Optimal Incentives for Patent Challenges in the Pharmaceutical Industry

Enrico Böhme · Jonas Severin Frank · Wolfgang Kerber

Accepted: 21 April 2021 / Published online: 9 May 2021
© The Author(s) 2021

Abstract
In this paper, we show that a provision in antitrust law to allow patent settlements with a later market entry of generics than the date that is expected under patent litigation can increase consumer welfare. We introduce a policy parameter for determining the optimal additional period for collusion that would incentivize the challenging of weak patents and maximize consumer welfare. While in principle, later market entry leads to higher profits and lower consumer welfare, this can be more than compensated for if more patents are challenged as a result.

Keywords Antitrust · Collusion · Patent challenges · Patent settlements · Pharma

1 Introduction

It is a well-established empirical insight that patent offices grant many patents that are later found invalid when challenged in court (Allison & Lemley, 1998; Lemley, 2001; Lemley & Shapiro, 2005). Patents that are granted erroneously might lead to

Jonas Severin Frank is economist and serves at the Bundeskartellamt (Federal Cartel Office). At present, he holds the position of press officer and deputy spokesperson. The statements made in this paper reflect the authors’ views and should not be attributed to the Bundeskartellamt.

Enrico Böhme
boehmee@staff.uni-marburg.de

Jonas Severin Frank
severin.frank@gmail.com

Wolfgang Kerber
kerber@wiwi.uni-marburg.de

1 School of Business and Economics, Research Group Institutional Economics, Philipps-University Marburg, Barfüßertor 2, 35037 Marburg, Germany
2 Bundeskartellamt (Federal Cartel Office), er-Friedrich-Str, 53113 Bonn, Germany
3 School of Business and Economics, Research Group Economic Policy, Philipps-University Marburg, Am Plan 2, 35037 Marburg, Germany
unjustified market power that harms consumers through higher prices. Invalidating patents through patent litigation is an important instrument for solving this problem (Ayres & Klemperer, 1999; Lemley & Shapiro, 2005; Shapiro, 2003); but this raises the question of whether firms have socially optimal incentives to challenge potentially invalid (‘weak’) patents.

Since challenging a patent is costly (with costs of, e.g., searching for prior art, attorneys, courts or patent office fees, lengthy proceedings), this creates an underlying externality that results in fewer patent challenges: Potential challengers are not fully internalizing the positive effects of their patent challenge for society—in particular, in terms of the additional consumer surplus that is generated from market entry. The literature has discussed instruments to address this issue such as: the subsidization of patent challenges; the creation of cash-bounty programmes to encourage private patent examinations; or the reduction of costs for patent revocation (Farrell & Merges, 2004; Miller, 2004; Thomas, 2001).

In addition to this general problem, a pattern has emerged in the pharmaceutical industry whereby patent holders pay firms (usually generics) not to challenge their weak patents (‘reverse payments’) through patent settlements. These ‘pay-for-delay’ agreements have the consequence of delaying market entry and price competition. Patent settlements can therefore be an instrument for undermining the usefulness of patent litigation as a tool to solve the weak patent problem. Competition authorities in the U.S. and the E.U. have challenged a number of patent settlements with reverse payments in the pharmaceutical industry for violating antitrust rules.

The objective of this paper is to analyse whether the incentives to challenge weak patents should be considered in the antitrust assessment of patent settlements with reverse payments. Beyond all controversial discussions of the details, the basic arguments on the potentially anticompetitive effects of patent settlements with high reverse payments are widely accepted in the economic and legal discussion (Elhauge & Krüger, 2012; Shapiro, 2003; Willig & Bigelow, 2004; Woodcock, 2016) and were accepted by the Supreme Court in its 2013 Actavis decision (FTC v. Actavis 2013). According to this view, patent settlements do not violate antitrust laws if they do not lead to lower consumer welfare (through later market entry of generics) than the level that could be achieved through patent litigation.

