the sixth stage, the entrant decides whether to exit the market.

The timing is then the following:

1. **Policy choice.** The Antitrust Authority implements a policy $\rho \in \{R, N\}$.

2. **Entrant's type.** The entrant learns his own type.

3. **Entrant's investment.** The weak entrant can pay $I$ to become competitive.

4. **Originator's offer.** The originator makes a settlement offer.

5. **Entrant's settlement decision.** The entrant accept or rejects the offer.

6. **Entrant's exit decision.** The entrant decides whether to stay or to exit.

The entrant has, therefore, two ways of being competitive: either he draws the strong signal, or, being weak, he invests. This represents the research process to get a viable product: or a good production process is found at the early stages of its development (strong entrant), or more efforts are necessary (weak entrant) and the entrant must invest to become competitive. This investment can only be done *before* the settlement-litigation stages: this represents a situation where the investment is a lengthy process that needs time to be completed. This represents the pharmaceutical industry, where the investments to improve the process of manufacturing a drug typically take time.[83]

If the originator accepts the offer, $D$ and $R$ are enforced. If he rejects it, the parties litigate. The originator's patent has strength $\theta$. For simplicity, we just refer to the "patent strength", which embeds both the probability of infringement by the entrant and the patent validity. In case of litigation, the originator and the entrant bear, respectively, litigation costs $C_O$ and $C_E$. In the pre-expiry period, denote $H$ the originator's profits if he is the sole supplier on the market for the entire patent period, $L$ if entry occurs immediately and $E$ the entrant's profits if it enters immediately. Like in the previous chapters, assume that $H > L + E$. If the entrant is weak and has invested, his net payoff is reduced by $I$. We assume that $I < [(1-\theta)E - C_E]$: this implies that the entrant is willing to invest if he is sure that litigation will take place. Denote $q$ the probability that the entrant invests.

If the parties *litigate*, they expect to obtain:

Incumbent: $\mu(1-q)H + [1-\mu(1-q)][\theta H + (1-\theta)L - C_O]$

Entrant: $(1-\theta)E - C_E$ if he is competitive, $\pi_E^{LM} = 0$ otherwise.

By litigating, the originator knows that with a probability $\mu(1-q)$ the entrant is not competitive, in which case refusing to offer a reverse payment drives him out

[83] Results would be robust also to allowing investment after the settlement-litigation stage, provided that the time required for the investment to be completed is sufficiently high compared to the patent length.

of the market and allows the originator to enjoy monopoly profits. The probability that the entrant is not competitive is the probability that he is weak $\mu$ *and* has not invested $(1-q)$. If the entrant is competitive, which occurs with probability $[1-\mu(1-q)]$, he remains on the market and litigates: the originator has a probability $\theta$ of winning the case, in which case it gets $H$, and probability $1-\theta$ of losing it and get $L$. Whether he wins or loses, provided that the entrant is competitive (and, therefore, litigation actually occurs), litigation costs are $C_O$. The competitive entrant knows that he has a probability $(1-\theta)$ of winning, in which case he gets $E$, otherwise he earns nothing. His litigation costs are $C_E$. The non-competitive entrant, on the other hand, knows that he is not be able to stay on the market, so in case of litigation he exits. His litigation payoff is therefore 0.

If parties *settle* without a reverse payment, they expect to obtain:
Originator: $\mu(1-q)H + [1-\mu(1-q)][DH + (1-D)L]$
Entrant: $(1-D)E$ if he is competitive, $\pi_E^{SM} = 0$ otherwise.

These settlement profits resemble the litigation payoffs closely. The only differences are in that we have (i) $D$ instead of $\theta$ and (ii) no litigation cost. Upon a settlement, indeed, the parties agree on an entry date $D$ and avoid paying the litigation costs. This makes a settlement more attractive than litigation for both the originator and the entrant. The non-competitive entrant gets 0 because he prefers to exit, as he would not be able to compete anyway, while the competitive one remains and gets $E$ in the post-entry period $(1-D)$.

If parties settle with a reverse payment, they expect to obtain:
Originator: $\mu(1-q)H + [1-\mu(1-q)][\hat{D}H + (1-\hat{D})L] - R$
Entrant: $(1-\hat{D})E + R$ if he is competitive, $\pi_E^{SM} = R$ otherwise.

Note that when the originator offers a reverse payment the non-competitive entrant gets a positive payoff $R > 0$. This is one of the driving forces of the results. In the following subsection I describe the equilibrium behavior.

### 4.2.1 Settlement stage

The competitive entrant accepts the originator's settlement offer as long as his payoff is higher than under litigation. Assume that the entrant accepts the offer when he is indifferent between litigation and settlement. When the originator offers a settlement *without* a reverse payment, the entrant accepts it when: $(1-D)E \geq (1-\theta)E - C_E$, i.e. as long as

$$D \leq D^* = \theta + \frac{C_E}{E} \tag{4.1}$$

When the originator offers a settlement *with* a reverse payment, we know from the previous chapters that (i) for this to be rational for the originator we need $\hat{D} \geq D^* = \theta + \frac{C_E}{E}$ and (ii) the originator will set the entry date equal to the latest possible date: $D = \hat{D}$.[84] The competitive entrant accepts it as long as $(1 - \hat{D})E + R \geq (1 - \theta)E - C_E$, i.e. as long as

$$R \geq R^* = (\hat{D} - \theta)E - C_E \tag{4.2}$$

The non-competitive entrant accepts any (non-negative) offer, as his litigation payoff is zero.

In equilibrium, when reverse payments are banned, the originator has a dominant strategy: to offer a settlement with an entry date $D = D^* = \theta + \frac{C_E}{E}$. With such a strategy, (i) if the entrant is competitive, he is left indifferent between accepting and litigating, and (ii) if he is not, the originator gets the full monopoly profits, as the non-competitive entrant will exit the market. Clearly no strategy is better than this.

In particular, litigating is weakly dominated by settling without reverse payments: if the entrant is non-competitive, both strategies yield the same outcome (the monopoly profits) and, if he is competitive, the originator (i) saves his litigation costs and (ii) obtains a later entry date ($D = \theta + \frac{C_E}{E}$) than the expected one under litigation ($\theta$).

When reverse payments are allowed, on the other hand, the optimal strategy is less trivial: offering a reverse payment such that the competitive entrant is indifferent between litigating and settling can be not optimal anymore. When the entrant is not competitive, indeed, the originator is wasting money. This creates a tension between maximizing the industry profits (setting $D = \hat{D}$ and using $R = R^* = (\hat{D} - \theta) - C_E$ to make the competitive entrant indifferent between settling and litigating) and betting on the weakness of the entrant. This bet consists of offering a settlement without reverse payment, in the hope that the entrant is not competitive and leaves the market. In turn, this gives the weak entrant an incentive to invest: when the probability that the originator does not offer a reverse payment is positive, the weak entrant may prefer to invest and get positive profits after the agreed entry date $D^* = \theta + \frac{C_E}{E}$.[85]

When reverse payments are allowed, the originator decides whether to use them or not in his settlement offer. The weak entrant has two strategies too: to invest to become competitive or not. Given that the entrant's type and his investment

[84] The entrant would accept an entry date equal to $D^* = \theta + \frac{C_E}{E}$ in a settlement without reverse payment. Therefore, an entry date smaller than $D^*$ would reduce the monopoly period, so a settlement with reverse payment and entry date $\hat{D} < D^*$ would never be used.

[85] Provided that these profits are greater than the investment cost.

decision are his private information, we can represent the parties' decisions as a normal-form game.

The payoff table between the originator and the *weak* entrant is given in the following table. "R" means a settlement with reverse payment and "D" means a settlement without it.

| | Invest | Not invest |
|---|---|---|
| R | $\underline{\hat{D}H + (1-\hat{D})L - (\hat{D}-\theta)E + C_E}$, $(1-\theta)E - C_E - I$ | $H - (\hat{D}-\theta)E + C_E$, $\underline{(\hat{D}-\theta)E - C_E}$ |
| D | $(\theta + \frac{C_E}{E})H + (1-\theta-\frac{C_E}{E})L$, $\underline{(1-\theta)E - C_E - I}$ | $\underline{H}$, 0 |

$\{R, D\}$ are the originator's strategies and $\{Invest,\ Not\ Invest\}$ are the entrant's ones. The first element in each cell is the payoff of the originator and the second one the payoff of the entrant. Recall that we assume that the weak entrant prefers to invest if he is sure that the originator will not offer a reverse payment: $I < (1 - \theta - \frac{C_E}{E})E$. The best responses are underlined. Consider now the case where the AA allows reverse payments and sets a latest entry date $\hat{D}$ such that $I > (1-\hat{D})E$. As long as $I > (1-\hat{D})E$, there is no pure strategy Nash equilibrium: if the originator offers a reverse payment, the best response for the entrant is not to invest - but if he does not invest, the originator prefers not to offer a reverse payment, pushing the weak entrant out of the market. If the originator does not offer it, the entrant will prefer to invest, but this makes the originator offer it. We only have a mixed strategy equilibrium. On the other hand, when $I \leq (1-\hat{D})E$ we have a pure strategy equilibrium where the originator offers a reverse payment and the entrant invests. It is easy to show that this equilibrium yields a lower welfare than the one when reverse payments are banned. This is the reason why we can limit our attention to values of $\hat{D}$ such that $I > (1-\hat{D})E$, i.e. $\hat{D} \geq 1 - \frac{I}{E}$.

The reason of the entrant's opportunism is that his incentives to invest shrink if the originator is going to offer a reverse payment - and this effect is stronger the later is the maximal entry date. Note, indeed, that when the originator offers a reverse payment the entrant gets $(1-\hat{D})E + (\hat{D}-\theta)E + C_E$, where $(1-\hat{D})E$ is the post-entry profit and $(\hat{D}-\theta)E + C_E$ the equilibrium reverse payment. A larger $\hat{D}$ makes the incentives to invest smaller, because the cost remains $I$ while the benefit $(1-\hat{D})E$ shrinks.

