Delay Competition to Increase Competition: Should Reverse Payments be Banned per se?*

Anton-Giulio Manganelli†
RBB Economics
December 8, 2014

Abstract

In the pharmaceutical industry, a reverse payment is a payment from an originator to a generic entrant in exchange of a delay in his entry. Reverse payments make the monopoly period longer and increase industry profits. In some recent cases, the US and the EU Antitrust Authorities have banned these agreements per se. This paper analyzes their dynamic effects and shows that this should not be the case under two scenarios: (i) when the parties’ investment decisions are taken into account and the information over the patent strength is asymmetric; (ii) when the entrant may go bankrupt. In case (i), reverse payments increase the entrant’s investment. This, coupled with the asymmetric information over the patent strength, increases the number of litigations. A fraction of them ends up in generic entry, which increases consumer surplus. Reverse payments also create a tension in the originator’s investment, absent from the traditional patent literature. In case (ii), banning reverse payments pushes the entrants with liquidity problems out of the market. This reduces consumer surplus through the reduction of the number of competitors. Banning reverse payments is worse when the economy is in a downturn and tacit collusion is sustainable among few players.

JEL Classification: K21, L12, L41.

Keywords: reverse payments, pay-for-delay, cartel, pharmaceutical industry, litigation, settlement, asymmetric information.

*I thank Klenio Barbosa, Roberta Dessì, Bruno Jullien, Gerard Llobet, Adrian Majumdar, Xavier Martinez-Giralt, Patrick Rey, the CRESSE (2013) and the EPIP (2013) participants for their useful comments and discussions. I am deeply indebted, in particular, to Yassine Lefouili for his invaluable help. The views expressed in this paper are those of the author and do not necessarily reflect the position of RBB Economics.

†Associate at RBB Economics. E-mail: ag.manganelli@gmail.com. This paper was developed during my PhD studies at Toulouse School of Economics.


1 Introduction

In the pharmaceutical industry, patents are the main assets of firms. Originators have the right to enforce them by litigating against potential infringers, in order to prevent them from exploiting their own inventive activities. Litigations, however, are costly and their outcome is uncertain. As an alternative to litigation, the parties can settle the (potential) dispute. Patent settlements, however, could cover cartel-like agreements: a firm with a weak patent may agree to share its monopoly profits with a rival and cover it behind a settlement agreement. The simplest way to share monopoly profits is a payment from the originator to the entrant in exchange of a delay in his entry, i.e. a reverse payment (also called "pay-for-delay"). This explains why Antitrust Authorities scrutinize this type of settlements carefully.

In 2003, in the Bristol-Myers\textsuperscript{1} case, the Cardizem\textsuperscript{2} case and the Valley Drug-Geneva Pharmaceuticals\textsuperscript{3} case, and in 2006 in the Tamoxifen\textsuperscript{4} case, the incumbent paid a potential generic competitor to avoid litigating over the patent and to stay out of the market until patent expiry. The FTC found these agreements anticompetitive. In 2008, the European Commission launched an inquiry into the pharmaceutical sector with dawn raids at the premises of several originator and generic companies, with particular attention to settlement agreements involving reverse payments. In June 2013, in the Lundbeck\textsuperscript{5} case, the European Commission fined Lundbeck and other companies for delaying generic entry through the use of reverse payments. In July 2014 the same happened with Servier and some generic firms.\textsuperscript{6} In Canada, the Competition Bureau stated that they will use criminal sanctions against reverse payments agreements.\textsuperscript{7} At the appeal level, however, in the US some of these decisions have been overturned. In the appeal of 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 illegal, as they did not extend beyond the original patent terms. This makes reverse payments legal on the basis that patents, as long as they are considered valid, give the full right of exclusion. In 2013, the Supreme Court of the US, in the Actavis case, affirmed that reverse payments should be treated under a rule of reason and that agreements over valid patents could be challenged by the FTC.\textsuperscript{8}

An economic approach towards patents suggests that they should be considered as probabilistic property rights (Shapiro 2003, Lemley and Shapiro 2005), as a patent may later be found invalid - or the entrant’s product may be found not to infringe it. In other words, the patent-holder should not have the right to exclude a party from using its own patent,

\begin{footnotes}
\item[3]Valley Drug Co. v Geneva Pharmaceuticals, Inc, 344 F.3d 1294 (11th Cir. 2003).
\end{footnotes}
but only to try to exclude it. This makes entry possible, if parties litigate. Shapiro (2003) suggests to allow settlements that leave consumers, in expectations, at least with the same surplus 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 payments 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, results crucially hing on some caveats. 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 paper is to contribute to the literature on reverse payments by analyzing their dynamic effects. In particular, this paper considers two independent cases. In the first one, it considers the impact that the Antitrust policy on reverse payments has on the parties’ incentives to invest, on the litigation rate and, finally, on consumer surplus. In the second one, it focuses on the possible liquidity problems of generic pharmaceutical firms. These two cases show that reverse payments, even if they typically make the originator’s monopoly period longer, should not be considered anticompetitive *per se*.

In the first case, reverse payments delay the generic manufacturer’s entry but also increases his incentives to invest. The higher industry profits due to reverse payments always increase the entrant’s profits and, at the same time, create a tension in the originator’s incentives to invest. This always increases generic entry and, when there is asymmetric information over the patent strength, also the litigation rate. A fraction of these additional litigations eventually ends up in generic entry, which increases consumer surplus. In other words, reverse payments create a trade-off between *delaying generic entry* and *inducing more generic entry*. The impact of reverse payments on consumer surplus is, therefore, not trivial and deserves a careful analysis. For several parameter sets, the positive effect of inducing entry offsets the negative effect of delaying it. This suggests that a rule of reason is more suited than a ban *per se*. The optimal policy is derived. Moreover, the originator’s incentives to invest may increase or decrease. They may decrease because the higher industry profits can make the entrant enter, which reduces the originator’s profits, but they may also increase because the higher industry profits make the entrant, provided that he has invested, keener on settling.

---

9 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.
on more favorable terms to the originator in order to avoid litigation.

The second case draws evidence from Hall (2002), who shows that small and new pharmaceutical firms, as generic producers often are, experience high costs of capital which can create and worsen liquidity problems, and Pisano (2006), who shows that new biotechnology firms are financially constrained and that some drug development failures can lead to bankruptcy. The analysis, in this case, is on the relationship among the Antitrust policy, reverse payments and consumer surplus when the generic producer may have liquidity problems. Banning reverse payments reduces the entrant’s expected profits, making weak entrants go bankrupt. This reduces consumer surplus both before and after patent expiry through the reduction of the number of competitors. For some parameter sets, this effect offsets the positive effect of banning reverse payments, which consists of earlier entry, provided that the entrant does not have liquidity problems. The negative effect of banning reverse payments is stronger when the economy is in a downturn and few players are able to tacitly collude. This suggests that generic manufacturers could ask for a "failing firm" defence. The EC recently opened the door to this type of defence, though in a case unrelated with the pharmaceutical industry.

This paper uses the analytical framework of the literature of litigations and settlements. Litigations and settlements have 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 of 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, while, 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 - which makes the equilibrium 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, this paper sticks to a one-sided incomplete information game, where one party makes a take-it-or-leave-it offer. Results, as will become clear, hold also for more general bargaining rules.

---

10 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.

2 Case I - The model

This section describes the first case, i.e. the one about the parties' investment decisions and litigation rate. 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.\footnote{Date 0 is the date when the entrant is ready to enter, which is the same as the one when the parties decide whether to litigate or to settle. This will be shown in Appendix 5.2.}

In the first stage, the AA decides the maximal allowed amount of reverse payment $\hat{R}$. A maximal reverse payment of 0 means that they are banned and one of $\infty$ means that no cap is set. In the second stage, the originator can 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, if both the originator and the generic manufacturer have invested, the entrant makes a take-it-or-leave-it settlement offer.\footnote{The fact that the entrant makes a take-it-or-leave-it offer (or, better, a take-it-or-leave-it request) is not necessary for the results. Any form of bargaining that leaves the entrant with some additional surplus from the settlement compared to his threat point (the litigation payoff) yields our qualitative results. 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.}

The offer consists of an entry date $0 \leq D \leq 1$ and a payment $R \leq \hat{R}$ from the originator to the entrant. In the fifth stage, the originator learns the true probability $\theta \in \{\underline{\theta}, \bar{\theta}\}$ of winning the litigation, where $0 < \underline{\theta} < \bar{\theta} < 1$. Call it "patent strength".\footnote{The patent strength represents both the probability that the patent is held valid and infringed - see the next footnote.}

The probability of drawing $\underline{\theta}$ is $\lambda$. This signal represents the information that arrives from the national patent office or from experts asked to evaluate the patent strength prior to the potential litigation. We assume that the originator has better information about the patent strength because he is the party that filed the patent application and, therefore, has a better knowledge of the possible problems related to his patent.\footnote{The set-up with the originator receiving the private signal suits to the case where the entrant challenges the validity of the patent - so it is reasonable to assume that the originator has better information. However, results are qualitatively the same when the entrant has better information - which suits to the case where the entrant seeks to invent around the patent, provided that the identity of the party making the take-it-or-leave-it offer is inverted.}

In the sixth stage, if the originator accepts the offer $D$ and $R$ are enforced, otherwise the 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 chooses a cap $\hat{R}$.

2. **Originator's investment.** The originator invests $I_O$ to enter the market or stays out.

3. **Entrant's investment.** The entrant invests $I_E$ to enter the market or stays out.
4. **Entrant’s offer.** If the entrant has invested, the entrant makes a settlement offer.