The basic idea of the model that we are presenting in this paper is to analyse the trade-off between allowing later market entry in patent settlements (with higher profits for the originator and the generic firm) and the positive effects on consumer welfare from the greater incentives for generics to challenge a larger number of weak patents. We can show that incorporating consideration of patent-challenge incentives into the antitrust assessment under certain conditions increases consumer welfare by allowing an additional period of collusion – which is stipulated by a newly introduced policy parameter.

1 The general problem of firms’ not taking into account externalities through their market entry was already analysed by Dixit and Stiglitz (1977). In their model, there is a trade-off between the gains in consumer welfare from an additional differentiated product and the loss of economies of scale from spreading production across a larger variety of products.
The paper is structured as follows: In Sect. 2, we present a brief overview of the legal and economic discussion of the antitrust assessment of patent settlements and explain our basic approach. The model is presented in Sects. 3–5 (model framework, equilibrium analysis, and welfare analysis). Section 6 provides our analyses of comparative statics, while Sect. 7 addresses the policy parameter’s impact on innovation incentives. In Sect. 8, we briefly discuss the results and some conclusions.

2 Patent Settlements, Antitrust, and Incentives for Patent Challenges in the Pharmaceutical Industry

Since the prices of pharmaceutical products sharply decrease after the entry of generics, any unjustified delay in generic firms’ entry can lead to high additional health costs for consumers and society. For that reason, the U.S. Federal Trade Commission (FTC) has challenged patent settlements with reverse payments in the pharmaceutical industry since 1999 (FTC 2002). In particular, patent settlements in which the (patent-holding) originator firms pay large sums to generic firms—‘reverse payments’—and agree on future entry dates of generics have been the object of competition and antitrust law proceedings.

The main discussion in U.S. antitrust law has focused on the question of whether the existence of a large reverse payment—larger than the patent-challenge litigation costs—should be sufficient to presume the illegality of a patent settlement or whether a broader rule-of-reason approach should be applied. Particularly important was the ruling of the U.S. Supreme Court in the Actavis case, in which it was decided that patent settlements with high unexplained reverse payments can be a signal of the weakness of patents and therefore of the anticompetitiveness and illegality of such patent settlements; the Court concluded that a rule-of-reason approach was appropriate (FTC v. Actavis 2013; Edlin et al., 2015).

The European Commission has also investigated patent settlements and viewed agreements with restrictions on market entry and reverse payments as potentially anticompetitive and requiring close scrutiny by competition law.2 The Commission decided in several patent settlement cases with reverse payments that such settlements violate Art. 101 TFEU. Especially important are the Lundbeck and Servier cases, in which the General Court of the E.U. confirmed that such patent settlements can be anticompetitive agreements according to Art. 101 TFEU.3

In the economic discussion, there is a broad consensus that patent settlements with reverse payments from originators to generics can be an effective instrument to protect weak patents from invalidation through patent challenges. In addition, the criterion of Shapiro (2003)—patent settlements should not lead to lower consumer welfare than the level under patent litigation—has been broadly accepted in

---

2 See European Commission DG Comp (2009, p. 254).
3 See General Court Case T-467/13—Arrow Group and Arrow Generics v Commission, General Court Case T-691/14—Servier and Others v Commission.
the economic and legal discussion as a relevant normative criterion for the antitrust assessment of patent settlements.  

Basic models of the price effects of patent settlements show that reverse payments that are larger than litigation costs lead to an entry date of generics that is delayed beyond the expected entry date under litigation: the benchmark of the litigation solution (Elhauge & Krüger, 2012; Frank & Kerber, 2018; Shapiro, 2003). While this result strongly supports the concerns about reverse payments, subsequent economic analyses have shown that under less simple and more realistic assumptions about the negotiation situation—e.g., information asymmetries, risk aversion, multiple entrants, etc.—the conclusions about reverse payments are less clear, although the basic entry-delaying effect of reverse payments still exists (Dickey et al., 2010; Frank & Kerber, 2018; Willig & Bigelow, 2004).