When $I > (1-\hat{D})E$, we have a matching penny game. Of course, in this category of games we only have a mixed strategy equilibrium.

When the entrant is strong, on the other hand, the game is simpler. The following table represents the payoffs of the originator and the *strong* entrant:

| | |
|---|---|
| | - |
| R | $\underline{\hat{D}H+(1-\hat{D})L-(\hat{D}-\theta)+C_E}$, $(1-\theta)E-C_E$ |
| D | $(\theta+\frac{C_E}{E})H+(1-\theta-\frac{C_E}{E})L$, $(1-\theta)E-C_E$ |

From the matrices above, note that when the originator offers a reverse payment a more lenient policy (higher $\hat{D}$) has the following effects: (i) higher originator's profits when the entrant is competitive; (ii) lower originator's profits when the entrant is non-competitive;[86] (iii) no impact on the entrant's profits when he is competitive;[87] (iv) higher entrant's profits when he is non-competitive.

It is common knowledge that the payoff matrix is the former with probability $\mu$ and the second one with probability $(1-\mu)$. Denote $p^{\rho}$ the probability that the originator settles and $q^{\rho}$ the probability that the entrant invests when the policy is $\rho$. Therefore, when reverse payments are allowed, the originator offers a reverse payment $R=(\hat{D}-\theta)E-C_E$ if:[88]

$$[1-\mu(1-q)][\hat{D}H+(1-\hat{D})L]-(\hat{D}-\theta)E+C_E \geq [1-\mu(1-q)][(\theta+\frac{C_E}{E})H+(1-\theta-\frac{C_E}{E})L]$$

which yields

$$\mu \leq \mu^*(q)=\frac{(H-L-E)}{(1-q)(H-L)}$$

When $q=0$, i.e. the weak entrant never invests, the originator offers a reverse payment if:

$$\mu \leq \mu^* = \frac{(H-L-E)}{(H-L)} \tag{4.3}$$

When the probability that the entrant is weak is too small, the originator always offers a reverse payment. The originator risks too much from a settlement without it, so he prefers to offer a reverse payment that makes the competitive entrant willing to settle, even if sometimes he wastes money with a non-competitive entrant. The idea is that it is so unlikely that the entrant is non-competitive, even when he never invests in case he is weak ($q=0$), that it is never worthwhile to take the risk of facing a competitive one. The probability of getting the monopoly profits by offering a settlement without reverse payment and pushing the weak entrant out of the market is so small that the originator always prefers to offer a reverse payment. Being $I>(1-\hat{D})E$, when $\mu \leq \mu^*$ the entrant never invests.

[86] It can be shown (see the Appendix) that the originator has no incentive to offer a reverse payment positive but smaller than the maximal one. Therefore, the originator will always offer it and ask for the latest entry date $\hat{D}$. This makes his profits smaller when he offers a reverse payment and the entrant is not competitive.

[87] This is due to the originator making a take-it-or-leave-it offer. The basic results, however, are robust to modifications of this bargaining rule, as shown in the Appendix.

[88] $\mu(1-q)H$ on both sides have been simplified.

On the other hand, when the probability that the entrant is weak $\mu$ is higher than $\mu^*$, we have a mixed strategy equilibrium. The originator, independently from the policy $\rho$, proposes a settlement with a reverse payment when:

$$q \geq q^* = 1 - \frac{(H - L - E)}{(H - L)\mu} \tag{4.4}$$

and the weak entrant invests if:

$$p^\rho[(1 - \hat{D})E + (\hat{D} - \theta)E - C_E] + (1 - p^\rho)(1 - \theta - \frac{C_E}{E}) - I \geq p^\rho[(\hat{D} - \theta)E - C_E]$$

In the LHS, the first addend is the probability $p^\rho$ that the originator offers a reverse payment times the competitive entrant's settlement profits (that are the market profits $(1 - \hat{D})E$ plus the reverse payment $(\hat{D} - \theta)E - C_E$). The second addend is the probability that the originator offers a settlement without reverse payment times the competitive entrant's settlement profits $(1-\theta-\frac{C_E}{E})$. The RHS is the payoff from not investing, that is simply the probability of getting a reverse payment times its size. The entrant, under policy $\rho$, invests when

$$p \leq p^{\rho*} = \frac{(1 - \theta)E - C_E - I}{(\hat{D} - \theta)E - C_E} \tag{4.5}$$

This leads to the following Proposition.

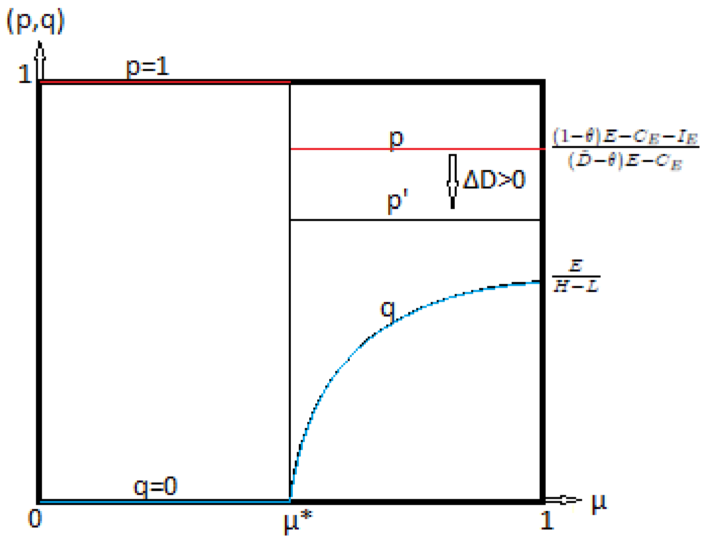
**Proposition 14** *When $\mu > \mu^*$, there is a unique mixed strategy equilibrium where the probability that the originator offers a reverse payment is $p^{\rho*} = \frac{(1-\theta)E-C_E-I}{(\hat{D}-\theta)E-C_E}$ and the probability that the entrant invests is $q^* = 1 - \frac{(H-L-E)}{(H-L)\mu}$.*

**Corollary 3** *The threshold $\mu^*$ and the probability that the entrant invests $q^*$ are independent from the policy $\rho$. The policy has an impact (only) on the probability that the originator offers a reverse payment $p^\rho$: a more lenient policy lowers the probability that the originator offers it.*

The probability $p^{\rho*} = \frac{(1-\theta)E-C_E-I}{(\hat{D}-\theta)E-C_E}$ that the originator offers a reverse payment depends positively on the entrant's litigation profits net of the investment cost (the numerator) and negatively on the size of the reverse payment (the denominator). A more lenient policy (a later entry date $\hat{D}$) increases the size of the reverse payment and, therefore, the entrant's incentive not to invest. This reduces the incentives to offer a reverse payment for the originator, thus reducing $p^{\rho*}$. This, in turn, increases the incentives to invest for the entrant. This effect exactly compensates the first effect: the equilibrium investment does not depend on the policy $\rho$.[89]

[89]This exact compensation is robust to changes in the rule that the originator makes a take-it-or-leave-it offer. With a bargaining rule that makes (i) the originator still decide whether to offer a reverse payment or not and (ii) the entrant get an equal share of the settlement surplus, the entrant's incentives to invest remain unchanged. This is shown in the Appendix.

A more lenient policy, therefore, only *reduces* the ratio of settlements involving a reverse payment. This counterintuitive result will prove useful in designing the optimal policy. The following graph represents the equilibrium settlement and investment probabilities $(p, q)$ for different values of $\mu$.




The x-axis represents $\mu$ and the y-axis represents $p^{\rho*}$ and $q^*$. When reverse payments are allowed and the originator uses them, the probability that the originator offers a reverse payment is $p^{\rho*} = 1$ for $\mu \leq \mu^*$ and $p^{\rho} = \frac{(1-\theta)E-C_E-I}{(\hat{D}-\theta)E-C_E}$ for $\mu > \mu^*$. The probability that the weak entrant invests is $q^* = 0$ for $\mu \leq \mu^*$ and $q^* = 1 - \frac{H-L-E}{(H-L)\mu}$ for $\mu > \mu^*$. We have, therefore, two areas: (i) $0 \leq \mu \leq \mu^*$: the originator always offers a reverse payment, as the probability that the entrant is strong is too high; this makes the weak entrant never invest. (ii) $\mu^* \leq \mu \leq 1$: both the originator offers a reverse payment and the entrant invests with a positive probability smaller than one; the probability that the weak entrant invests is increasing with $\mu$. A more lenient policy (later allowed entry date $\hat{D}$ and therefore higher maximal reverse payment) increases the profits of the non-competitive entrant, which reduces his incentives to invest. This makes the originator offer a reverse payment less often, as it is more likely that the entrant is not able to effectively compete, which in turn increases the entrant's incentive to invest. These two effects offset each other and the entrant's probability of investing remains at the initial level, while the probability that the originator offers a reverse payment becomes lower. Therefore, under a lenient policy the originator offers a reverse payment fewer times and the equilibrium investment does not change. When reverse payments are banned, the originator cannot offer them and the entrant invests with probability one, because $I < (1 - \theta - \frac{C_E}{E})E$. The welfare implications are discussed in the next subsection.

### 4.2.2 Welfare implications

Denote $\bar{W}$ the (total) duopoly welfare when the entrant is on the market, $\underline{W}$ the monopoly welfare and $W^{\rho}$ the welfare when the parties settle under policy $\rho$ and the entrant is competitive. When the entrant is not competitive, independently from whether a reverse payment has occurred, the originator remains the monopolist, so welfare is $\underline{W}$. The policy $\rho$ has an impact on the settlement welfare: a more lenient policy allows the parties to delay the generic manufacturer's entry and, therefore, to make the monopoly period longer. So, total welfare is $W^R = \hat{D}\underline{W} + (1-\hat{D})\bar{W}$ when the parties settle with a reverse payment and the entrant is competitive, $W^N = (\theta + \frac{C_E}{E})\underline{W} + (1-\theta-\frac{C_E}{E})\bar{W}$ when the parties settle without a reverse payment and the entrant is competitive and $\underline{W}$ when the entrant is not competitive. We assume, of course, $\bar{W} > \underline{W}$. When the entrant is weak and invests, total welfare is reduced by $I$.