5. **Originator’s signal.** The originator receives the private signal $\theta \in \{\underline{\theta}, \bar{\theta}\}$.\(^{16}\)

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

In case of litigation, the originator and the entrant bear, respectively, litigation costs $C_O$ and $C_E$.\(^{17}\) Define $H$ the originator’s profits if he is the monopolist 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 the entrant if entry occurs immediately. Assume that $H > L + E$: 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. By backward induction, the originator first computes whether the entrant will enter or not; if he will, both players form expectations over the outcome in the litigation-settlement subgame.

When reverse payments are allowed, it can be easily shown that industry profits are higher (intuitively, because they are used only if they delay entry, so that the pledgeable profits become higher - see Lemma 1). Therefore (i) the entrant always has more incentives to invest, (ii) the originator may have more or less incentives to invest (more if the entrant would have entered anyway and the higher profits make him offer more favorable settlement terms; less if they induce a generic manufacturer that would have otherwise stayed out to enter) and (iii) $CS$ is lower for a given investment level. We will show that there exist several parameter sets where $CS$ increases under $^R > 0$, even when no cap is set ($^R = \infty$), thanks to the pro-investment effects.

The following subsection analyzes the last stage of the game and computes the litigation and settlement profits.

### 2.1 Litigation-Settlement stage

The model is solved backwards. Denote $\hat{\theta}$ the (hypothetical) realization for which the originator would be indifferent between litigating and accepting the entrant’s offer and $E_o[\theta|\theta > \bar{\theta}]$ the entrant’s Bayesian updating of $\theta$ given that the originator refuses a proposal based on $\bar{\theta}$.

If the parties litigate, they expect to obtain:

\(^{16}\)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.

\(^{17}\)They can be seen as the incremental legal costs of litigation (i.e. those in excess of any legal costs associated with settlement).
Originator: \( \theta H + (1 - \theta)L - C_O \)
Entrant: \( (1 - E_\theta(\theta))E - C_E \) before the settlement offer;
\( (1 - E_\theta(\theta|\theta > \bar{\theta}))E - C_E \) 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 \( (1 - E_\theta(\theta|\theta > \bar{\theta})) \). His litigation costs are \( C_E \).

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 at least as profitable as litigating, i.e. if \( DH + (1 - D)L - R \geq \theta H + (1 - \theta)L - C_O \), which yields:

\[
D \geq D^* = \theta + \frac{R - C_O}{H - L}. \tag{1}
\]

The minimal entry date that the originator is willing to accept is increasing in \( R \) and \( \theta \) and decreasing in \( C_O \) and \( (H - L) \) (as long as \( R \) is larger than \( C_O \)). A higher patent strength makes the originator more confident of winning the litigation and, therefore, less willing to accept an early entry. A higher \( R \) means that accepting the settlement is more costly, which also makes the originator less keen on settling. However, a settlement makes the originator save the litigation costs \( C_O \). The net cost of the settlement \( R - C_O \) is weighted over the gain of making monopoly profits longer, \( (H - L) \).

When reverse payments are allowed, the following Lemma shows that the only problem is the choice of the reverse payment.

**Lemma 1** If the entrant has invested, he asks for the maximal possible payment \( \hat{R} \) and enters at \( D^* \) if \( D^* \) is not greater than 1, and asks for \( R = (1 - \theta)(H - L) + C_O \) and enters at the patent expiry otherwise.

**Proof.** See Appendix 5.1 for the first part of the Lemma. The intuition is that a larger reverse payment more than compensates the profit loss due to the later entry needed to keep

---

The fact that each party bears his own litigation costs is the so called American rule. Results are robust to changes in the allocation of litigation costs.
the originator willing to settle. A higher reverse payment implies a later entry date - see (1): a marginally higher $R$ makes the entrant gain $dR$ through $R$ and lose $\frac{E}{H-L}dR$ through $D$. Being the loss in the originator’s profits $(H - L)$ higher than the entrant’s profits $E$, the optimal $R$ is the maximal possible one: $R = \hat{R}$. The second part comes from the fact that that the parties cannot agree on an entry date after patent expiry $D > 1$. Therefore, $D = 1$ is an implicit constraint on the maximal reverse payment: the entrant cannot ask for a reverse payment so high that the originator would need a monopoly period even longer than the patent duration to accept the settlement. Any policy $\hat{R} > (1 - \theta)(H - L) + C_O$, therefore, is equivalent to $\hat{R} = (1 - \theta)(H - L) + C_O$. For the sake of exposition, we assume in the following that $\hat{R} \leq (1 - \theta)(H - L) + C_O$.

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

**Lemma 2** *The choice of the maximal reverse payment $\hat{R}$ is a perfect substitute for the choice of the latest entry date $\hat{D}$.*

**Proof.** See (1). There is a biunivocal correspondence between $R$ and $D^*$, so setting a cap on $R$ or on $D$ is equivalent.

We can therefore restrict our attention to a cap on $R$. Importantly, this Lemma implies that the model can be reinterpreted in terms of latest entry date, instead of maximal reverse payment.

### 2.2 Optimal entrant’s offer

In principle, the entrant could behave as the principal in a principal-agent relationship with the originator, i.e. he could write a menu of contracts and make the originator truthfully reveal his type. However, this is impossible, as the originator’s settlement profits do not depend on his type (see Appendix 5.5 for a more detailed discussion). Therefore, the contract for the high type $\bar{\theta}$ must yield the same profits for the originator as the one for the low type $\underline{\theta}$, in order to make the originator not lie about his own type. The entrant can decide to make the participation constraint of the high type binding, in which case 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 it is not possible to elicit the originator’s type and that $\theta$ has 2 possible realizations, the entrant (provided that he has invested) has two potential optimal strategies: one that leaves the weak originator (i.e. when he draws $\theta = \underline{\theta}$) indifferent between litigating and settling and one that leaves the strong one (i.e. when he draws $\theta = \bar{\theta}$) indifferent. Call the realization $\tilde{\theta} \in \{\underline{\theta}, \bar{\theta}\}$ that leaves the originator indifferent between litigating and settling the realization "targeted" by the entrant. Denote $\hat{\theta}$ the equilibrium targeted realization under the policy $\hat{R}$. From Lemma 1, we know that we can substitute $R$ with $\hat{R}$. Upon entry, the entrant will optimally propose a reverse payment $\hat{R}$ and an entry date $D(\hat{\theta})$ such that the
originator is indifferent between accepting the settlement and litigating when the realized \( \theta \) is \( \hat{\theta} \).

Denote \( \pi^R_i(\hat{\theta}) \) the expected profits of party \( i \), where \( i = E \) is the entrant and \( i = O \) is the originator, under policy \( \hat{R} \) when the entrant targets the realization \( \hat{\theta} \). In other words, \( \pi^R_i(\hat{\theta}) \) represents the expected profits after stage 4 (the entrant makes the offer) and before stage 5 (the originator does not know his private signal yet). Under policy \( \hat{R} \), in equilibrium the originator will accept the request \( \{ R = \hat{R}, D = D(\hat{\theta}) = \hat{\theta} + \frac{\hat{R} - CO}{H-L} \} \) if and only if the realized patent strength is not larger than the targeted one: \( \theta \leq \hat{\theta} \) - see (1).

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

\[
\pi^E_i(\bar{\theta}) = \lambda[(1 - \theta - \frac{\hat{R} - CO}{H-L})E + \hat{R}] + (1 - \lambda)[(1 - \bar{\theta})E - CE].
\]  

(2)

(2)

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

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

\[
\pi^E(\bar{\theta}) = (1 - \bar{\theta} - \frac{\hat{R} - CO}{H-L})E + \hat{R}. 
\]  

(3)

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 exchange for the certainty of avoiding litigation.

The following Lemma describes the optimal entrant’s offer, given the AA’s policy.

**Lemma 3** The entrant targets \( \hat{\theta} = \bar{\theta} \) under policy \( \hat{R} \) if and only if

\[
\lambda > \frac{\hat{R} + CE - \frac{\hat{R} - CO}{H-L}E}{\hat{R} + CE - \frac{\hat{R} - CO}{H-L}E + (\bar{\theta} - \bar{\theta})E},
\]

\[
\frac{d\lambda^{\hat{R}}}{d\hat{R}} = \frac{(1 - \frac{E}{\hat{R} + CE - \frac{\hat{R} - CO}{H-L}E + (\bar{\theta} - \bar{\theta})E})}{(1 - \frac{E}{\hat{R} + CE - \frac{\hat{R} - CO}{H-L}E + (\bar{\theta} - \bar{\theta})E})^2} > 0.
\]

This important lemma implies that the AA can affect the realization targeted by the entrant, as \( \lambda^{\hat{R}} \) is strictly increasing in \( \hat{R} \). This means that an entrant that targets \( \bar{\theta} \) switches target to \( \bar{\theta} \) when the allowed reverse payment is sufficiently larger - in other words, the entrant offers a better offer to the originator, in order to avoid litigation. 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.
Consider now consumer surplus (CS). If a settlement takes place, CS is simply the monopoly CS until the generic manufacturer’s entry date \( D \) and the duopoly CS from that moment until patent expiry \( 1 - D \). Denoting \( \bar{S} \) the monopoly CS and \( \bar{S} > S \) the duopoly CS, we have \( CS = D\bar{S} + (1 - D)\bar{S} \). If litigation occurs, we 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 \( CS_L = \theta\bar{S} + (1 - \theta)\bar{S} \).