A serious problem with this discussion is that it has so far almost exclusively focused on the effects of patent settlements on consumer welfare via price effects: on the question of whether patent settlements lead to delayed generic entry, with negative effects on consumer welfare. However, patent settlements can also affect consumer welfare through two additional channels: effects on innovation incentives; and effects on patent-challenging incentives (Frank & Kerber, 2018).

To date, only Elhauge and Krüger (2012), Manganelli (2014), and Woodcock (2016, 2017) have taken the innovation incentive effect into account when analysing the effects of patent settlements. Although we should be cautious about drawing excessively far-reaching conclusions from their models, their results suggest that competition authorities do not need to worry too much about negative effects on innovation incentives.

With respect to patent-challenging incentives, some scholars are concerned about an excessively restrictive policy towards patent settlements. The argument is that the possibilities for generics to make profits from patent settlements through reverse payments would decrease their incentives to challenge patents or hinder procompetitive settlements (e.g., Dickey et al., 2010, p. 399). Additionally, in the abovementioned Actavis ruling of the U.S. Supreme Court, Chief Justice Roberts states in his dissenting opinion that putting limits on the possibility to engage in patent settlements with certain entry dates reduces the incentives to challenge patents (FTC v. Actavis 2013, Roberts dissenting, p. 17). Manganelli (2014) develops a model in which allowing reverse payments can lead to additional generic entry and could offset the direct entry delay effect of reverse payments because higher industry profits imply higher incentives for generics to enter the market.

What is the economic intuition behind our approach? According to the Shapiro criterion, patent settlements violate antitrust rules if they lead to a negotiated market entry of generics that is later than the expected entry date under patent litigation. The basic idea of our model is to introduce a policy parameter that allows longer

---

4 In competition and antitrust policy in the U.S. and the E.U., the effect on consumer welfare (and not total welfare) is the relevant normative criterion.
collusion between patent holders and generic entrants beyond this expected entry date under litigation in antitrust assessments.

This means that we address the externality problem of costly patent challenges through larger profits for the patent holders that are shared via reverse payments with the generic entrant. As a result, patents that previously would not have been challenged are challenged. We show in our model that under certain conditions, the increase in consumer welfare through the challenging of additional weak patents can be larger than the additional costs to consumers from later generic entry than what would be expected under litigation.

The structure of our model is partly based upon the model of Gratz (2012), who was the first to offer an integrated analysis of price and challenge incentive effects with regard to patent settlements. In her model, judicial errors (due to information problems in the application of the rule of reason) lead to overall positive effects on consumer welfare through erroneously allowing later generic entry than that in the benchmark litigation solution.

In contrast to her approach, we introduce a policy parameter and analyse the optimal additional delay (compared to the litigation solution) that would maximize consumer welfare if challenging incentives are also taken into account. In addition, we analyse the effect of such an optimal policy parameter on innovation incentives and, in a comparative statics analysis, how this optimal policy parameter depends, e.g., on the quality of patents and on the challenge costs.

### 3 Model Framework

Our model follows the basic framework introduced by Gratz (2012). In this framework, an originator \((O)\) holds a patent that expires at \(t = 1\), where \(t \in [0, \infty)\) denotes time. The patent strength is described by \(\gamma \in [0, 1]\), which allows us to interpret \(\gamma\) as reflecting the probability that the patent is found valid in court. We assume that \(\gamma\) is a random variable that follows a twice continuously differentiable CDF \(H(\gamma)\) with density \(h(\gamma)\). It is assumed that all patents have the same value and that the realizations of \(\gamma\) are common knowledge.

A generic company \((G)\) acts as a potential entrant in the market by considering a challenge to a patent at \(t = 0\) with challenge costs \(f_g > 0\).\(^5\) The challenge costs include the costs of preparing the challenge as well as firm investments for entering the market: technology and marketing investments to overcome market entry barriers.\(^6\) Potential litigation costs are assumed to be equal to zero. In the event of a challenge, the outcome depends on the court’s decision: If the patent is declared invalid, \(G\) enters at \(t = 0\); if the patent is held valid, market entry occurs at \(t = 1\).