The following table shows the six possible cases.

| *Outcome* | *Probability* | *Welfare* |
|---|---|---|
| R - Weak&Invest | $\mu p^{\rho *} q^*$ | $\hat{D}\underline{W} + (1-\hat{D})\bar{W} - I$ |
| R - Weak&No inv | $\mu p^{\rho *}(1-q^*)$ | $\underline{W}$ |
| R - Strong | $(1-\mu)p^{\rho *}$ | $\hat{D}\underline{W} + (1-\hat{D})\bar{W}$ |
| D - Weak&Inv | $\mu(1-p^{\rho *})q^*$ | $(\theta + \frac{C_E}{E})\underline{W} + (1-\theta-\frac{C_E}{E})\bar{W} - I$ |
| D - Weak&No inv | $\mu(1-p^{\rho *})(1-q^*)$ | $\underline{W}$ |
| D - Strong | $(1-\mu)(1-p^{\rho *})$ | $(\theta + \frac{C_E}{E})\underline{W} + (1-\theta-\frac{C_E}{E})\bar{W}$ |

When $\mu < \mu^*$, we know that the originator always offers a reverse payment ($p = 1$) and the entrant never invests ($q = 0$). Therefore, when the entrant is strong, during a fraction of time $\hat{D}$ we have a monopoly and during a fraction $(1-\hat{D})$ we have a duopoly; when he is weak, he never invests, so we just have the monopoly welfare. This yields:

$$E[W|R, \mu < \mu^*] = (1-\mu)[\hat{D}\underline{W} + (1-\hat{D})\bar{W}] + \mu\underline{W} \tag{4.6}$$

When $\mu \geq \mu^*$, we have the mixed strategy equilibrium with $p^{\rho *} = \frac{(1-\theta)E - C_E - I}{(\hat{D}-\theta)E - C_E}$ and $q^* = 1 - \frac{(H-L-E)}{(H-L)\mu}$. Using the welfare values from the previous table, after some rearrangements the expected welfare is:

$$E[W|R, \mu \geq \mu^*] = \underline{W} + \frac{I}{H-L}[(1-\mu)(H-L) + \bar{W} - \underline{W} - E] \tag{4.7}$$

This leads to the following proposition.

**Proposition 15** *When $\mu < \mu^*$, the degree of policy leniency $\hat{D}$ has a negative impact on welfare; when $\mu \geq \mu^*$, it has no effect.*

When $\mu < \mu^*$, $\hat{D}$ has a negative impact on $W$ because it just delays entry without reducing the investment. A higher $\hat{D}$ increases the non-competitive entrant's payoff and reduces his incentives to invest, but he never invests also when reverse payments are banned. There is no point, therefore, in reducing his incentives to invest: a higher $\hat{D}$ only makes the monopoly period longer.

When $\mu \geq \mu^*$, on the other hand, this negative effect is counterbalanced by a positive one: a longer $\hat{D}$ makes the originator settle without a reverse payment more often and, in these cases, entry occurs earlier. Under quite general bargaining rules, these two effects exactly compensate each other.[90] We have, therefore, an irrelevance result: a longer $\hat{D}$ makes the settlements with a reverse payment more harmful, but also less frequent.

**Proposition 16** *When $\mu \geq \mu^*$, the investment cost $I$ has a positive impact on welfare as long as the difference in consumer surplus from passing from monopoly to duopoly is greater than the originator's profit loss multiplied by the probability that the entrant is weak: $\bar{S}- \underline{S} > \mu(H-L)$.*

**Proof.** Denote $W = S + \Pi$, where $S$ is consumer surplus and $\Pi$ the industry profits. We have $\bar{W} = \bar{S} + L + E$ and $\underline{W} = \underline{S} + H$. We have $\frac{dE[W|R,\mu\geq\mu^*]}{dI} = \frac{(1-\mu)(H-L)+\bar{W}-\underline{W}-E}{H-L}$, whose numerator can be rewritten as $\bar{S}- \underline{S} -\mu(H-L)$, which gives the result.

This counterintuitive result is due to the fact that a higher investment cost implies a higher social cost for a given ratio of settlements with reverse payments, but also it reduces the probability $p^{\rho*}$ that the originator offers it (and does not change the probability $q^*$ that the entrant invests). This makes it more likely that a settlement without reverse payments takes place, which makes entry occur earlier. A higher investment cost can inrease welfare not through a reduction of investment, but through a reduction of the use of reverse payments. This positive effect is more important when the difference in consumer surplus ($\bar{S}- \underline{S}$) is larger. The probability $\mu$ that the entrant is weak makes it more difficult that the investment cost has a positive impact because a higher $\mu$ means that the total investment cost $\mu q^* I$ will be higher. $(H-L)$ is a measure of the type of competition in the industry: the higher the profit loss, the more difficult it is that a higher investment cost increases total welfare. Overall, a higher investment cost is likely to increase welfare when consumers gain a lot from the increased competition, when the technology is not too difficult and the monopolist would not lose much because of the generic entry.

[90] In the Appendix I show that this is robust to different sharing rules, as long as the originator decides whether to give a reverse payment or not.

When reverse payments are banned ($p^N$ is constrained to be 0), we know that the entrant invests ($q = 1$). This implies a social cost of $\mu I$. The agreed entry date for the generic is $D = \theta + \frac{C_E}{E}$. So we have:

$$E[W|N] = (\theta + \frac{C_E}{E})\underline{W} + (1 - \theta - \frac{C_E}{E})\bar{W} - \mu I \tag{4.8}$$

Equations (4.6), (4.7) and (4.8) give us the welfare under the different policies and under different parameter sets. Through a simple comparison of (4.8) with (4.6) and (4.7) one can derive the optimal policy $\rho^*$. The following Proposition considers the case of an easy technology.

**Proposition 17** *When $\mu < \mu^*$,*
*if $\theta + \frac{C_E}{E} \geq 1 - \frac{I}{E}$ and $I \geq I^*_\theta = (\bar{W} - \underline{W})[1 - \theta - \frac{C_E}{E}]$, then $\rho^* = (R, \hat{D} = \theta + \frac{C_E}{E})$;*
*if $\theta + \frac{C_E}{E} \geq 1 - \frac{I}{E}$ and $I < I^*_\theta$, then $\rho^* = N$.*
*if $\theta + \frac{C_E}{E} \leq 1 - \frac{I}{E}$ and $I \geq I^*_I = \frac{(\bar{W} - \underline{W})[1 - \theta - \frac{C_E}{E}]E}{(1-\mu)(\bar{W} - \underline{W}) + \mu E}$, then $\rho^* = (R, \hat{D} = 1 - \frac{I}{E})$;*
*if $\theta + \frac{C_E}{E} \leq 1 - \frac{I}{E}$ and $I < I^*_I$, then $\rho^* = N$.*

**Proof.** From a comparison between (4.6) and (4.8), we see that, when the technology is easy ($\mu < \mu^*$), allowing reverse payments with entry date $\hat{D}$ yields higher welfare as long as

$$I \geq (\bar{W} - \underline{W})[\frac{\hat{D} - \theta - \frac{C_E}{E}}{\mu} + 1 - \hat{D}]$$

When this is the case, this policy makes the entrant cease investing, which improves welfare because the investment cost is higher than the welfare gain due to higher competition. Raising $\hat{D}$ further, however, only reduces welfare - see (4.6). Therefore, the optimal policy consists in allowing reverse payments and choosing the minimal entry date $\hat{D}$ such that the originator offers it ($\hat{D} \geq \theta + \frac{C_E}{E}$) and the entrant does not invest when the originator offers it ($\hat{D} \geq 1 - \frac{I}{E}$). Then, the latest entry date $\hat{D}$ must be equal to $\max\{(\theta + \frac{C_E}{E}), (1 - \frac{I}{E})\}$ and the threshold for the investment cost is

$$I^* = (\bar{W} - \underline{W})[\frac{\max\{(\theta + \frac{C_E}{E}), (1 - \frac{I}{E})\} - \theta - \frac{C_E}{E}}{\mu} + 1 - \max\{(\theta + \frac{C_E}{E}), (1 - \frac{I}{E})\}]$$

When $\theta + \frac{C_E}{E} \geq 1 - \frac{I}{E}$, this threshold is

$$I^* = (\bar{W} - \underline{W})[1 - \theta - \frac{C_E}{E}]$$

and, when $\theta + \frac{C_E}{E} < 1 - \frac{I}{E}$, the threshold is

$$I^* = \frac{(\bar{W} - \underline{W})[1 - \theta - \frac{C_E}{E}]E}{(1 - \mu)(\bar{W} - \underline{W}) + \mu E}$$

This is the entry date such that (i) the investment is reduced to zero and (ii) the monopoly period is minimized.[91]

Note that even $\hat{D} = 1$ can be better than banning reverse payments when the investment cost is sufficiently high: $I \geq (\bar{W} - \underline{W})[\frac{1-\theta-\frac{C_E}{E}}{\mu}]$. The threshold for the investment cost is decreasing in $\mu$, which means that the harder the technology, the better it is to allow for reverse payments. Intuitively, the easier the technology, the less a policy allowing reverse payments benefits welfare, because only few entrants need to invest to become competitive. The cost of allowing them remains the same while the benefit is reduced, so it is more difficult that reverse payments increase welfare for an easy technology. The following Proposition shows the optimal policy when the technology is difficult.