Disregarding the investment decisions, it is easy to see that CS is higher when reverse payments are banned \( (\hat{R} = 0) \). A "laissez faire" policy \( (\hat{R} = \infty) \) makes firms choose \( D = 1 \), while banning them makes the entrant propose \( D(\hat{\theta}) = \hat{\theta} - \frac{cO}{(H-L)} \), which is smaller than 1. Being \( \bar{S} > S \), 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 EC’s opinion that reverse payments should be banned per se. However, when considering both parties’ incentives to invest, banning reverse payments can reduce CS. A ban on reverse payments, indeed, always reduces the entrant’s incentives to invest and creates a tension in the originator’s ones. The originator’s incentives can be reduced because the smaller industry profits can make the entrant more aggressive in his settlement proposal, while the entrant’s incentives are always reduced because of the smaller industry profits. When the originator does not invest the CS falls to zero, while when the entrant does not invest there is no entry before the patent expiry nor litigation, which keeps CS at the monopoly level. For several parameter sets, as will be shown, the entry-enhancing effects of reverse payments dominate the entry-delaying ones. The following subsection describes all the subgame perfect equilibria of the game.

### 2.3 Subgame perfect equilibria

The possible outcomes of the game depend on (i) the originator’s investment, (ii) the entrant’s investment and (iii) the entrant’s settlement offer.

1) If \( I_O > H \), the originator does not invest and the game ends. Consumer surplus is \( CS(O_{out}) = 0 \). This case is trivial: if monopoly profits are smaller than the investment cost, the originator stays out of the market.

2) If \( \pi_O(\hat{\theta}) \leq I_O \leq H \), then the originator invests if and only if the entrant does not enter, i.e. when \( I_E > \pi^R_E(\hat{\theta}) \). In this case consumer surplus is \( CS(E_{out}) = \bar{S} \), the monopoly outcome. If \( I_E \leq \pi^R_E(\hat{\theta}) \), 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 \( \pi_O(\hat{\theta}) \leq I_O \leq \pi_O(\hat{\theta}) \), then the originator invests if the entrant does not enter or if he enters and targets \( \hat{\theta} \). The entrant enters if and only if \( I_E \leq \pi^R_E(\hat{\theta}) \), where \( \hat{\theta} \) is the targeted patent strength when the policy is \( \hat{R} \). If \( I_E > \pi^R_E(\hat{\theta}) \), the entrant stays out and consumer
surplus is \( CS(E_{\text{out}}) = S \). If the entrant enters, he targets \( \theta \) if \( \lambda > \lambda^R \) and \( \tilde{\theta} \) if \( 0 < \lambda \leq \lambda^R \).

Recall that the targeted realization depends on \( \hat{R} \), as \( \hat{R} \) has a positive impact on \( \lambda^R \). For any policy, if the entrant would enter and target \( \theta \), consumer surplus is \( CS(O_{\text{out}}) = 0 \), as the originator would not invest. On the other hand, if the entrant enters and targets \( \tilde{\theta} \), then

\[
CS^R(\tilde{\theta}) = (\tilde{\theta} + \frac{\hat{R} - C_0}{H - L}) S + (1 - \tilde{\theta} - \frac{\hat{R} - C_0}{H - L}) \tilde{S}.
\]

Note that the policy choice has an impact not only on CS, but also on the entrant’s profits and therefore on the originator’s incentives to invest.

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

If the entrant targets \( \theta \), consumer surplus is

\[
CS^R(\theta) = \lambda[(\theta + \frac{\hat{R} - C_0}{H - L}) S + (1 - \theta - \frac{\hat{R} - C_0}{H - L}) \tilde{S}] + (1 - \lambda)[\tilde{\theta} S + (1 - \tilde{\theta}) \tilde{S}]
\]

The originator draws \( \theta \) with probability \( \lambda \), in which case he settles, and \( \tilde{\theta} \) 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, \( \tilde{\theta} \).

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

**Proposition 1** 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. We show these two cases in detail. Recall that the originator’s profits depend only on the realization \( \theta \) targeted by the entrant: we have \( \pi_O(\theta) \equiv \pi^0_O(\theta) = \pi^\infty_O(\theta) < \pi_O(\tilde{\theta}) \equiv \pi^0_O(\tilde{\theta}) = \pi^\infty_O(\tilde{\theta}) \).

**Case 1: more generic entry.** Consider now \( \lambda^0 < \lambda^\infty < \lambda \). This means that the probability that the patent is weak is so high that the entrant targets the low realization even when no cap is set on reverse payments. Consider now the investment costs \( I_E \) and \( I_O \). If \( I_O > \pi^R_O(\tilde{\theta}) \), the originator never invests and the game ends. If \( I_O \leq \pi^R_O(\tilde{\theta}) \), we have three cases:

If (I) \( 0 \leq I_E \leq \pi^0_E(\theta) \), then the entrant invests under both policies. This makes CS higher under \( \hat{R} = 0 \), because this makes entry occur as soon as possible - see (1).

\(^{19}\)i.e. his settlement proposal is \( R = \hat{R} \) and \( D = \theta + \frac{\hat{R} - C_0}{H - L} \) if \( \lambda \geq \lambda^R \) and \( D = \tilde{\theta} + \frac{\hat{R} - C_0}{H - L} \) otherwise.

\(^{20}\)The equalities \( \pi^0_O(\theta) = \pi^\infty_O(\theta) \) and \( \pi^0_O(\tilde{\theta}) = \pi^\infty_O(\tilde{\theta}) \) are due to the fact that the entrant makes a take-it-or-leave-it offer. A more general bargaining rule without this feature does not change the qualitative results - see Section 4.
If (II) \( \pi^E_0(\theta) < I_E \leq \pi^E_{\infty}(\theta) \), the entrant invests if and only if the allowed reverse payment is sufficiently high. CS is therefore higher under such a cap, because it makes the entrant enter the market before patent expiry when the entry date associated with \( \hat{R} \) is smaller than 1 and because, when the originator draws \( \theta = \tilde{\theta} \), litigation occurs. Note in a more general framework, where the possible realizations of \( \theta \) are more than two, the only necessary condition is that the entrant does not target the highest realization of \( \theta \) (or its upper bound, in the case of the continuous distribution). In other words, it is sufficient that there be some realization that triggers litigation.

If (III) \( \pi^E_{\infty}(\theta) < I_E \), the entrant does not invest under any policy. CS is then the same under any policy.

The existence of case (II) completes the proof. The intuition is that if the originator’s investment cost is sufficiently small and the entrant’s cost is intermediate, the originator invests, but the entrant will if and only if reverse payments are allowed and sufficiently high. This increases the size of the pledgeable profits, making the entrant willing to invest. Some of these investments will end up in litigations and the ones that end up in a settlement will make the entrant enter before patent expiry. Both effects increase CS.

**Case 2: more originator’s investment.** Consider now \( \lambda^0 < \lambda < \lambda^\infty \). The probability of the low realization of the patent strength is such that the entrant targets \( \underline{\theta} \) under \( \hat{R} = 0 \) and \( \tilde{\theta} \) when \( \hat{R} \) is sufficiently high. Therefore, the originator obtains higher profits when reverse payments are allowed and \( \hat{R} \) is sufficiently high. When \( \pi_0(\theta) < I_0 < \pi_O(\tilde{\theta}) \), allowing reverse payments makes the originator invest, which increases CS. The intuition is that reverse payments increases industry profits, which makes the entrant keener on settling on more favorable terms to the originator, in order to reduce the risk of litigation. This increases the originator’s profits and triggers the originator’s investment - which would not have taken place otherwise. This raises CS, as it creates a market that would not have existed otherwise. This is another reason why reverse payments have the potential of increasing CS.

These two cases show that a ban *per se* is suboptimal. Section 2.5 will provide a numerical example for both cases. Next subsection derives the impact of the AA policy on CS and derives the optimal policy.

### 2.4 Ranking of CS and optimal policies

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

1) The best outcome for CS is that both parties invest, the entrant targets \( \underline{\theta} \) and the policy is \( \hat{R} = 0 \). In this case entry occurs as soon as possible, if parties settle, and litigation is possible. In this case, consumer surplus is

\[
CS^0(\theta) = \lambda[(\theta - \frac{C_O}{H - L})\bar{S} + (1 - \theta + \frac{C_O}{H - L})\tilde{S}] + (1 - \lambda)[\tilde{\theta}\bar{S} + (1 - \tilde{\theta})\tilde{S}].
\]
2) The second-best outcome is that, under \( \hat{R} = 0 \), the entrant enters and targets \( \bar{\theta} \). In this case consumer surplus is

\[
CS^0(\bar{\theta}) = (\bar{\theta} - \frac{CO}{H - L})S + (1 - \bar{\theta} + \frac{CO}{H - L})\bar{S}. \tag{21}
\]

3) The third-best outcome is that, under \( \hat{R} = 0 \), the entrant would not enter because his profits are smaller than his investment cost, but the AA can equalize them by choosing the appropriate maximal reverse payment. In this case consumer surplus is

\[
CS^R(\bar{\theta}) = \lambda[(\bar{\theta} + \frac{\hat{R} - CO}{H - L})S + (1 - \bar{\theta} - \frac{\hat{R} - CO}{H - L})\bar{S}] + (1 - \lambda)[\bar{\theta}S + (1 - \bar{\theta})\bar{S}], \tag{22}
\]

when the entrant still targets \( \bar{\theta} \), and \( CS^R(\tilde{\theta}) = (\tilde{\theta} + \frac{\hat{R} - CO}{H - L})S + (1 - \tilde{\theta} - \frac{\hat{R} - CO}{H - L})\bar{S} \) if the cap \( \hat{R} \) is such that the entrant now targets \( \tilde{\theta} \).