---

5 In our model we assume that only one entrant exists. This implies that we do not take into account the ‘multiple challenger problem’: the public good problem that emerges if several generic entrants could simultaneously challenge a patent, or any strategic interaction between the generics.

6 Note that we assume that any market entry prior to the expiration of the patent implies a challenge.
This implies that $G$’s expected entry date under litigation is

$$t_{\text{lit}}^g = \gamma \cdot 1 + (1 - \gamma) \cdot 0 = \gamma,$$

so that the firms’ expected litigation profits are given by

$$\pi^o|_g = t_{\text{lit}}^g \pi^m + \left(1 - t_{\text{lit}}^g\right) \pi^d|_g$$

and

$$\pi^g|_g = \left(1 - t_{\text{lit}}^g\right) \pi^d|_g = (1 - \gamma) \pi^d|_g,$$

where $\pi^m$ describes the originator’s monopoly profit, while $\pi^d|_g$ and $\pi^d|_g$ denote the firms’ profits after $G$’s market entry. Hence, the generic challenges all patents for which

$$\pi^g|_g(\gamma) - f_g = (1 - \gamma) \pi^d|_g - f_g \geq 0.$$

If $G$ challenges a patent, both parties can settle their patent dispute through a patent agreement on a specific (potentially delayed) generic entry date—$t_{\text{set}}^g \in [0, 1]$—and a reverse payment—$r_{\text{set}}^g \geq 0$—from the originator to the generic. We assume that there are no settlement costs and that $t_{\text{set}}^g$ as well as $r_{\text{set}}^g$ are endogenously determined during the settlement negotiations.

If the firms agree on a settlement, the originator can reap monopoly profits $\pi^m$ for period $[0, t_{\text{set}}^g)$, whereas for $[t_{\text{set}}^g, 1]$, the originator and generic generate duopoly profits $\pi^d|_o + \pi^d|_g = \pi^d$. After patent expiration, all profits are assumed to be equal to zero because of perfect competition: $\pi^c = 0$. Consumer welfare under the different market structures is denoted by $CW^m$, $CW^d$, and $CW^c$. We assume that $\pi^m > \pi^d > \pi^c$ and that $CW^m < CW^d < CW^c$, which corresponds to standard conditions under competitive markets. Fig. 1 summarizes this situation.

The firms’ profits under a settlement are given by

$$\pi^{\text{set}}|_o = t_{\text{set}}^g \pi^m + \left(1 - t_{\text{set}}^g\right) \pi^d|_o - r_{\text{set}}^g$$

and

$$\pi^{\text{set}}|_g = \left(1 - t_{\text{set}}^g\right) \pi^d|_g + r_{\text{set}}^g,$$

which corresponds to a joint settlement profit of

$$\Pi^{\text{set}} = t_{\text{set}}^g \pi^m + \left(1 - t_{\text{set}}^g\right) (\pi^d|_o + \pi^d|_g).$$

(1)

If we assume that market entry is delayed beyond the date under the litigation outcome—that $t_{\text{set}}^g > t_{\text{lit}}^g$—the joint settlement profit exceeds its litigation counterpart:

$$\Pi^{\text{lit}} = t_{\text{lit}}^g \pi^m + \left(1 - t_{\text{lit}}^g\right) (\pi^d|_o + \pi^d|_g).$$

In particular, this ‘settlement surplus’—the gain from settling instead of litigating—is given by

$$s = \Pi^{\text{set}} - \Pi^{\text{lit}} = \left(t_{\text{set}}^g - t_{\text{lit}}^g\right) \left(\pi^m - \left(\pi^d|_o + \pi^d|_g\right)\right) > 0.$$

(2)