**Proposition 18** *When $\mu \geq \mu^*$,*
*if $I \geq \hat{I} = (\bar{W} - \underline{W})\frac{(1-\theta-\frac{C_E}{E})(H-L)}{(\bar{W}-\underline{W}+H-L-E)}$, then $\rho^* = (R, \hat{D} \geq \max\{(\theta + \frac{C_E}{E}), (1-\frac{I}{E})\})$;*
*if $I < \hat{I}$, then $\rho^* = N$.*

**Proof.** Same as above, by comparing (4.7) and (4.8).

When $\mu$ is high, the parties randomize their strategies. Note that, here, any $\hat{D}$ greater or equal than $\max\{(\theta + \frac{C_E}{E}), (1 - \frac{I}{E})\}$ is optimal, differently from the case of the easy technology. The reason is that a higher $\hat{D}$ makes settlements with reverse payment worse, but it makes them occur less often because it reduces $p^{\rho*}$. When the technology to produce the good is more difficult than the threshold $\mu^*$, an early limit on the entry date (or, conversely, a limit on the size of the reverse payment) does not impact welfare. Note also that the threshold for the investment cost does not depend on $\mu$. As long as the technology is sufficiently difficult, the fact that it is very difficult or just above the threshold does not change the benefits and the costs of allowing reverse payments.

**Corollary 4** *When $\theta + \frac{C_E}{E} \geq 1 - \frac{I}{E}$, a necessary and sufficient condition for $I^*$ to be higher than $\hat{I}$ is $\bar{S} - \underline{S} > H - L$.*

**Proof.** Just compare $I^*_\theta$ and $\hat{I}$. Recall that $\bar{W} = \bar{S} + L + E$ and $\underline{W} = \underline{S} + H$. After some simple algebra we get $I^*_\theta \geq \hat{I}$ when $\bar{S} - \underline{S} > H - L$.

When the patent is sufficiently strong, the threshold for the investment cost is decreasing over the difficulty of the technology as long as the difference in consumer

[91] One can easily check that if $\theta + \frac{C_E}{E} < \hat{D} < 1 - \frac{I}{E}$ the originator offers a reverse payment and welfare is lower than when they are banned, because the entrant does not reduce his investment. With such a $\hat{D}$, the unique equilibrium is in pure strategies where the originator offers a reverse payment and the entrant invests. The entry date is delayed compared to a settlement without reverse payments without any change on the equilibrium investment.

surplus from passing from monopoly to duopoly is larger than the originator's profit loss. This condition is more easily fulfilled when competition increases consumer surplus substantially with respect to the profit loss, e.g. when goods are perceived as differentiated.[92] When this is the case, a product with a more difficult technology always requires that reverse payments be treated in a more lenient way. In other words, a more difficult technology makes it more likely that a lenient policy increases welfare.[93]

**Corollary 5** *A higher patent strength $\theta$ makes it more likely that allowing reverse payments increases total welfare.*

**Proof.** The derivatives of $I_\theta^*$, $I_I^*$ and $\hat{I}$ with respect to $\theta$ are negative.

The intuition is clear: a higher patent strength makes the loss due to reverse payments smaller (entry would be late in any case), while the benefit does not change (reducing the duplication of the investment). This applies to both easy and difficult technologies.

In the following section I provide a numerical example.

### 4.2.3 Numerical example

Assume that $H = 40$, $L = 20$, $E = 10$, $\bar{S} = 15$, $\underset{\sim}{S} = 0$. Such a low consumer surplus under monopoly can be due, for example, to an identical consumers' willingness to pay up to a given quantity.[94] Given these values, we have $\mu^* = \frac{1}{2}$.

If $\mu > \frac{1}{2}$, we know from Proposition 18 that the optimal policy is

$$\rho = (R, \hat{D} \geq \max\{(\theta + \frac{C_E}{E}), (1 - \frac{I}{E})\})$$

when

$$I \geq \hat{I} = (\bar{W} - \underline{W})\frac{(1 - \theta - \frac{C_E}{E})(H - L)}{(\bar{W} - \underline{W} + H - L - E)}$$

provided that the entrant invests when he is sure that the originator will not offer a reverse payment: $I < (1 - \theta - \frac{C_E}{E})E$.

Given the values assumed above, we have $\bar{W} = \bar{S} + L + E = 45$, $\underline{W} = \underset{\sim}{S} + H = 40$. Substituting these values in the threshold for the investment cost, we

[92] This is the case for many drugs. Notwithstanding the same formulation, consumers are often ready to pay a much higher price for the branded drug.

[93] When $\theta + \frac{C_E}{E} < 1 - \frac{I}{E}$, the condition for $I^*$ to be higher than $\hat{I}$ is (i) $\frac{E}{\bar{W} - \underline{W}} > \frac{H-L}{\bar{S} - \underset{\sim}{S}}$ when $\bar{S} - \underset{\sim}{S} \geq H - L$ and (ii) it is always higher when $\bar{S} - \underset{\sim}{S} < H - L$ (and equal to $\hat{I}$ when $\mu = \mu^*$).

[94] The monopolist sets a price equal to the consumers' willingness to pay and gets all the surplus.

get that allowing reverse payments is optimal when

$$I > \hat{I} = 5 * \frac{(1-\theta-\frac{C_E}{10})20}{15} = \frac{20}{3}(1-\theta-\frac{C_E}{10})$$

The investment cost must also be smaller than $(1-\theta-\frac{C_E}{E})E$ to make the entrant invest with positive probability, so $I < 10(1-\theta-\frac{C_E}{10})$. There is, therefore, a range between $\frac{20}{3}(1-\theta-\frac{C_E}{10})$ and $10(1-\theta-\frac{C_E}{E})$ for the investment cost such that the optimal policy is to allow reverse payments, even for $\hat{D}=1$.

If $\mu \leq \frac{1}{2}$, the optimal policy is given by Proposition 17. When $\theta+\frac{C_E}{10} \geq 1-\frac{I}{10}$, then it is optimal to allow reverse payment (with entry date $\hat{D}=\theta+\frac{C_E}{10}$) if $I \geq I^* = 5[1-\theta-\frac{C_E}{10}]$, while when $\theta+\frac{C_E}{10} < 1-\frac{I}{10}$ it is optimal to allow them (with entry date $\hat{D}=1-\frac{I}{10}$) if

$$I \geq I^* = \frac{50[1-\theta-\frac{C_E}{10}]}{5(1-\mu)+10\mu} = \frac{10}{1+\mu}[1-\theta-\frac{C_E}{10}]$$

The investment cost must still be smaller than $10(1-\theta-\frac{C_E}{E})$ to make the entrant invest when the originator does not offer a reverse payment. This implies that the range of values for $I$ such that allowing reverse payments is optimal spans from $5[1-\theta-\frac{C_E}{10}]$ to $10[1-\theta-\frac{C_E}{10}]$ when $\theta+\frac{C_E}{10} \geq 1-\frac{I}{10}$ and from $\frac{10}{1+\mu}[1-\theta-\frac{C_E}{10}]$ to $10(1-\theta-\frac{C_E}{E})$ when $\theta+\frac{C_E}{10} < 1-\frac{I}{10}$.

## 4.3 Discussion and conclusions

When the opportunism in the entrant's investment decision to become competitive is taken into account, banning reverse payments can *reduce* total welfare.[95] The reason is that allowing them increases the entrant's profits when he is not competitive. This reduces his incentives to duplicate the originator's investment, which has a positive impact on welfare (R&D expenses are not replicated) and a negative one (the entrant competes less often). As long as the investment cost is sufficiently high, the former effect dominates. Under quite general conditions, the threshold for the investment cost is decreasing in the difficulty of the technology, so reverse payments for a drug with a more difficult technology should be treated in a more lenient way.

When allowing reverse payments is optimal and the technology is easy, the optimal allowed entry date is the minimum that makes the originator offer a reverse payment and the entrant not invest. When allowing reverse payments is optimal and the technology is difficult, then the entry date is irrelevant, because a later entry date increases the originator's settlement profits when the entrant is

[95] This holds also when they are allowed but with a too low cap, e.g. a cap equal to the avoided litigation costs of the originator, as has been proposed.

competitive, reduces them when he is not and increase the entrant's profits when he is not competitive. This reduces the entrant's incentives to invest, but also it gives the originator an incentive to make the entrant competitive, so he offers a reverse payment less often. This makes the entrant invest more, which brings the entrant's incentives to invest back to the initial level. The (only) consequence of a later entry date is a lower frequency of reverse payments. A later entry date has therefore a negative but also a positive effect on welfare: the negative one is that a settlement with a reverse payment delays entry, while the positive one is that settlements involving reverse payments occur less often. When the originator chooses whether to offer a reverse payment or not, these two effects perfectly offset each other - even when he does not have the full the bargaining power. Assuming that the originator offers a take-it-or-leave-it offer avoids the well known problem of multiple equilibria in signalling games. These conclusions are robust to modifications on the bargaining rule (see the Appendix).

Patent strength has a role too: the stronger is the patent, the more likely it is that reverse payments increase welfare. The intuition is straightforward: the cost of allowing reverse payments is smaller when the patent is strong (expected entry would occur lately anyway), while the benefit is the same (avoiding the social cost of replicating the investment).

Allowing the entrant to invest after the settlement-litigation stages would make the originator always offer a reverse payment, if they are allowed, making the problem less interesting but with the same policy implications as under the easy technology: it is still optimal to allow reverse payments when the investment cost is sufficiently high. The timing proposed in this paper, where the investment is possible only before the litigation-settlement stage, is consistent with the pharmaceutical industry, where investments take time to become productive. However, allowing investment after the litigation-settlement stage would not change the qualitative results, as long as the remaining patent length is sufficiently small compared to the time needed for the investment to be completed. A shorter remaining patent length, compared to the time needed for the investment to be completed, implies smaller entrant's profits from investing after the settlement-litigation stages. Smaller profits imply a higher incentive for the entrant to invest *before* the settlement-litigation stages. The longer time for the investment to be completed, compared to the remaining patent length, the closer the results to the ones presented in this essay.

A more general model could make the investment cost of the generic manufacturer endogenous. The government can, indeed, choose a patent policy such that it is more difficult to "invent around" a patent - higher $I$ and $\theta$ - or increase the standard, in terms of bioequivalence studies and marketing authorizations, required for a generic drug to be marketed - higher $I$. From Proposition 16, we know that a higher investment cost is beneficial for total welfare when consumers' benefit from competition is large, the technology is easy and the profit loss is small.