4) If, under \( \hat{R} = 0 \), the entrant would enter and target \( \bar{\theta} \) but the originator’s investment cost \( I_O \) is between the profits he would get when the entrant targets the low realization and the ones when the entrant targets the high realization, the originator does not invest. Consumer surplus, therefore, would be \( CS^0(O_{out}) = 0 \). The fourth-best outcome consists, therefore, of the AA allowing reverse payments with a cap such that the entrant targets \( \bar{\theta} \). The entrant targets \( \tilde{\theta} \) if and only if the probability \( \lambda \) that the patent is weak is smaller than \( \lambda^\hat{R} \), which occurs when the entry date is sufficiently high. Under such a policy, consumer surplus is

\[
CS^R(\bar{\theta}) = (\bar{\theta} + \frac{\hat{R} - CO}{H - L})S + (1 - \bar{\theta} - \frac{\hat{R} - CO}{H - L})\bar{S}.
\]

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 would eventually enter the market. If he would eventually enter, therefore, consumer surplus would be \( CS^0(O_{out}) = 0 \). The fifth-best outcome, therefore, consists of the AA implementing a policy such that the entrant’s entry cost \( I_E \) is larger than the profits he makes by entering. This makes the generic producer stay out of the market and, therefore, it makes the originator invest. Therefore \( CS^0(E_{out}) = S \). 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) The worst outcome occurs when the originator’s investment cost \( I_O \) is higher than the monopoly profits \( H \). In this case, nothing can be done to push (at least) the originator to invest, therefore \( CS^0(O_{out}) = 0 \).

The objective of the AA is to maximize CS. The full characterization of the optimal policy is complex (see the Appendix 5.3). The following proposition focuses on the cases where allowing reverse payments is optimal.

---

\( CS^0(\bar{\theta}) \) is smaller than \( CS^0(\bar{\theta}) \) if and only if \( \lambda \geq \frac{CO}{(\bar{\theta} - 2)(H - L)} \). We assume that this is the case.

The ranking \( CS^0(\bar{\theta}) > CS^R(\bar{\theta}) \) is true as long as \( \hat{R} > (1 + \bar{\theta} - \bar{\theta})(H - L) - \frac{(1 - \lambda)CO}{\lambda} \), otherwise it is reversed.
Proposition 2 Allowing reverse payments is optimal when:

(1) \( \pi_0(\bar{\theta}) < I_0 \leq \pi_0(\bar{\theta}) \), (1.a) \( \lambda \leq \lambda^0 \) and (1.a.i) \( I_E > \pi^R_E(\bar{\theta}) \): the optimal reverse payment \( \hat{R} \) is the one such that \( I_E = \pi^R_E(\bar{\theta}) \); (1.b) \( \lambda^0 < \lambda \leq \lambda^\infty \): the optimal reverse payment is \( \hat{R} = \min \{ R \text{ s.t. } I_E \leq \pi^R_E(\bar{\theta}) \text{ and } \lambda \leq \lambda^R \} \).

(2) \( 0 \leq I_0 \leq \pi_0(\bar{\theta}) \) and (2.a) \( \lambda \leq \lambda^0 \) and \( I_E > \pi^0_E(\bar{\theta}) \): the optimal reverse payment \( \hat{R} \) is the one such that \( I_E = \pi^R_E(\bar{\theta}) \); (2.b) \( \lambda^0 < \lambda \leq \lambda^\infty \) and \( I_E > \pi^0_E(\bar{\theta}) \), then the optimal reverse payment is \( \hat{R} = \min \{ R \text{ s.t. } I_E \leq \pi^R_E(\bar{\theta}) \text{ and } \lambda \leq \lambda^R \} \); if (2.c) \( \lambda > \lambda^\infty \) and \( I_E > \pi^0_E(\bar{\theta}) \), then the optimal reverse payment is the one such that \( I_E = \pi^R_E(\bar{\theta}) \).

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 entrant enter and target \( \bar{\theta} \). In subcase (1.a), the entrant would target the high realization under any policy, but if reverse payments are banned he does not enter. Therefore, reverse payments must be allowed and their size should just make the entrant’s profits match his investment cost. This makes the entrant enter as soon as possible. In case (1.b), the probability of the low realization is such that the entrant targets the low realization if reverse payments are banned, but he targets the high one if the allowed reverse payment is sufficiently large. Therefore, it is optimal to allow reverse payments with a size such that the entrant can recoup his investment cost and he targets the high realization. This policy makes the originator invest, the entrant enter and the monopoly period is minimized.

In case (2), the originator’s investment cost is so small that he invests for any realization the entrant eventually targets. Therefore, the only problem is to make the entrant enter. Subcase (2.a) is the same as (1.a) above. For a higher probability of the low realization and an investment cost that does not make the entrant enter when reverse payments are banned (2.b), the optimal policy is \( \hat{R} = \min \{ R \text{ s.t. } I_E \leq \pi^R_E(\bar{\theta}) \text{ and } \lambda \geq \lambda^R \} \). This policy makes the entrant recoup his investment cost with the earliest entry date. For the case (2.c), where the probability of the low realization is so high that the entrant always targets it, the optimal policy is \( \hat{R} = R \text{ s.t. } I_E = \pi^R_E(\bar{\theta}) \); so that the entrant can recoup it and enter as soon as possible.

A natural question is what is the role of patent strength. Patent strength has two opposite impacts on the optimal policy. On one hand, reverse payments with a strong patent make the additional generic entry ending up in litigation have a lower probability of increasing consumer surplus; on the other one they also induce just a small delay, because the expected entry under litigation is already late. In other words, a strong patent makes both the benefits and the costs of reverse payments smaller. These two opposite forces make the impact of patent strength on the optimal treatment of reverse payments ambiguous. Appendix 5.4 discusses this.
2.5 Numerical examples

This section provides a numerical example for each of the two scenarios where reverse payments increase CS.

Scenario 1: more generic entry. Assume $I_O = 25$, $E = 10$, $H = 40$, $L = 20$, $C_E = C_O = 2$, $\theta = \frac{1}{3}$, $\tilde{\theta} = \frac{2}{3}$, $\lambda = \frac{4}{5}$. These values imply that $\lambda R = \frac{R+6}{R+12.56}$. For any $\tilde{R} < 20.66$ we have $\lambda > \lambda R$, i.e., the entrant targets the low realization if he enters. The originator will (i) accept $R = \tilde{R}$, $D(\tilde{\theta}) = \tilde{\theta} + \frac{R-C_O}{H-L}$ if and only if $\theta = \tilde{\theta}$ and (ii) accept $R = \tilde{R}$, $D(\bar{\theta}) = \bar{\theta} + \frac{R-C_O}{H-L}$ for any realization of $\theta$. The effective cap on $\tilde{R}$, i.e., the one that makes the agreed entry date equal to the patent expiry, is equal to $\tilde{R} = (1-\theta)(H-L) + C_O$ (see Lemma 1), which is equal to 15.33 when the entrant targets the low realization and 8.66 when he targets the high one.

From (2) and (3) we have

(i) $\pi^R_E(\bar{\theta}) = \lambda[(1-\theta - \frac{R-C_O}{H-L})E + \tilde{R}] + (1-\lambda)[(1-\tilde{\theta})E - C_E]$;

(ii) $\pi^E_E(\bar{\theta}) = (1-\theta - \frac{R-C_O}{H-L})E + \tilde{R}$.

Given the parameters assumed above, we get $\pi^R_E(\bar{\theta}) = \frac{4R}{5} + 4.33$ and $\pi^E_E(\bar{\theta}) = \frac{4R}{5} + 6.4$.

The profits of the originator are $\pi_O(\bar{\theta}) = 31.33$ and $\pi_O(\theta) = 26.89$. When $\tilde{R} < 20.66$, the entrant targets $\bar{\theta}$, so, being the effective cap on reverse payments equal to 15.33 $< 20.66$, the relevant comparison is between $I_O$ and $\pi_O(\theta)$ for any policy. We have $I_O < \pi_O(\bar{\theta})$, therefore the originator invests under both policies. So we have three cases, depending on $I_E$:

(I) $0 < I_E \leq 6.4$ : the entrant invests even when $\tilde{R} = 0$, therefore the CS is highest under $\tilde{R} = 0$;

(II) $6.4 < I_E \leq 15.33$: the entrant invests only for a sufficiently high $\tilde{R} > 0$: the CS is highest under $\tilde{R} = 2.5I_E - 16$, as this is the lowest reverse payment that allows the entrant to recoup his investment cost, entry occurs as soon as possible (before patent expiry for any $I_E < 15.33$) and litigation occurs with positive probability (the probability $(1-\lambda) = \frac{1}{3}$ of drawing $\tilde{\theta}$);

(III) $I_E > 15.33$: the entrant does not invest in any case.

Case II shows that reverse payments can increase CS by increasing generic entry before patent expiry and by increasing the probability of litigation.

Scenario 2: more originator’s investment. Assume the same parameters as above except for $I_O$, which is now in the range $\pi_O(\bar{\theta}) < I_O \leq \pi_O(\theta)$ and $\lambda = \frac{1}{2}$.