The delayed market entry reduces the generic’s profit by $\left(t_{\text{set}}^g - t_{\text{lit}}^g\right) \pi^d|_g$. Hence, we assume that the generic demands compensation for this loss. In addition, we
explicitly assume that the firms equally share the settlement surplus. The reverse payment from $O$ to $G$ is therefore given by

$$r_{set}^O = \frac{t_{set}^O - t_{lit}^O}{2} \pi_g^d + s/2,$$

so that individual settlement profits become

$$\pi_{set}^O = t_{set}^O \pi_m + \left(1 - t_{set}^O\right) \pi_o^d - \left(t_{set}^O - t_{lit}^O\right) \pi_g^d + s/2 = \pi_{lit}^O + s/2 \tag{3}$$

$$\pi_{set}^g = \left(1 - t_{set}^g\right) \pi_g^d + \left(t_{set}^g - t_{lit}^g\right) \pi_g^d + s/2 = \pi_{lit}^g + s/2 \tag{4}$$

Given Equation (4), we know that the generic challenges all patents for which we have

$$\left(1 - t_{set}^g\right) \pi_g^d + \left(t_{set}^g - t_{lit}^g\right) \pi_g^d + s/2 - f_g \geq 0 \tag{5}$$

Note that in a litigation-replicating settlement, the firms would avoid any entry delay compared to the entry date under litigation: They would agree on $t_{set}^g = t_{lit}^g$. On the other hand, under the most collusive settlement, they would maximize the settlement surplus by deferring market entry until the end of the original patent duration; that is, we would have $t_{set}^g = 1$. This would be the natural outcome if firms were allowed to settle without any legal restrictions. According to the criterion of Shapiro, competition authorities would prohibit all patent settlements through which consumers are harmed by generic entries that are later than the date expected under litigation. This implies that the firms are not allowed to specify entry dates beyond $t_{lit}^g = \gamma$.

However, deviating from Gratz (2012), we introduce a policy parameter $\tilde{t}$ that explicitly allows competition authorities to grant the parties an additional time period for collusion. In particular, $O$ and $G$ can agree to share monopoly profits until $t_{set}^g = t_{lit}^g + \tilde{t} = \gamma + \tilde{t}$, which implies that the firms can choose any $t_{set}^g \leq t_{lit}^g + \tilde{t}$. We assume that the competition authority has to commit to $\tilde{t}$ before $\gamma$ is realized. As the original remaining patent duration is equal to one, $\tilde{t}$ can be interpreted as a percentage share of the original remaining patent duration that is additionally granted for collusion. We illustrate this in Fig. 2.

---

7 If the bargaining process results in a smaller share of the settlement surplus’s being given to the generic, the challenging incentives would be negatively affected. However, the corresponding impact would only quantitatively change our results. Hence, the implications of our model would not change.
As long as we strictly interpret $\bar{t}$ as being an additional time period of collusion, we have to restrict $\bar{t}$ to values between zero and one because we cannot have negative amounts of collusion. Later, however, we will take a broader analytical perspective that allows for a more general interpretation of $\bar{t}$: As $\bar{t} = 0$ clearly corresponds to the litigation benchmark, we know that $\bar{t} > 0$ generally represents all cases in which we have generic entry delays beyond the expected entry date under litigation. For the entrant, the effect of a positive policy parameter can be compared to the instrument of directly subsidizing the patent challenge so that more patents are challenged as a result. The difference from our solution would be that a subsidy does not lead to collusion between the generic entrant and the originator firm and the corresponding generic entry delay.

In our model, negative values of $\bar{t}$ also have a meaningful interpretation: $\bar{t} < 0$ generally refers to all cases in which generic firms would be forced to (hypothetically) enter the market prior to the expected entry date under litigation. This refers to situations where many patents are already challenged naturally—challenge costs are sufficiently low—and forcing earlier entry would be more beneficial than encouraging even more patent challenges. Using this interpretation, we will describe some important policy implications in Sect. 5.