## 4.4 Appendix

### 4.4.1 Maximal Reverse Payments

This appendix shows that, when the originator offers a reverse payment, he offers the maximal possible amount $R^* = (\hat{D} - \theta)E - C_E$. In principle, he could prefer to offer a smaller amount and adjust the entry date accordingly, as he is losing money with certainty when the entrant is non-competitive.

Given the duality between $D$ and $R$, we can use $D$ as the originator's decision variable. His profits when he offers a settlement date $\tilde{D} < \hat{D}$ and its dual reverse payment $R(\tilde{D}) = (\tilde{D} - \theta)E - C_E < R^*$.

The following table represents the profits the originator gets for each possible outcome.

| *Outcome* | *Probability* | *O's profits* |
|---|---|---|
| R - Weak&Invest | $\mu p^{\rho *} q^*$ | $\tilde{D}H + (1-\tilde{D})L - (\tilde{D} - \theta - \frac{C_E}{E})E$ |
| R - Weak&No inv | $\mu p^{\rho *}(1-q^*)$ | $H - (\tilde{D} - \theta - \frac{C_E}{E})E$ |
| R - Strong | $(1-\mu)p^{\rho *}$ | $\tilde{D}H + (1-\tilde{D})L - (\tilde{D} - \theta - \frac{C_E}{E})E$ |
| D - Weak&Inv | $\mu(1-p^{\rho *})q^*$ | $(\theta + \frac{C_E}{E})H + (1 - \theta - \frac{C_E}{E})L$ |
| D - Weak&No inv | $\mu(1-p^{\rho *})(1-q^*)$ | $H$ |
| D - Strong | $(1-\mu)(1-p^{\rho *})$ | $(\theta + \frac{C_E}{E})H + (1 - \theta - \frac{C_E}{E})L$ |

This table implies that the originator's profits are equal to

$$\begin{aligned}\pi_O^R &= p^{\rho *}\{[\mu q^* + 1 - q^*][\tilde{D}H + (1-\tilde{D})L - (\tilde{D} - \theta - \frac{C_E}{E})E] + \\ &\quad +\mu(1-q^*)[H - (\tilde{D} - \theta - \frac{C_E}{E})E]\} + \\ &\quad +(1-p^{\rho *})\{[\mu q^* + 1 - q^*][(\theta + \frac{C_E}{E})H + (1 - \theta - \frac{C_E}{E})L] + \mu(1-q^*)H\}\end{aligned}$$

Note, first, that the threshold $\mu^*$ below which we are in a pure strategy equilibrium does *not* depend on the size of the reverse payment ($\tilde{D}$ cancels out). Therefore, for any size of the reverse payment, when $\mu \leq \mu^*$ we have a pure strategy equilibrium and when $\mu \leq \mu^*$ we have the mixed strategy equilibrium.

When $\mu \leq \mu^*$, we the pure strategy equilibrium is such that the originator offers a reverse payment ($p^{\rho *} = 1$) and the entrant never invests ($q^* = 0$).This yields

$$\pi_O^R = \mu H + (1-\mu)[\tilde{D}H + (1-\tilde{D})L] - (\tilde{D} - \theta - \frac{C_E}{E})E$$

The derivative of this with respect to $\tilde{D}$ is

$$\frac{d\pi_O^R}{d\tilde{D}} = (1-\mu)(H-L) - E$$

When this derivative is not negative, the originator has no incentive to offer a smaller reverse payment. Note that the bigger is $\mu$ the smaller is the derivative. We are in the area $\mu \leq \mu^*$, so the maximum value that $\mu$ can get is $\mu^* = \frac{H-L-E}{H-L}$. It is sufficient, therefore, to check the sign of the derivative for $\mu = \mu^*$. When $\mu = \mu^* = \frac{H-L-E}{H-L}$, the derivative is zero (and it is positive for any $\mu < \mu^*$). Therefore, when we are in the area of the pure strategy equilibrium, the originator has no incentive to offer a smaller reverse payment.

When $\mu > \mu^*$, we know that we are in the mixed strategy equilibrium area. From (4.5), we know that the equilibrium probability that the originator offers a reverse payment is $p^{\rho *} = \frac{(1-\theta)E-C_E-I}{(\tilde{D}-\theta)E-C_E}$, that depends negatively on $\tilde{D}$. A lower reverse payment implies a lower entry date $\tilde{D}$, which implies a higher probability of offering a reverse payment. In other words, if the originator offers a lower reverse payment, he must offer it more often. After plugging the values of $p^{\rho *}$ and $q^*$ in the profit function and after some rearrangements, we get

$$\pi_O^R = H - (1-\theta-\frac{C_E}{E})E$$

This profit function is clearly independent from $\tilde{D}$, so the originator has no incentive to lower the equilibrium reverse payment. The intuition is that if he lowers it, he must offer it more often ($p^{\rho *}$ depends negatively on $\tilde{D}$) and this effect perfectly offsets the advantage of paying a smaller reverse payment.

We can conclude that the originator has no incentive to offer a positive reverse payment smaller than $R^*$, both when $\mu < \mu^*$ and $\mu \geq \mu^*$. We can therefore restrict our attention to the polar cases of (i) a settlement with a reverse payment equal to $R^*$ and (ii) a settlement without it.

### 4.4.2 Sharing of the Settlement Surplus

This appendix shows that results are robust to allowing the entrant to get a share of the settlement surplus. For simplicity, consider a rule that (i) still makes the originator choose whether to offer a reverse payment or not and (ii) gives the originator a share $\sigma$ and the entrant a share $1-\sigma$ of the settlement surplus, with respect to each party's litigation payoffs.

When the originator gets all the surplus, we know that $R_O = (\hat{D}-\theta)E - C_E$ and, when reverse payments are not used, $D_O = \theta + \frac{C_E}{E}$. When the entrant gets it (i.e. the entrant makes the take-it-or-leave-it offer), the originator is kept at his litigation payoff, so $R_E = (\hat{D} - \theta)(H - L) + C_O$ and $D_E = \theta - \frac{C_O}{H-L}$. The sharing rule explained above implies a reverse payment equal to

$$\breve{R} = \sigma[(\hat{D} - \theta)E - C_E] + (1 - \sigma)[(\hat{D} - \theta)(H - L) + C_O]$$

and an entry date, when reverse payments are not used, equal to

$$\breve{D} = \sigma[\theta + \frac{C_E}{E}] + (1 - \sigma)[\theta - \frac{C_O}{H - L}]$$

With such a reverse payment, the originator chooses to use it in the settlement offer as long as

$$[1 - \mu(1 - q)][\hat{D}H + (1 - \hat{D})L] - \breve{R} \geq [1 - \mu(1 - q)][\breve{D}H + (1 - \breve{D})L]$$

This yields

$$\mu \leq \breve{\mu}^* = \frac{(H - L - E)\sigma(\hat{D} - \theta - \frac{C_E}{E})}{(H - L)\{(\hat{D} - \theta) + \frac{C_O}{H-L} - \sigma[\frac{C_O}{H-L} + \frac{C_E}{E}]\}}$$

For an easier comparison with $\mu^*$ of expression (4.3), ignore litigation costs. This is a good approximation of $\hat{\mu}^*$ for sufficiently high market profits $H$, $L$ and $E$. In this case, we have

$$\mu \leq \breve{\mu}^* = \frac{\sigma(H - L - E)}{(H - L)} \tag{4.9}$$

Therefore, we have $\breve{\mu}^* \simeq \sigma\mu^*$. A sharing rule such that the entrant gets a positive share of the settlement surplus reduces the threshold below which the originator always offers a reverse payment. In particular, the smaller the bargaining power of the originator, the smaller the threshold. Given that now the originator has to relinquish a part of the surplus, he has a smaller incentive to offer a reverse payment. The area of the mixed-strategy equilibrium widens.

The originator proposes a settlement with a reverse payment $\hat{R}$ when

$$q \geq \breve{q}^* = 1 - \frac{(H - L - E)\sigma(\hat{D} - \theta - \frac{C_E}{E})}{\mu(H - L)\{(\hat{D} - \theta) + \frac{C_O}{H-L} - \sigma[\frac{C_O}{H-L} + \frac{C_E}{E}]\}}$$

Ignoring litigation costs, we get

$$q \geq \breve{q}^* = 1 - \frac{\sigma(H - L - E)}{\mu(H - L)} \tag{4.10}$$

Note that $\breve{q}^*$ is always larger than $q^*$ of expression (4.4). This means that, in equilibrium, the entrant invests more: a sharing rule such that the entrant gets a positive share of the settlement surplus makes the originator less willing to offer a reverse payment and the entrant more willing to invest.

The weak entrant invests when

$$p[(1-\hat{D})E+\breve{R}]+(1-p)(1-\breve{D})E-I \geq p\breve{R}$$

which yields

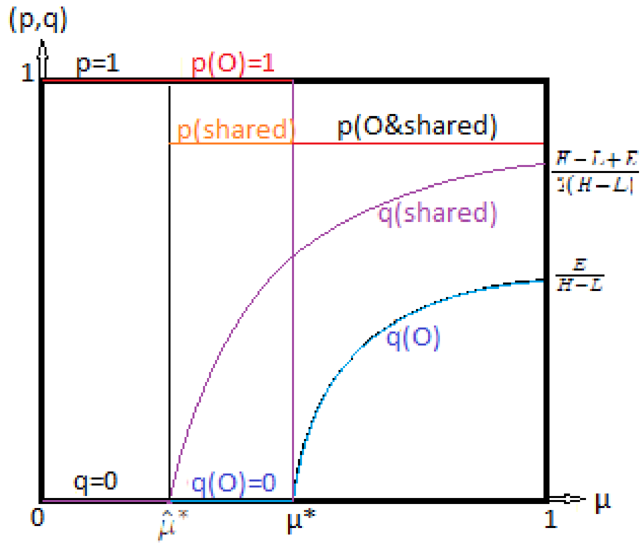
$$p \leq p^{\rho} = \frac{(H-L)[(1-\theta)E-\sigma C_E - I]+(1-\sigma)EC_O}{(H-L)[(\hat{D}-\theta)E-\sigma C_E]+(1-\sigma)EC_O}$$

Ignoring litigation costs, we get

$$p \leq \breve{p}^{\rho *} = \frac{(1-\theta)E-I}{(\hat{D}-\theta)E} \tag{4.11}$$

This is exactly $p^{\rho *}$ of expression (4.5) when we disregard litigation costs. The reason why the threshold $\hat{p}^{\rho *}$ is independent from the size of the reverse payment is that $R$ appears both in the LHS and in the RHS multiplied by $p$. In other words, the only things the entrant considers when deciding whether to invest or not are (i) his litigation payoff minus the investment cost (the numerator) and (ii) the expected market profit reduction due to the settlement (the denominator: entry occurs in $\hat{D}$ instead than in $\theta$).