We have $\pi^R_E(\bar{\theta}) = \frac{4R}{5} + 4.33$ and $\pi^E_E(\bar{\theta}) = \frac{4R}{5} + 4.5$. The entrant, therefore, targets $\bar{\theta}$ if and only if $\tilde{R} \geq 0.66$. The originator’s profits are $\pi_O(\bar{\theta}) = 31.33$ and $\pi_O(\theta) = 28$. Therefore, being $\pi_O(\bar{\theta}) < I_O < \pi_O(\theta)$, allowing reverse payments with $\tilde{R} \geq 0.66$ makes the originator invest, which increases CS. The optimal policy is exactly $\tilde{R} = 0.66$: the originator invests and the monopoly period is minimized. Note that even a "laissez-faire" policy yields a strictly higher consumer surplus than $\tilde{R} = 0$, as it makes the originator invest.
3 Case II - The model

This section describes the second case, i.e. the one treating the liquidity problems of generic pharmaceutical firms. The model is very similar to the previous case, except for the following modifications. In the second stage, the entrant learns his type, which can be weak or strong. It is weak with probability \( \mu \in [0, 1] \). When the entrant is weak, if the originator and the entrant do not settle with a reverse payment at least equal to a threshold \( k \), the entrant is not able to compete and exits the market immediately;\(^{23}\) 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 litigate or settle their dispute. The bargaining process, like in the previous case, 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.\(^{24}\) The offer consists of an entry date \( D \) and a reverse payment \( R \) from the originator to the entrant. If the originator accepts the offer \( D \) and \( R \) are enforced, while if he rejects it parties litigate. In the fifth stage, if the 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 chooses a cap \( \hat{R} \).
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 case is that here, before the settlement offer, (i) the entrant draws a private signal on his financial strength (in the first case, the originator draws a private signal over the patent strength) and (ii) investment decisions are not taken into account.\(^{25}\) The notation is like in the previous case. The originator’s patent strength \( \theta \) is now common knowledge. In the pre-expiry period, like in the previous case, let \( H \) be 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

---

\(^{23}\) This represents a situation where the entrant goes bankrupt or simply prefers to abandon that market.

\(^{24}\) The fact that the entrant makes the offer is, like in the previous case, 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 can make 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 other party (in this case, the originator) making the take-it-or-leave-it offer, like in the previous case.

\(^{25}\) If we had kept the investment decisions, the possibility of the entrant’s liquidity constraints lowers the threshold for the originator’s investment, but does not add any relevant insight.
enters immediately. Hence, $L + E$ are the joint profits of the originator and entrant if entry occurs immediately. We assume that $H > L + E$. 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. As in the previous case, $\bar{S}$ is the consumer surplus (CS) in the pre-expiry period when only the originator is active and $\bar{S}$ the pre-expiry CS when also the entrant is on the market; moreover, denote $\bar{s}$ the post-expiry CS when the entrant has exited the market and $\bar{s}$ when the entrant is on the market. Of course we have $\bar{S} \geq \bar{S}$ and $\bar{s} \geq \bar{s}$. The weight of the post-expiry period, in both the parties’ profits and CS, is $\delta$. Note that $\delta$ may be higher than 1, as the post-expiry period of the drug can be much longer than the pre-expiry one.\footnote{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.}

The basic tradeoff for CS is that, when reverse payments are sufficiently large, a part of the higher industry profits can be reaped by the entrant, who is therefore able to remain on the market even if he is weak. This increases CS both in the pre-expiry and in the post-expiry period. On the other hand, high reverse payments typically make the parties agree on a later entry date, making the originator’s monopoly period longer, which lowers the CS in the pre-expiry period if the entrant is strong. This case will show that there exist parameter sets where the first (positive) effect offsets the second (negative) one, even when entry occurs at patent expiry.

The following subsection computes 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. Appendix 6.1 will discuss the existence of separating equilibria.

### 3.1 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:** $\left[(1 - \theta)E - C_E\right] + \delta e$ 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 gets $l$ in the post-expiry period. Whether he wins or loses, provided that the entrant is strong (and, therefore, litigation actually occurs), litigation costs are $C_O$. The strong entrant, instead, 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 would not be able to stay on the market if they litigate, so his litigation payoff is 0.

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

\[
\text{Originator: } \mu(H + \delta h) + (1 - \mu)[DH + (1 - D)L + \delta l] - R
\]

\[
\text{Entrant: } (1 - D)E + R + \delta e \text{ if he is strong, } R \text{ otherwise.}
\]

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

\[
\text{Originator: } DH + (1 - D)L - R + \delta l
\]

\[
\text{Entrant: } (1 - D)E + R + \delta e \text{ independently from his type.}
\]

By settling, the originator enjoys \(DH\) in the period before the agreed entry date and \((1 - D)L\) in the period until patent expiry. He pays \(R\) to the entrant and, finally, obtains \(\delta l\) from the post-expiry period. The entrant earns \((1 - D)E\) if he enters at date \(D\), receives the payment \(R\) and obtains \(\delta e\) after patent expiry. The same occurs when the reverse payment is smaller than \(k\), provided that the entrant is strong - otherwise, he only gets \(R\) and exits the market.

When the entrant’s settlement offer includes a reverse payment smaller than \(k\), the originator accepts it as long as

\[
D \geq D^{R<k} = \theta - \frac{C_O}{H - L} + \frac{R}{(1 - \mu)(H - L)}.
\]

When the reverse payments required by the entrant is at least equal to \(k\), the originator knows that, by accepting, he is making the weak entrant stay on the market. Therefore he will accept the settlement offer if and only if

\[
D \geq D^{R\geq k} = \theta + \frac{R - C_O}{H - L} + \mu(1 - \theta + \frac{C_O + \delta(h - l)}{H - L}).
\]

The minimal entry date acceptable for the originator is, like in the first case, increasing in \(R\) and \(\theta\) and decreasing in \(C_O\) and \((H - L)\), as long as \(R > C_O\). It is also increasing in the probability \(\mu\) that the entrant is weak: the higher this probability, the less the originator is willing to pay a reverse payment that keeps him on the market. For the sake of exposition, consider the parameter sets where both \(D^{R<k}\) and \(D^{R\geq k}\) are smaller than 1.\(^{27}\)

\(^{27}\)If any of the entry dates were higher than 1, the entrant would just reduce the reverse payment he asks (with respect to \(k - \varepsilon\) when he asks for \(D^{R<k}\) and with respect to \(\hat{R}\) when he asks for \(D^{R\geq k}\) in order to keep the originator indifferent between accepting and litigating with \(D = 1\)). This would complicate the exposition without changing any of the results.
Consider now the strong entrant’s incentives. By comparing his settlement profits with $R \geq k$ and $R < k$, given the originator’s minimal required entry dates $D_{R<k}$ and $D_{R\geq k}$, the strong entrant will propose a settlement with a reverse payment larger or equal to $k$ if and only if $D_{R\geq k} \leq 1$ and $(1 - D_{R\geq k})E + \hat{R} + \delta \epsilon \geq (1 - D_{R<k})E + k + \delta \epsilon$, which yields

$$
\hat{R} \geq R^* = \frac{(1-\mu)(H-L)}{(1-\mu)(H-L) + \mu E} \{k + \mu[1 - \theta + \frac{CO + \delta(h-l)}{H-L}]E\}. \tag{6}
$$

Intuitively, the minimal differential between the maximal reverse payment $\hat{R}$ and the amount $k$ required by the weak entrant to remain on the market increases with the probability $\mu$ that the entrant is weak. If $\hat{R} < R^*$, then the strong entrant prefers to ask for $k - \epsilon$ in order to reduce the originator’s expected cost of settling by making an offer that pushes the weak entrant out of the market. This makes the originator willing to accept the earlier entry date $D_{R<k}$.

Consider now the weak entrant’s incentives. He gets positive profits only through $R$ before patent expiry and remains on the market after patent expiry only if $R$ is at least equal to $k$, so his incentive to ask for a reverse payment is stronger than for the strong entrant. But the originator knows that, if $\hat{R} < R^*$, the strong entrant prefers to ask for $k - \epsilon$, so the originator would understand that the entrant is weak if he asked 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 larger or equal to $k$ 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, therefore, no better option than just mimicking the strong entrant’s strategy. That is why we just refer to the "entrant", regardless of his type, in the following analysis.

When $R^* > k$, we have three possible outcomes depending on the maximal reverse payment $\hat{R}$: (1) $\hat{R} < k$ : the entrant asks for $\hat{R}$ and the weak type exits the market; (2) $k \leq \hat{R} < R^*$ : the entrant asks for $k - \epsilon$ and the weak type exits the market; (3) $\hat{R} \geq R^*$ : the entrant asks for $\hat{R}$ and the weak type survives. Note that the signaling feature of this case changes Lemma 1: in the parameter set $k \leq \hat{R} < R^*$, the strong entrant asks for a reverse payment smaller than $\hat{R}$ and, therefore, get higher profits through an earlier entry date. The weak entrant uses the same strategy, as he cannot do anything better.

This leads to the following Lemma.

**Lemma 5** Any policy $\hat{R} < R^*$ makes the weak entrant exit the market.

---

28 We assume that the weak entrant needs a sufficiently high reverse payment to avoid bankruptcy - an early entry date is not sufficient. This assumption simplifies the analysis but does not change the main conclusions. The only necessary feature for the result is that a settlement with a reverse payment yields higher profits for the entrant than a settlement without it.

29 We assume, for the sake of exposition, that the entry date $D$ implied by $\hat{R}$ is not larger than 1. Any $\hat{R}$ larger than that will make the entrant ask for a reverse payment $R$ such that $D = 1$ (like in Lemma 1).

30 When $R^* \leq k$, case (2) disappears. In the following, to make the analysis more complete, assume that $R^* > k$.  

19
This Lemma will turn out to be useful in the analysis of the optimal policy of the AA. The next section analyzes consumer surplus under different policies.

3.2 Consumer surplus and optimal policies

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

When the parties settle with $R < k$, we have

$$CS^{R < k} = \mu (\bar{s} + \delta s) + (1 - \mu) [D^{R < k} S + (1 - D^{R < k}) \tilde{S} + \delta \bar{s}].$$ \hspace{1cm} (7)

With a probability $\mu$ the generic is weak and exits the market, leaving the monopoly to the originator until patent expiry, after which 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 $D^{R < k}$, to the duopoly CS after $D^{R < k}$ before patent expiry and to the competitive CS after it.

When the parties settle with $R \geq k$, which occurs when $\tilde{R} \geq R^*$, we have

$$CS^{R \geq k} = D^{R \geq k} S + (1 - D^{R \geq k}) \tilde{S} + \delta \bar{s}.$$ \hspace{1cm} (8)

This is the sum of the monopoly CS $S$ until $D^{R \geq k}$, the duopoly CS $\tilde{S}$ after $D^{R \geq k}$ until patent expiry and the competitive CS with the entrant still active, $\bar{s}$, after patent expiry.

By analyzing $CS^{R < k}$ we have the following Lemma.

**Lemma 6** $CS^{R < k}$ decreases with $\tilde{R}$.

Therefore, for any settlement with $0 < R < k$, entry is delayed compared to a settlement without a reverse payment, which reduces CS. 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 used but they are too small to make the weak entrant stay on the market.

*Disregarding the entrant’s bankruptcy problem,* it is easy to see that CS is higher under $\tilde{R} = 0$. A "laissez faire" policy makes firms choose a later entry date (equal to $D^{R < k}$ or 1, depending on the choice that grants more profits to the entrant), while banning them makes the entrant propose $D^{R = 0} = \theta - \frac{C_O}{(H - L)}$, that is strictly smaller than $D^{R < k}$ and 1. Being CS decreasing in the generic’s entry date, it is clear that CS is higher under a ban on reverse payments. Under a ban on reverse payments, the entrant asks to enter earlier. This
supports the FTC and EC’s opinion that reverse payments should be banned \textit{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 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 overwhelms the negative effect of late entry, making CS lower than if reverse payments had been allowed.

This leads, like in case I, to the following Proposition.

\textbf{Proposition 3} \textit{There exist parameter sets where banning reverse payments reduces consumer surplus.}

\textbf{Proof}. First of all, given Lemmas 5 and 6, we can disregard the policies $\hat{R} < R^*$, because the entrant does not ask for a reverse payment that allows the weak entrant to remain on the market and entry is delayed compared to a ban scenario. We can, therefore, just focus on a policy $\hat{R} = 0$ and on the policies $\hat{R} > R^*$.

From (7) and (8), the former with $\hat{R} = 0$, we get that allowing reverse payments increases consumer surplus when

$$\hat{R} < \tilde{R} = \mu (H - L) \delta \left[ \frac{(\tilde{s} - s)}{(S - S)} - \frac{(h - l)}{(H - L)} \right], \quad (9)$$

as long as $R^* < \hat{R}$. In particular, CS is maximal when $\hat{R} = R^*$ if $R^* < \hat{R}$ and when $\hat{R} = 0$ otherwise. Inequalities (6) and (9) yield a number of policy implications.

\textbf{Corollary 1} \textit{If the ratio \textit{of the increase in future CS over the increase of current CS from having an additional competitor $\frac{(\tilde{s} - s)}{(S - S)}$ is lower than the ratio of the increase in future profits over current profits from having a competitor less $\frac{(h - l)}{(H - L)}$, reverse payment do not increase consumer surplus.}

\textbf{Proof}. When $\frac{(\tilde{s} - s)}{(S - S)} < \frac{(h - l)}{(H - L)}$, we have $\hat{R} < 0$, so no reverse payment can increase CS.

When $\frac{(\tilde{s} - s)}{(S - S)} > \frac{(h - l)}{(H - L)}$, we have the following corollaries.

\textbf{Corollary 2} \textit{The higher the probability $\mu$ that the entrant is weak, the larger the parameter set where it is optimal to allow for reverse payments.}

\textbf{Proof}. Both in (6) and in (9) a higher $\mu$ makes the inequalities easier to fulfill.

A consequence is that during an economic downturn it is better to be \textit{more lenient} towards reverse payments, as $\mu$ would be higher. An economic downturn means, \textit{inter alia}, lower profitability and worse credit crunch - conditions that negatively impact 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.
Corollary 3  (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, as long as the number of firms is sufficiently low, is possible. The idea is that one additional competitor, further than the traditional pro-competitive effect, could 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. 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 firms to stay on the market through a lenient policy can increase CS, as it may reduce the sustainability of collusion.

Finally, note that even a “laissez-faire” policy of no cap to reverse payments may increase CS compared to a ban scenario.

Corollary 4  A "laissez-faire" policy yields higher CS than \(\bar{R} = 0\) if and only if \((1 - \mu)((1 - \theta)(H - L) + C_O) - \mu \delta (h - l) < \mu (H - L) \delta \left[\frac{(\bar{s} - s)}{s - \underline{s}}\right] - \frac{(h - l)}{(H - L)}\).

Proof. When no cap is set on reverse payments, the parties agree on the one such that the entry is upon parent expiry \((D = 1)\), as it is the one that maximizes the industry - and the entrant’s - profits. By substituting \(D^{R\ge k} = 1\) in (5), we get that the associated reverse payment is \(R(\bar{D} = 1) = (1 - \mu)((1 - \theta)(H - L) + C_O) - \mu \delta (h - l)\). By plugging it into (9) we get the result.

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 patent strength (case I), banning reverse payments may reduce consumer surplus. The reason is two-fold. First, banning reverse payments reduces the industry profits, which reduces the entrant’s incentives to invest and implies less generic entry. This, together with the asymmetric information over the patent strength, reduces the litigation rate and consumer surplus. This result is robust to any bargaining rule between the originator and the entrant, 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 receives a private signal, this result holds also if it is the entrant who gets the private signal. This result just requires the entrant to get some, however small, additional surplus from the settlement compared to litigation. The only necessary features for the results
are some asymmetric information between the parties and that the informed party does not make a take-it-or-leave-it offer. To see that this result holds also when the entrant gets the private signal, consider the following example. Now the entrant gets the private signal and the originator makes a take-it-or-leave-it settlement offer. The entrant now earns strictly more than his expected litigation payoffs not because of his bargaining power, but because of his information rent. Reverse payments increase the industry profits and, therefore, the originator’s willingness to avoid litigation. This makes the originator keener on settling on more favorable terms to the entrant. This increases the entrant’s profits, thus increasing his incentive to enter the market. Litigation is still possible and, therefore, the main result that reverse payments delay entry to increase entry still holds. Second, reverse payments creates a tension in the originator’s incentives to invest. They make industry profits larger, which can make (i) the entrant willing to enter, which reduces the originator’s incentives to invest, but also (ii) the entrant (provided that he invests) keener on settling on more favorable terms to the originator, which increases the originator’s incentives to invest. This result is robust to any bargaining rule between the originator and the entrant.\footnote{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 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 the tension for the originator would still exist.}

The main result is that allowing reverse payments delays generic entry but increases 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, moreover, that even a laissez-faire policy that allows complete freedom over the use of reverse payments (i.e. no cap is set) may be superior to a ban per se. Finally, note that patent strength has an ambiguous impact on the optimal policy. Two forces are present: on one hand, a reverse payment on a strong patent makes it unlikely that litigation ends up in generic entry; on the other one, a reverse payment on a strong patent also involves a small cost, given that the additional delay with respect of the expected one is small. For a more detailed discussion of patent strength and optimal policy, see Appendix 5.4.

For a practical enforcement, in order to know which agreements should be banned, the AA should have a rough estimate of $I_E$, $I_O$, $\pi^R_E(\hat{\theta})$ and $\pi^O(\hat{\theta})$. These estimates can be recovered from the expenses in R&D (for $I_O$), bioequivalence studies and marketing authorisations (for $I_E$), from the market profits and from the settlement agreements for similar products (for $\pi^O(\hat{\theta})$, $\pi^R_E(\hat{\theta})$ and $\pi^O(\hat{\theta})$). The AA can also estimate the parties’ profits by analysing the expected price decline due to competition. This can be done by analysing the price evolution of similar products when generic entry took place. With this estimate of the competitive price, the AA can estimate the profits $\pi^R_E(\hat{\theta})$ 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 reverse payments.
Banning reverse payments can reduce consumer surplus also because of the possibility of bankruptcy of generic entrants (case II). The reason is that banning reverse payments makes the entrant get less than he would with a settlement including a reverse payment, therefore exacerbating his possible liquidity constraints. If the entrant is liquidity constrained, a settlement without reverse payments does not allow him to raise the funds he needs and forces him to exit the market. This reduces consumer surplus both before and after patent expiry. This effect goes in the opposite direction to 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 static consumer surplus loss due to the later entry associated with reverse payments, including when this is the patent expiry date. This suggests that a rule of reason is more suited than a ban \emph{per se} on reverse payments.

In the first case, the main result was that reverse payments can have a positive effect on the originator’s incentives to invest and always have a positive effect on the entrant’s incentives to invest. In that case, therefore, reverse payments allow for more entry, while here they allow for less exit. In both cases, they keep a higher competitive pressure on the market. This effect overwhelms the negative effect of the late entry due to reverse payments under many parameter sets, making CS higher.