The mixed strategy equilibrium with such a bargaining rule is given by $\breve{p}^{\rho *} = \frac{(1-\theta)E-I}{(\hat{D}-\theta)E}$ and $\breve{q}^{*} = 1-\frac{\sigma(H-L-E)}{\mu(H-L)}$. The following graph compares it to the equilibrium found in the main text, where the originator makes the take-it-or-leave-it offer. For the graphical representation, assume that $\sigma = \frac{1}{2}$: this yields $\breve{p}^{\rho *} = \frac{(1-\theta)E-I}{(\hat{D}-\theta)E}$ (the orange segment) and $\breve{q}^{*} = 1-\frac{(H-L-E)}{2\mu(H-L)}$ (the violet curve).




The two changes are that (i) the threshold for $\mu$ below which the originator always offers a reverse payment is lower, (ii) the entrant always invests more after the threshold. This sharing rule reduces the originator's surplus from offering a reverse payment: this reduces the originator's incentives to offer it which, in turn, increases the entrant's incentive to invest. This, finally, raises the originator's incentives to offer the reverse payment back: in particular, when litigation costs are arbitrarily small compared to market profits, the originator offers a reverse payment with the same frequency as when he gets all the settlement surplus (the case of the main text).[96] The range for $\mu$ where the players play the mixed-strategy equilibrium widens.

Conclusions remain qualitatively unchanged. In particular:

1) the probability $\breve{p}^{\rho *}$ that the originator offers a reverse payment is 1 until the threshold and equal to the probability $p^{\rho *} = \frac{(1-\theta)E-I}{(\hat{D}-\theta)E}$ of expression (4.5) when litigation costs are arbitrarily small compared to market profits. It is, therefore, still discontinuous in the threshold, independent from the difficulty of the technology $\mu$ and decreasing in the leniency of the policy $\hat{D}$.

2) the probability $\breve{q}^{*}$ that the entrant invests is zero until the threshold and then continuously increasing over $\mu$ after it. It is also still independent from the leniency of the policy $\hat{D}$.

3) the policy implications are unchanged. In particular, when we ignore litigation costs, also the irrelevance result holds: $\hat{D}$ has no effect on total welfare.[97]

The main results, therefore, are robust to a change in the sharing rule that makes the entrant enjoy a share of the settlement surplus.

### 4.4.3 The Entrant makes the Take-it-or-Leave-it Offer

This appendix considers the case where the entrant makes the take-it-or-leave-it offer. Therefore, the timing becomes

1. **Policy choice.** The Antitrust Authority implements a policy $\rho \in \{R, N\}$.
2. **Entrant's type.** The entrant learns his own type.
3. **Entrant's investment.** The weak entrant can pay $I$ to become competitive.
4. **Entrant's offer.** The entrant makes a settlement offer.

[96] When we include litigation costs, the overall effect makes the probability of offering a reverse payment *higher* than when the originator gets all the surplus. This is consistent with the negative relationship between the equilibrium probabilities of offering a reverse payment and $\hat{D}$: a higher profitability of offering a reverse payment for the originator (through a higher $\hat{D}$ or a bargaining rule that allows the originator to get all the settlement surplus) makes him offer them *less often*, because of the interaction with the entrant's incentives to invest.

[97] One can check that by repeating the calculations that led to (4.7) by using $\breve{p}^{\rho *}$ and $\breve{q}^{*}$.

5. **Originator's settlement decision.** The entrant accept or rejects the offer.

6. **Entrant's exit decision.** The entrant decides whether to stay or to exit.

Consider the pooling equilibrium where both entrant's types after the investment decision (competitive and non-competitive) make the same offer. The originator accepts it as long as his payoff is higher than under litigation. When the entrant asks for a settlement *without* a reverse payment, the originator accepts it when:

$$\mu(1-q)H+[1-\mu(1-q)][DH+(1-D)L] \geq \mu(1-q)H+[1-\mu(1-q)][\theta H+(1-(\theta)L]$$

i.e. as long as

$$D \geq \check{D} = \theta - \frac{C_O}{H-L} \tag{4.12}$$

When the entrant asks for a settlement *with* a reverse payment, we know from the previous chapters that (i) for this to be rational for the entrant we need $\hat{D} \geq \check{D} = \theta - \frac{C_O}{H-L}$ and (ii) the entrant will set the entry date equal to the latest possible one: $D = \hat{D}$. The originator accepts it as long as

$$[1-\mu(1-q)][\hat{D}H + (1-\hat{D})L] - R \geq [1-\mu(1-q)][\theta H + (1-(\theta)L - C_O]$$

, i.e. as long as

$$R \leq \check{R} = [1-\mu(1-q)][(\hat{D}-\theta)(H-L) + C_O] \tag{4.13}$$

Note that here, differently from the main text, the size of the reverse payment depends on $\mu(1-q)$. The reason is that the entrant must keep into account that the originator is less willing to accept a reverse payment when the difficulty of the technology is higher ($\mu$) and when the entrant invests less $(1-q)$. The higher the probability $\mu(1-q)$ that the entrant is non-competitive, the less attractive it is for the originator to give a reverse payment. When the entrant makes the take-it-or-leave-it offer, this means a lower reverse payment; when the originator makes it, as in the main text, this means a (weakly) lower probability of a settlement with a reverse payment.[98]

Consider now the entrant's incentives to ask for a reverse payment. The non-competitive entrant always prefers to ask for a reverse payment: any $R > 0$ gives him a positive payoff, while a settlement over the entry date $D$ gives him a zero

[98] In the main text, the originator offers a reverse payment with probability one if $\mu < \mu^*$ and only with a positive probability $p^\rho$ smaller than one if $\mu \geq \mu^*$.

payoff. The competitive entrant will also ask for a reverse payment as long as his profits are higher than the ones from a settlement without it:

$$(1-\hat{D})E + [1-\mu(1-q)][(\hat{D}-\theta)(H-L) + C_O] \geq (1-\theta+\frac{C_O}{H-L})E$$

$(1-\hat{D})E$ represents the market profits the entrant makes after the agreed entry date, $[1-\mu(1-q)][(\hat{D}-\theta)(H-L)+C_O]$ is the reverse payment and $(1-\theta+\frac{C_O}{H-L})E$ the entrant's profits in a settlement without a reverse payment, when the agreed entry date is $\check{D} = \theta - \frac{C_O}{H-L}$. A *sufficient* condition for this inequality to be true is $\hat{D} \geq \theta - \frac{C_O}{H-L}$, i.e. the condition needed for the use of reverse payments to be rational for the entrant.[99] Therefore, the competitive entrant always asks for a reverse payment. Both entrant's types, then, ask for a reverse payment $R = \check{R}$.

Consider now the weak entrant's incentives to invest. Given that he will ask for a reverse payment $R = \check{R}$ and the originator will accept it, the weak entrant decides to invest when

$$(1-\hat{D})E+[1-\mu(1-q)][(\hat{D}-\theta)(H-L)+C_O]-I \geq [1-\mu(1-q)][(\hat{D}-\theta)(H-L)+C_O]$$

This inequality yields $I \leq (1-\hat{D})E$. Note that the size of the reverse payment has no effect: given that both entrant's types ask for $\check{R}$ and the originator accepts it, the utility of investing lies only in the fact that it makes the entrant get the market profits $(1-\hat{D})E$ after the entry date. When this is the case, the weak entrant always invests, otherwise he never does. As in the main text, it can be shown that a policy allowing reverse payments with an entry date $\hat{D}$ such that $I \leq (1-\hat{D})E$ is always suboptimal, so we focus on $I > (1-\hat{D})E$. This means that the entry date must be smaller than $1-\frac{I}{E}$.

Consider therefore a policy $\rho = (R, \hat{D} > 1 - \frac{I}{E})$. Under such a policy, the entrant always asks for a reverse payment and never invests. Total welfare is

$$E[W|R] = \mu \underline{W} + (1-\mu)[\hat{D}\underline{W} + (1-\hat{D})\bar{W}] \quad (4.14)$$

When the entrant is weak (which occurs with probability $\mu$), given that he never invests, total welfare is just the monopoly welfare $\underline{W}$. When the entrant is strong, total welfare is just the monopoly welfare until $\hat{D}$ and the duopoly welfare after it.

[99] This condition comes from setting $\mu(1-q) = 1$, so that the second addend on the LHS is zero. When $\mu(1-q) < 1$, the LHS is even larger.

When reverse payments are banned ($\rho = N$), the weak entrant invests as long as

$$(1-\theta+\frac{C_O}{H-L})E - I \geq 0$$

When the entrant invests, he gets market profits $(1-\theta+\frac{C_O}{H-L})E$ after his entry and pays $I$. When he does not invest, he just stays out of the market and his profits are zero. Therefore, he invests when $I \leq (1-\theta+\frac{C_O}{H-L})E$. Assume that this is the case. Total welfare is

$$E[W|N] = (\theta - \frac{C_O}{H-L})\underline{W} + (1-\theta+\frac{C_O}{H-L})\bar{W} - \mu I \tag{4.15}$$

This is the sum of the monopoly welfare in the pre-entry period and the duopoly welfare in the post-entry period, minus the cost of investing for the weak entrants.