For a practical enforcement, the estimates the AA needs 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 profit 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 decisions of the local Patent Office, etc) and the litigation costs of the parties. Some importance should be given to the general economic situation too, as a period of economic downturn makes the optimal policy more lenient.

References


5 Case I - Appendix

5.1 Reverse payments and late entry

Proof of Lemma 1. Compare the profits the entrant obtains from offering the reverse payment \( R = \hat{R} \) and the associated entry date \( D(\hat{\theta}) = \hat{\theta} + \frac{\hat{R} - C_{O}}{H - L} \) with the ones from asking for a smaller reverse payment \( R < \hat{R} \) and the associated entry date \( D(\hat{\theta}) = \hat{\theta} + \frac{\hat{R} - C_{O}}{H - L} \). Note that, from (1), \( D(\hat{\theta}) \) is the optimal entry date given the lower reverse payment, as it keeps the originator indifferent between accepting and refusing the offer for the (possibly new) targeted realization \( \hat{\theta} \). Consider, first, the case where the entrant targets the same realization: \( \hat{\theta} = \hat{\theta} \).

In this case the probability of litigation is the same, so we can just compare the entrant’s settlement profits from the offer \( \{\hat{R}, D(\hat{\theta})\} \) with the ones from the alternative offer \( \{R < \hat{R}, D(\hat{\theta})\} \). Note that a lower reverse payment implies an earlier entry date: a marginally lower \( R \) makes the entrant lose \( dR \) through \( R \) and gain \( \frac{E}{H - L} dR \). Being the loss in the originator’s profits \( (H - L) \) higher than the entrant’s profits \( E \), the optimal \( R \) is the maximal possible one: \( R = \hat{R} \). Consider now the case where the entrant, as a consequence of lowering \( R \) from \( \hat{R} \) to \( \hat{R} \), targets a different realization: \( \hat{\theta} \neq \hat{\theta} \). This represents a further distortion with respect to the optimal offer: even under this different target, \( \hat{R} < \hat{R} \) is not optimal, because the entrant could now raise the reverse payment \( \hat{R} \) to \( \hat{R} \) and extract the additional originator’s surplus through it. Again, therefore, the optimal \( R \) is \( \hat{R} \). This Lemma holds under any bargaining rule.\(^{32}\)

5.2 No Entry Delay

Here we show that the entrant enters as soon as he can (footnote 12). Change the notation in the following way: \( 0 \) is now the date when the entrant is ready to enter and \( T \) the entry date he actually chooses. \( D \) and 1 remain, respectively, the entry date and the patent expiry date.

\(^{32}\)The larger is \( \hat{R} \), the longer the monopoly period - it is in both parties’ interest to make it as long as possible.
The entrant now can choose the moment \( T \) when he will make an offer to the originator. In order to simplify the notation, just assume that the patent strength is common knowledge.\(^{33}\)

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 \( R \) with \( \hat{R} \), we have:

\[
D \geq D^*(T) = \theta + \frac{\hat{R} - C_O}{(1 - T)(H - L)}.
\]

The minimal entry date that the originator is ready to accept is increasing in \( T \), as long as \( \hat{R} > C_O \). In this case, the larger the time elapsed between the moment when the entrant is ready to enter and the moment when he discusses the settlement with the originator, the later the entry date the originator is willing to accept. In other words, waiting is counterproductive for the entrant, as it only reduces the amount he can get. On the other hand, when \( \hat{R} < C_O \), the minimal entry date that the originator is willing to accept is decreasing in \( T \). The larger the time elapsed between the moment when the entrant is ready to enter and the moment when he discusses the settlement with the originator, the earlier he can actually enter (in case of settlement) *relatively to the (lower) patent validity*. This positive effect for the entrant must be weighted with the later date \( T \) when the settlement is discussed. Substituting \( D^*(T) \) in the entrant’s profits, we get \( \pi_E^\hat{R} = (1 - T)(1 - \theta - \frac{\hat{R} - C_O}{(1 - T)(H - L)})E + \hat{R} \). Its derivative with respect to \( T \) is \( \frac{d\pi_E^\hat{R}}{dT} = -(1 - \theta)E < 0 \), therefore the entrant prefers to enter as soon as possible.

In conclusion, the entrant chooses \( T = 0 \) and discusses the settlement as soon as possible.

### 5.3 Optimal Policy

This proposition gives the full characterization of the optimal policy.

\(^{33}\)Results, as will become clear, do not depend on this.
Proposition 4 An optimal policy\textsuperscript{34} is the following:

if (1) $I_O > H$, any policy is equivalent;
if (2) $\pi_O(\theta) < I_O \leq H$, then $\hat{R} = 0$;
if (3) $\pi_O(\theta) < I_O \leq \pi_O(\bar{\theta})$, then if (3.a) $\lambda \leq \lambda^0$ and (3.a.i) $I_E \leq \pi_E^0(\theta)$ then $\hat{R} = 0$; if (3.a.ii) $\pi_E^0(\theta) < I_E \leq \pi_E^\infty(\theta)$, then $\hat{R} = R$ s.t. $I_E = \pi_E^R(\theta)$; if (3.a.iii) $I_E > \pi_E^\infty(\theta)$ then any policy is equivalent; if (3.b) $\lambda^0 \leq \lambda \leq \lambda^\infty$, then $\hat{R} = \min\{R \text{ s.t. } I_E \leq \pi_E^R(\theta) \& \lambda \leq \lambda^R\}$; if (3.c) $\lambda > \lambda^\infty$, then any policy is equivalent;

if (4) $0 \leq I_O \leq \pi_O(\bar{\theta})$, (4.a) $\lambda < \lambda^0$ and (4.a.i) $I_E \leq \pi_E^0(\theta)$, then $\hat{R} = 0$; if (4.a.ii) $\pi_E^0(\theta) < I_E \leq \pi_E^\infty(\theta)$, then $\hat{R} = R$ s.t. $I_E = \pi_E^R(\theta)$; if (4.a.iii) $I_E > \pi_E^\infty(\theta)$, then any policy is equivalent; if (4.b) $\lambda^0 < \lambda < \lambda^\infty$ and (4.b.i) $I_E \leq \pi_E^0(\theta)$ then $\hat{R} = 0$; if (4.b.ii) $\pi_E^0(\theta) < I_E \leq \pi_E^\infty(\theta)$ then $\hat{R} = \min\{R \text{ s.t. } I_E \leq \pi_E^R(\theta) \& \lambda > \lambda^\infty, R \text{ s.t. } I_E \leq \pi_E^R(\theta) \& \lambda < \lambda^R\}$; if (4.b.iii) $I_E > \pi_E^\infty(\theta)$, any policy is equivalent; if (4.c) $\lambda > \lambda^\infty$ and (4.c.i) $I_E \leq \pi_E^0(\theta)$, then $\hat{R} = 0$; if (4.c.ii) $\pi_E^0(\theta) < I_E \leq \pi_E^\infty(\theta)$ then $\hat{R} = R$ s.t. $I_E \leq \pi_E^R(\theta)$; if (4.c.iii) $I_E > \pi_E^\infty(\theta)$, any policy is equivalent.

\textbf{Proof and Explanation.} In case (1) 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 profit but higher than the profits he makes when the entrant is present, the AA needs to deter the generic manufacturer’s entry. 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 $\bar{\theta}$. Reducing the entrant’s profits as much as possible, through $\hat{R} = 0$, is the best the AA can make in order to make the entrant not invest and, therefore, to make the originator invest. 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 $\hat{R} = 0$. If the entrant’s investment cost is intermediate (3.a.ii), then the AA must set a policy that allows the entrant to recoup it. In this case the optimal policy is $\hat{R} = R$ s.t. $I_E = \pi_E^R(\bar{\theta})$. This policy 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 (3.b), the entrant would target $\bar{\theta}$ under $\hat{R} = 0$ and $\bar{\theta}$ for a sufficiently high reverse payment: the optimal policy is therefore to allow reverse payments with a cap such that the entrant still targets the high realization and recoups his investment cost. In case (4), the originator’s investment cost is so small that he would invest for any realization the entrant may target.

\textsuperscript{34}It is "an" optimal policy because there can be other policies that yield the same CS. For example, in case (2), any policy other than $\hat{R} = 0$ yields the same CS if $I_E > \max\{\pi_E^R(\bar{\theta}),\pi_E^R(\theta)\}$, because the generic producer would not enter in any case.
Therefore, when the probability of the low realization is sufficiently small (4.a), an optimal policy is $\bar{R} = 0$. The entrant would target the high realization under both policies, so it is better to make him target it when reverse payments are banned. If the entrant’s investment cost is higher (4.a.ii and 4.a.iii), then the optimal policy is to allow reverse payments in order to make the entrant recoup his investment cost, when possible. When the probability of the low realization is slightly higher (4.b), the optimal policy is similar to the previous case. The only difference is that, when the profits are intermediate (4.b.ii), the optimal reverse payment makes the entrant target the low realization, if this allows him to recoup the investment cost. In other words, this policy makes the entrant recoup his investment cost with the earliest entry date and keeps him willing to target the low realization, or just implies the earliest possible entry date if it is impossible to make the entrant recoup his investment and target the low realization. For the case (4.c), where the probability of the low realization is so high that the entrant always targets the low one, then the optimal policy is very similar to the previous case, with the only difference that $\pi_{E}(\tilde{\theta})$ may be disregarded and only $\pi_{E}(\bar{\theta})$ needs to be considered when the entrant’s investment cost is intermediate.