Allowing reverse payments is optimal when $E[W|R] \geq E[W|N]$, which yields

$$I \geq \check{I} = \frac{[\mu(1-\hat{D}) + \hat{D} - \theta - \frac{C_E}{E}](\bar{W} - \underline{W})}{\mu}$$

The policy implications are very similar to the main text: (i) for a sufficiently high investment cost, it is better to allow reverse payments, so that investment is reduced; (ii) the threshold for the investment cost is decreasing over $\mu$: the more difficult is the technology, the better it is to allow reverse payments. Policy implications are, therefore, robust also to changing the party that makes the take-it-or-leave-it offer.

### 4.4.4 Incentives to Invest

This Appendix discusses the policy implications when we consider the incentives to invest to be on the market - the focus of the second chapter of the thesis. All the profits in this Appendix are the *ex ante* profits that each party expects to earn.

Consider first the case where (1) the originator makes the take-it-or-leave-it offer.

Consider the case (1.a) $\mu \leq \mu^*$. When reverse payments are allowed, the originator always uses them and the entrant never invests. Therefore, the originator's profits are

$$\pi_O^R = \mu H + (1-\mu)[\hat{D}H + (1-\hat{D})L] - [(\hat{D}-\theta)E - C_E]$$

The first addend represents the possibility that the entrant is weak, in which case the originator gets the monopoly profits $H$; the second one the possibility that he is strong, in which case he gets the monopoly profits until $\hat{D}$ and the duopoly profits afterwards; the third one is the reverse payment. The entrant's profits are

$$\pi_E^R = (1-\mu)[(1-\hat{D})E] + [(\hat{D}-\theta)E - C_E]$$

The first addend represents the possibility that the entrant is strong, in which case he gets market profits $(1-\hat{D})E$; the second one is the reverse payment. When reverse payments are banned, the weak entrant always invests. The originator's profits are, therefore,

$$\pi_O^N = (\theta + \frac{C_E}{E})H + (1-\theta-\frac{C_E}{E})L$$

and the entrant's profits

$$\pi_E^N = (1-\theta-\frac{C_E}{E})E - \mu I$$

The originator's profits are higher when reverse payments are allowed as long as $(\hat{D}-\theta-\frac{C_E}{E})[H-L-E] > 0$, which is true because $\hat{D} \geq \theta+\frac{C_E}{E}$ (otherwise reverse payments would not be used) and $H > L+E$ by assumption (monopoly profits are higher than the sum of duopoly profits). The entrant's profits are higher when $I \geq (1-\hat{D})E$, which is true because allowing reverse payments with an entry date $\hat{D} \leq 1-\frac{I}{E}$ is never optimal (therefore the AA will never choose such an entry date). When $\mu \leq \mu^*$, therefore, both the originator and the entrant's profits increase. This means that we are in the same situation described in the second chapter: reverse payments increase the parties' profits[100], therefore they always increase the entrant's incentives to invest to be on the market. The originator's incentives can be higher or lower depending on how much entry is increased. We have the same tension for the originator: his incentives to invest increase *given the entrant's entry*, but may be reduced if the higher entrant's incentives increase entry sufficiently.

Consider now the case (1.b) $\mu > \mu^*$. When reverse payments are allowed, the originator uses them with a positive probability $p^{\rho*}$ (4.5) and the entrant invests with a probability $q^*$ (4.4). Therefore, the originator's profits are

$$\begin{aligned}\pi_O^R &= p^{\rho*}\{\mu(1-q^*)H + [1-\mu(1-q^*)][\hat{D}H + (1-\hat{D})L] - [(\hat{D}-\theta)E - C_E]\} + \\ &\quad +(1-p^{\rho*})\{\mu(1-q^*)H + [1-\mu(1-q^*)][(\theta+\frac{C_E}{E})H + (1-\theta-\frac{C_E}{E})L]\}\end{aligned}$$

Substituting for $p^{\rho*}$ and $q^*$, we get

$$\pi_O^R = H - (1-\theta)E + C_E$$

[100] In that chapter, the originator's profits increase because of his information rent (he receives a private signal over his patent strength), while in the main text of this chapter they increase because he is the party making the take-it-or-leave-it offer.

The originator's profits are the difference between the monopoly profits and the entrant's expected litigation payoff. The entrant's profits are

$$\pi_E^R = (1-\theta)E - C_E - \mu I$$

The first two addends represent the expected litigation payoff, while the third one is the expected investment cost. When reverse payments are banned, the weak entrant always invests, so the parties profits are the same as in case (1.a): the originator's profits are $\pi_O^N = (\theta + \frac{C_E}{E})H + (1-\theta-\frac{C_E}{E})L$ and the entrant's profits $\pi_E^N = (1-\theta)E - C_E - \mu I$. The originator's profits are higher when reverse payments are allowed when $(1-\theta-\frac{C_E}{E})[H-L-E] > 0$, which is always true because $1 \geq \theta + \frac{C_E}{E}$,[101] and $H-L-E>0$ by assumption. The entrant's profits are unchanged. When $\mu \leq \mu^*$, therefore, only the originator's profits increase, while the entrant's profits remain unchanged. That means that allowing reverse payments when (i) the originator has the bargaining power and (ii) the technology is difficult has the additional positive effect of fostering the originator's investment to be on the market. There is no tension in the originator's incentives here, because the entrant's profits do not increase. Keeping this into account strengthens the conclusion that allowing reverse payments can increase total welfare. The conclusion in the main text is, therefore, a conservative condition for reverse payments to enhance welfare.

Consider now the case (2) where the entrant makes the take-it-or-leave-it offer. We know, from the previous Appendix, that the entrant will always ask for a reverse payment independently from his type. So, when reverse payments are allowed, the entrant uses them and never invests. Therefore, the originator's profits are

$$\pi_O^R = \mu H + (1-\mu)[\hat{D}H + (1-\hat{D})L] - [1-\mu][(\hat{D}-\theta)(H-L) + C_O]$$

The first addend represents the possibility that the entrant is weak, in which case the originator gets the full monopoly profits $H$; the second one the possibility that he is strong, in which case he gets the monopoly profits until $\hat{D}$ and the duopoly profits afterwards; the third one the reverse payment when the entrant never invests ($q = 0$). The entrant's profits are

$$\pi_E^R = (1-\mu)[(1-\hat{D})E] + [1-\mu][(\hat{D}-\theta)(H-L) + C_O]$$

The first addend represents the possibility that the entrant is strong, in which case he gets market profits $(1-\hat{D})E$; the second one is the reverse payment. Profits

[101] If $\theta + \frac{C_E}{E} > 1$, then the entrant would simply prefer to stay out than litigating, so the problem would not be interesting.

are similar to case (1.a): the differences are that $q = 0$ and the size of the reverse payment is higher. When reverse payments are banned, the weak entrant always invests. The originator's profits are, therefore, $\pi_O^N = (\theta - \frac{C_O}{H-L})H + (1-\theta-\frac{C_O}{H-L})L$ and the entrant's profits $\pi_E^N = (1-\theta+\frac{C_O}{H-L})E - \mu I$. The originator's profits are higher when reverse payments are allowed when

$$\mu[(1-\theta)(H-L)+C_O] > 0$$

which is true unless $\mu = 0$. This leads to the following Proposition:

**Proposition 19** *The entrant's profits are higher when* $\mu < \tilde{\mu} = \frac{(H-L-E)[(\hat{D}-\theta)(H-L)+C_O]}{(H-L)[(\hat{D}-\theta)(H-L)+C_O+(1-\hat{D})E-I]}$.[102]

For the entrant's profits to increase, the technology must be sufficiently easy: a hard technology reduces the originator's willingness to pay a reverse payment, because he knows that the entrant has no incentive to invest and, therefore, is probably non-competitive. When the difficulty is higher than $\tilde{\mu}$, then the entrant prefers reverse payments to be banned - even if he prefers to ask for a reverse payment as soon as he has invested to enter the market.

When the entrant makes the offer, therefore, allowing reverse payments always increases the originator's profits and increase the entrant's ones when the technology is sufficiently easy. This means that when the technology is difficult, the originator's incentives to invest always increase, both because his profits increase in case of generic entry and because generic entry occurs less often. When the technology is difficult, we have the well-known tension on the originator's side and higher incentives to invest for the entrant.

In the intermediate case (3) where the originator chooses whether to offer a reverse payment or not and both parties get a positive share of the surplus, reverse payments increase both parties' profits (like in case 1.a), both when $\mu \leq \hat{\mu}^*$ and when $\mu > \hat{\mu}^*$. The entrant's incentives to invest increase and the originator's ones show the tension explained above.

[102] *Ex ante*, when the entrant has not learnt yet whether he is weak or strong, he can be hurt by reverse payments. A high probability $\mu$ that he is weak reduces the originator's willingness to pay (at the limit, when $\mu = 1$, the originator's willingness to pay is zero). For the entrant, when he is on the market and has learnt his type, it is always convenient to ask for a reverse payments, but *ex ante* this reduces his profits when the technology is sufficiently difficult.

# Conclusion

This section links the results of essays about the pharmaceutical industry. The second and the third chapters show that, when reverse payments are used, they actually delay entry, but the very possibility of using them increases the competition on the market. This effect is neglected in the FTC and the Commission's decisions of banning reverse payments *per se*.

Considering the results of these essays together, we can also draw some conclusions over the patent strength and the incentives to invest. A stronger patent usually calls for a more lenient policy. The reason is that, both in the third and fourth essay, a stronger patent implies a lower cost from allowing reverse payments - entry would be late anyway. On the investment side, we have the tension explained in the second essay on the originator's investment when the technology is easy (while the entrant always has higher incentives to invest), but the tension fades away when the technology is difficult, because the entrant's profits do not increase. Only the originator's profits increase, which increases his investment to be on the market. When the technology is difficult, therefore, the optimal policy described in the fourth chapter is only a conservative estimate of the "real" optimal one - the one that considers both (i) the parties' investment to be on the market and (ii) the productive investment of the entrant. Under both types of technology, when reverse payments are allowed the entrant has an incentive to invest *more* in order to *be* on the market, while the originator has the tension described in the second chapter when the technology is easy and always has an incentive to invest more when it is difficult. This is a further reason why a lenient policy towards reverse payments, when the technology is difficult, can improve welfare.