5.4 Patent Strength and Optimal Policy

This paragraph discusses the ambiguous impact of the patent strength $\theta$ on the optimal policy. Recall that there are two opposite forces: a strong patent makes the generic entry ending up in litigation have a lower probability of increasing consumer surplus, but also it makes reverse payments induce just a small delay, because the expected entry under litigation is already late. We show here how each area of the optimal policy changes after an increase in patent strength. Assume, for simplicity, that an increase in patent strength consists of an increase of the same size of the two realizations $H$ and $L$.

In Area (1) of Proposition 4 nothing can be done to induce the originator to invest, so the patent strength has no impact.

In Area (2) it is optimal to ban reverse payments: the condition $\pi_{O}(\tilde{\theta}) < I_{O} \leq H$ can be rewritten as $\tilde{\theta}H + (1 - \tilde{\theta})L - C_{O} < I_{O} \leq H$, so an increase in $\theta$ makes the area smaller.

In Area (3) the optimal policy depends on the size of $\lambda$ and $I_{E}$: the condition of this area can be rewritten as $\tilde{\theta}H + (1 - \tilde{\theta})L - C_{O} < I_{O} \leq \tilde{\theta}H + (1 - \tilde{\theta})L - C_{O}$. Note, first, that an increase in $\theta$ has no impact on its size - both sides increase by the same amount. When (3.a) $0 < \lambda \leq \lambda_{0}$, which can be rewritten as $0 < \lambda \leq \frac{EC_{O}+(H-L)C_{E}}{(\tilde{\theta} - \theta)E(H-L) + EC_{O}+(H-L)C_{E}}$, an increase in $\theta$ has no impact either. When (3.a.i) $0 \leq I_{E} \leq \pi_{E}(\tilde{\theta})$, which can be rewritten as $0 \leq I_{E} \leq (1 - \tilde{\theta} + \frac{C_{O}}{H-L})E$, an increase in $\theta$ reduces the area, so, like in area (2), an increase in $\theta$ reduces the parameter region where a ban on reverse payments is optimal. However, this is not the case when (3.a.ii) $\pi_{E}(\tilde{\theta}) < I_{E} \leq \pi_{E}(\bar{\theta})$, which can be rewritten as $(1 - \tilde{\theta} + \frac{C_{O}}{H-L})E \leq I_{E} \leq (1 - \bar{\theta} + \frac{C_{O}}{H-L})(H - L)$. An increase in $\theta$ makes this area smaller, like above, but here it is optimal to allow reverse payments. The area that grows is the one where (3.a.iii) $I_{E} > \pi_{E}(\bar{\theta})$, 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) $\lambda^0 < \lambda \leq \lambda^\infty$, which can be rewritten as

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

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

In Area (4), when (4.a) $\lambda \leq \lambda^0$ the optimal policy is to ban reverse payments and an increase in $\theta$ has no impact. In areas (4.a.i), where it is optimal to ban reverse payments, and in (4.a.ii), where it is optimal to allow them, the impact is ambiguous, because a higher $\theta$ makes the condition for $I_O$ larger but the one for $I_E$ smaller. In area (4.a.iii), where the policy has no effects, a larger $\theta$ makes the area larger. When (4.b) $\lambda^0 < \lambda \leq \lambda^\infty$, it is optimal to allow reverse payments and an increase in $\theta$ makes the area smaller. In the subcase of (4.b) the impact is always ambiguous because $\theta$ makes the area of $I_O$ larger but the one of $\lambda$ smaller.

When (4.c) $\lambda > \lambda^\infty$, an increase in $\theta$ has an ambiguous effect too: when (4.c.i) $I_E \leq \pi^0_E(\tilde{\theta})$, the optimal policy is to ban reverse payments 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) $I_E > \pi^0_E(\tilde{\theta})$, the optimal policy is to allow reverse payments and an increase in $\theta$ makes the areas of the originator and the entrant’s investment costs larger, but the area of $\lambda$ smaller. In (4.c.iii) $\theta$ makes the parameter area larger and the policy is ineffective.

The overall interaction between the optimal policy and the patent strength is, therefore, ambiguous and no robust policy implications on patent strength can be derived.

The following table resumes the results. The columns for $I_O$, and $I_E$ represent the range of these parameters in each area, while column $\hat{R}^\ast$ represents the optimal policy. Ind means that any policy is indifferent on CS, yes means that the optimal policy sets a positive cap on $R$ and no means that the optimal policy is a ban on reverse payments. 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 $I_E$. The question marks mean that an increase in patent strength has an ambiguous impact on the considered area.

The overall interaction between the optimal policy and the patent strength is, therefore, ambiguous and no robust policy implications on patent strength can be derived.

The following table resumes the results. The columns for $I_O$, and $I_E$ represent the range of these parameters in each area, while column $\hat{R}^\ast$ represents the optimal policy. Ind means that any policy is indifferent on CS, yes means that the optimal policy sets a positive cap on $R$ and no means that the optimal policy is a ban on reverse payments. 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 $I_E$. The question marks mean that an increase in patent strength has an ambiguous impact on the considered area.
5.5 No Menu of Contracts

This subsection shows that the entrant cannot write a menu of contracts to make the originator truthfully reveal his type. Consider a candidate menu of contracts \{(\overline{D}, \overline{R}), (\overline{D}, \overline{R})\}, where \((\overline{D}, \overline{R})\) is designed for the low type and \((\overline{D}, \overline{R})\) for the high type. The constraints to fulfill are:

\[
\overline{D}H + (1 - \overline{D})L - \overline{R} \geq DH + (1 - D)L - R, \quad (IC_b)
\]

\[
DH + (1 - D)L - R \geq DH + (1 - \overline{D})L - \overline{R}, \quad (IC_g)
\]

\[
\overline{D}H + (1 - \overline{D})L - \overline{R} \geq \overline{\theta}H + (1 - \overline{\theta})L - CO, \quad (PC_b)
\]

\[
\overline{D}H + (1 - \overline{D})L - \overline{R} \geq \theta H + (1 - \theta)L - CO, \quad (PC_g)
\]

The first two inequalities are the incentive compatibility constraints to make each originator’s type prefer not to pretend to be the other type. Note that the originator’s true type does not enter these equations - it only enters the originator’s litigation payoff. Therefore the only way to fulfill these incentive compatibility constraints is to make them have the same value. We have, therefore, \(\overline{D}H + (1 - \overline{D})L - \overline{R} = DH + (1 - D)L - R\), that yields

\[
(\overline{D} - D)(H - L) = \overline{R} - R. \quad (IC_\overline{b} = IC_g)
\]
The third and the fourth inequalities are the participation constraints that make each type prefer not to litigate. Given that the left hand sides of the four inequalities above must be the same \( IC_\theta = IC_{\bar{\theta}} \), only the inequality with the larger right hand side can bind. This inequality is \( PC_{\bar{\theta}} \), as \( \bar{\theta}H + (1 - \bar{\theta})L - C_O \) is larger than \( \theta H + (1 - \theta)L - C_O \) because \( (\bar{\theta} - \theta)(H - L) > 0 \). Therefore, type \( \bar{\theta} \) is left with no rent and type \( \theta \) enjoys an information rent. Consider now the entrant’s profits. Recall that \( \lambda \) is the probability that the originator’s type is \( \theta \). The entrant’s problem is:

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

\[s.t. \ (\bar{D} - D)(H - L) = \bar{R} - \bar{R}.
\]

The derivative of \( \pi_E \) with respect to \( R \) is \( \lambda \frac{H - L - E}{H - L} > 0 \) and the one with respect to \( \bar{R} \) is \( (1 - \lambda) \frac{H - L - E}{H - L} > 0 \), therefore the entrant asks for the maximal allowed reverse payment. This makes \( \bar{R} = \bar{R} = \bar{R} \). This implies (for \( IC_{\bar{\theta}} = IC_{\theta} \)) that we also have \( D = \bar{D} \). The entry dates \( \bar{D} \) and \( \bar{D} \) associated with these reverse payments are the ones that make the \( PC_{\bar{\theta}} \) binding: \( \bar{D} = \bar{D} = \bar{\theta} + \frac{\bar{R} - C_O}{H - L} \). Given that \( D = \bar{D} \) and \( R = \bar{R} \), the candidate menu of contracts reduces to a single contract that leaves an information rent to the \( \theta \)-type. This contract is the settlement offer that targets \( \bar{\theta} \) in the main text. When the probability that the originator’s type is \( \bar{\theta} \) is too small, it is optimal to "shutdown" this type, offer a contract that extracts all the rent of the \( \theta \)-type and causes litigation when the originator draws the high realization \( \bar{\theta} \). This is exactly the settlement offer that targets \( \theta \) in the main text.

6 Case II - Appendix

6.1 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 stage. In a separating equilibrium, by definition, the originator perfectly knows what type of entrant he is facing. 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, the weak entrant always gets 0, because no settlement offer would ever be accepted by the originator. Therefore, if the strong entrant asks for a reverse payment, the weak entrant has an incentive to deviate from his candidate equilibrium strategy and to mimic the strong entrant’s strategy, to get some positive reverse payment and, therefore, some positive payoff. Therefore, no separating equilibrium with positive reverse payment can exist. There can only be a separating equilibrium without reverse payment. In such an equilibrium, the weak entrant still receives 0, but has no incentive to mimic the strong entrant’s strategy, as he would receive 0 in that case too. Given that
reverse payments are not used in this type of equilibrium, the policy of the AA has no impact on CS.