A possible *caveat* lies, however, in the investment to be competitive: a lenient policy makes the entrant invest *less* to *be competitive*. These two results mean that a lenient policy possibly increases potential entry (the investments to be able to produce), but can reduce the competitive capacity of the industry (the investments to be able to produce efficiently). A potentially dangerous effect of a lenient policy is incentivising "not useful" investments, i.e. investments whose objective is to show (to the originator) one's own existence on the market, but not one's own ability to actually compete.

# References


[1] Abrantes R., Froeb L., Geweke J. and Taylor C. (2006), “A Variance Screen for Collusion”, International Journal of Industrial Organization, Vol. 24, Issue 3, May, 467-486.

[2] Arve M. (2012), "Dynamic Procurement, Costly Bidding and Bankruptcy Risk", mimeo.

[3] Bebchuk L.A (1984), "Litigation and Settlement under Imperfect Information", *RAND Journal of Economics* 15, 404-415.

[4] Besanko, D., Spulber, D. F., (1989), “Antitrust Enforcement under Asymmetric Information”, Economic Journal, Royal Economic Society, vol. 99(396), pages 408-25, June.

[5] Bessen J.E. and Meurer M.J., 2006. "Patent Litigation with Endogenous Disputes," American Economic Review, American Economic Association, vol. 96(2), pages 77-81, May.

[6] Binmore K.G. and Herrero M.J. (1988a), "Matching and Bargaining in Dynamic Markets", *Review of Economic Studies* 55, 17-31.

[7] Binmore K.G. and Herrero M.J. (1988b), "Security Equilibrium", *Review of Economic Studies* 55, 33-48.

[8] Blinder A.S., Canetti E.R.D., Lebow D.E. and Rudd J.B. (1998), “Asking about Prices: a New Approach to Understanding Price Stickiness”, New York, Russell.

[9] Bowman D., Minehart D., Rabin M., (1999), “RDPs in a Consumption Savings Model”, Journal of Economic Behavior & Organization, Vol. 38 (1999) 155-178.

[10] Chen J. and Harrington J. Jr. (2006), “Cartel Pricing Dynamics with Cost Variability and Endogenous Buyes Detection”, International Journal of industrial Organization, 24, 1185-1212.

[11] Connor J.M. (1998), “What Can We Learn from the ADM Global Price Conspiracies?”, Staff Paper.

[12] Connor J.M. (2001), “Our Customers Are Our Enemies: the Lysine cartel of 1992-1995”, Review of Industrial Organization 18, 5-21

[13] Daughety A.F. and Reinganum J.F. (1994), "Settlement Negotiations with Two-Sided Asymmetric Information: Model Duality, Information Distribution and Efficiency", *Game Theory and Information*, 9403009, EconWPA.

[14] Ellison G. (2006), "Bounded Rationality in Industrial Organization", in Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, ed. Richard Blundell, Whitney Newey and Torsten Persson. Cambridge: Cambridge University Press.

[15] Erickson G.M. and Johansson J. K. (1985), "The Role of Price in Multi-Attribute Product Evaluations", Journal of Consumer Research, 12(2): 195-199.

[16] Fox C.R., Poldrack R.A., Tom S.M. and Trepel C., (2007), "The Neural Basis of RDPs in Decision-Making Under Risk", Science, Vol. 315 26 January.

[17] Gale D. (1986a), "Bargaining and Competition Part I: Characterization", *Econometrica* 54, 785-806.

[18] Gale D. (1986b), "Bargaining and Competition Part II: Existence", *Econometrica* 54, 807-818.

[19] Gale D. (1987), "Limit Theorems for Markets with Sequential Bargaining", *Journal of Economic Theory*, 43, 20-54.

[20] Genesove D. and Mayer C. (2001), "RDPs and Seller Behavior: Evidence from the Housing Market", The Quarterly Journal of Economics, 116 (4): 1233-1260.

[21] Gill D. and Prowse V. (2012), "A Structural Analysis of Disappointment Aversion in a Real Effort Competition", American Economic Review, 102(1): 469-503.

[22] Gill D. and Stone R. (2010), "Fairness and Desert in Tournaments", Games and Economic Behavior, 69, 346-364.

[23] Gratz L. (2012), "Economic Analysis of Pay-for-delay Settlements and Their Legal Ruling", Munich Discussion Paper No. 2012-6.

[24] Green E.J. and Porter R.H. (1984), "Noncooperative Collusion under Imperfect Information", Econometrica, Vol. 52, No. 1. Jan., 87-100.

[25] Haigh M.S., List J.A. (2006), "Do Professional Traders Exhibit Myopic RDPs? An Experimental Analysis", The Journal of Finance, Vol. 60, No. 1, February.

[26] Hall B. (2002), "The Financing of Research and Development", Oxford Review of Economic Policy, Vol. 18, no.1, pp.35-51.

[27] Haltiwanger J. and Harrington J. Jr. (1991), “The Impact of Cyclical Demand Movement on Collusive Behavior”, RAND Journal of Economics, 22, 89-106.

[28] Harrington J. Jr. (2004), “Cartel Pricing Dynamics in the Presence of an Antitrust Authority”, RAND Journal of Economics, 35, 651-673.

[29] Harrington J. Jr. (2005), “Optimal Cartel Pricing in the Presence of an Antitrust Authority”, International Journal of Industrial Organization, 46, 145-169.

[30] Heidhues P. and Koszegi (2008), “Competition and Price Variation When Consumers Are Loss Averse”. American Economic Review, 98(4): 1245-1268.

[31] Kahneman, Daniel; Knetsch, Jack L. and Thaler, Richard H. (1991), “The Endowment Effect, RDPs, and Status Quo Bias: Anomalies.” Journal of Economic Perspectives, Winter 5(1): 193-206.

[32] Kahneman, Daniel and Tversky, Amos (1979). “Prospect Theory: An Analysis of Decision under Risk.” Econometrica, March, 47(2): 263-91.

[33] Kalwani M. U. and Chi Kin Yim (1992), “Consumer Price and Promotion Expectations: An Experimental Study”, Journal of Marketing Research, 29(1): 90-100.

[34] Kandori M. (1991), “Correlated Demand Shocks and Price Wars During Booms”, Review of Economic Studies, 58, 171-180.

[35] Koszegi B. and Rabin M. (2006), “A Model of Reference-Dependent Preferences”. Quarterly Journal of Economics, 121(4): 1133-1165.

[36] Koszegi B. and Rabin M. (2007), “Reference-Dependent Risk Attitudes”. American Economic Review, 97(4): 1047-1073.

[37] Lemley M.A. and Shapiro C. (2005), "Probabilistic Patents", *Journal of Economic Perspectives,* Vol. 19, Number 2 — Spring 2005—Pages 75–98

[38] Levenstein M. and Suslow V. Y. (2001), “What Determines Cartel Success?” Journal of Economic Literature, Vol. 44, No. 1, March, 43-95(53).

[39] Levenstein M. (1997), “Price Wars and the Stability of Collusion: A Study of the Pre-World War I Bromine Industry?” The Journal of Industrial Economics, Vol. 45, No. 2, Jun., 117-137.

[40] Macera R., (2011), “Intertemporal Incentives with Expectation-Based Reference-Dependent Preferences”, mimeo.

[41] Novemsky N. and Kahneman D. (2005), “The Boundaries of RDPs”, Journal of Marketing Research, Vol. 17, May, 119-128.

[42] Pisano G. (2006), "Science Business: the Promise, the Reality and the Future of Biotech", Harvard Business School press, Boston.

[43] P’Ng I.P.L. (1983), "Strategic Behavior in Suit, Settlement and Trial", *Bell Journal of Economics*, 14, 539-550.

[44] Reinganum J.F. and Wilde L.L. (1986), "Settlement, Litigation and the Allocation of Litigation Costs", *RAND Journal of Economics*, 17, 557-566.

[45] Rotemberg J. (2004), “Fair Pricing”, National Bureau of Economic Research Working Paper 10915.

[46] Rotemberg J. and Saloner G. (1986), “A Supergame Theoretic Model of Price Wars during Booms”, American Economic Review, 76, 390-407.

[47] Rubinstein A. and Wolinsky A. (1985), "Equilibrium in a Market with Sequential Bargaining", *Econometrica* 53, 1133-1150.

[48] Rubinstein A. and Wolinsky A. (1990), "Decentralized Trading, Strategic Behavior and the Walrasian Outcome", *Review of Economic Studies* 57, 63-78.

[49] Salant S.W. (1984), "Litigation of Settlement Demands Questioned by Bayesian Defendants", working paper 516, California Institute of Technology, Div. of the Humanities and Social Sciences.

[50] Salant S.W. and Rest G. (1982), "Litigation of Questioned Settlement Claims: A Bayesian Nash-Equilibrium Approach", Rand Corporation Discussion Paper p-6809.

[51] Salop S. (1979), “Monopolistic Competition with Outside Goods”, The Bell Journal of Economics, Vol. 10, No. 1, Spring, 141-156.

[52] Schweizer Urs (1989), "Litigation and Settlement under Two-Sided Incomplete Information", *The Review of Economic Studies*, Vol. 56, No. 2, pp. 163-177.

[53] Shapiro, C. (2003) " Antitrust Limits to Patent Settlements," *RAND Journal of Economics*, The RAND Corporation, Vol. 34(2), pages 391-411, Summer.

[54] Thaler R., (1980), “Toward a Positive Theory of Consumer Choice”, Journal of Economic Behavior and Organization, Vol. 1, Issue 1, 39-60.

[55] Tversky A. and Kahneman D., “RDPs in Riskless Choice: A Reference-Dependent Model”, The Quarterly Journal of Economics, Vol. 106, No. 4. (Nov., 1991), pp. 1039-1061.

[56] Willig R. and Bigelow J.P. (2004), "Antitrust Policy towards Agreements that settle Patent Litigation", *The Antitrust Bulletin*, Fall.

[57] Winer R. S. (1986), “A Reference Price Model of Brand Choice for Frequently Purchased Products”, Journal of Consumer Research, 13(2):250-256.