Overall, our model setup corresponds to a sequential game with the following timing of actions: (1) The competition authority commits ex ante to policy parameter $\bar{t}$ without knowing the realization of $\gamma$; (2) Nature draws $\gamma$, and its realization becomes common knowledge; (3) The generic decides whether to challenge at $t = 0$; (4) If a patent is challenged, the generic makes a settlement offer that contains $rp_{\text{set}}$ and $t_{\text{set}} \leq \gamma + \bar{t}$; and (5) The originator accepts or rejects the offer. If $O$ rejects the offer, the litigation outcome is realized.8

### 4 Equilibrium Analysis: The Optimal Policy Parameter

Since competition authorities do not challenge settlements with $t_{\text{set}} \leq t_{\text{lit}} + \bar{t}$, $G$ optimally offers $O$ to choose the corner solution, that is, to choose $t_{\text{set}} = t_{\text{lit}} + \bar{t} = \gamma + \bar{t}$ in equilibrium. This maximizes the settlement surplus, which is then equally shared between the firms. As we know from Equation (3) that the originator’s settlement profit is higher than its litigation counterpart, the originator accepts the offer. This implies that all cases are settled, so the settlement subgame immediately ends, and litigation never occurs. Our results are to some extent in line with the findings of Priest and Klein (1984), where settlement equilibria result from savings in litigation costs. However, the mechanism in our model is different, as the incentives to settle exclusively result from the allowed market entry delay.

---

8  With respect to the treatment of time in our model, we stick to the assumption of Gratz (2012) that a patent dispute is triggered and resolved at the same point in time. This is in contrast to, e.g., Shapiro (2003) and Elhauge and Krueger (2012), who assume that there is a litigation period. Introducing a litigation period would postpone the firms’ entry dates accordingly. This, however, would have no qualitative impact on our results.
Using (1) and \( r_g^\text{set} = \gamma + \tilde{t} \), we know that the joint settlement profit under \( \tilde{t} \) is described by

\[
\Pi^\text{set} = (\gamma + \tilde{t})\pi^m + (1 - (\gamma + \tilde{t}))\left(\pi^d_o + \pi^d_g\right)
\]

(6)

This holds for all challenged patents: for \( \gamma \in \left[0, \gamma^\text{set}_g\right] \), where \( \gamma^\text{set}_g \) denotes the critical level of patent strength at which \( G \) is indifferent between challenging a patent or not. Following (2), the generated surplus compared to litigation is therefore given by

\[
s = \tilde{t}\left(\pi^m - \left(\pi^d_o + \pi^d_g\right)\right)
\]

(7)

which allows us to determine the critical level of patent strength.

As we know that \( r_g^\text{set} = \gamma + \tilde{t} \), we can show that Condition (5) becomes

\[
(1 - \gamma)\pi^d_g + s/2 - f_g \geq 0.
\]

Hence, we find that \( \gamma^\text{set}_g \) is determined by

\[
(1 - \gamma)\pi^d_g + s/2 - f_g = 0 \quad \Leftrightarrow \quad \gamma^\text{set}_g = 1 - \frac{f_g - s/2}{\pi^d_g}
\]

(8)

Note that for \( \tilde{t} > 0 \), we have that \( G \)’s net expected profit from litigation is smaller than zero, which may raise concerns about the credibility of the firm’s incentive to challenge patents. However, we can easily show that due to the settlement surplus, the originator’s settlement profit strictly exceeds the expected profit from litigation. Following our assumption that negotiations between the parties take place before trial, we know that even for \( \tilde{t} > 0 \), generic challenges constitute a credible strategy because the generic anticipates that the originator will in any case agree to a settlement once a patent has been challenged.

Following Equation (8), we can conclude that \( G \) challenges patents for \( \gamma \in \left[0, \gamma^\text{set}_g\right] \), which leads to the following market structures: If \( G \) does not challenge any patents—for \( \gamma \in \left(\gamma^\text{set}_g, 1\right] \)—the originator’s monopoly covers the entire remaining patent duration. For \( \gamma \in \left[0, \gamma^\text{set}_g\right] \), \( G \) enters the market, so the originator holds a monopoly for \( t^\text{set}_g - 0 = \gamma + \tilde{t} \). Then, \( G \) enters at \( r^\text{set}_g \), creating a duopoly for the time period \( 1 - t^\text{set}_g = 1 - \gamma - \tilde{t} \).

Following our assumption that the policy-maker has to commit to a policy without being informed about \( \gamma \), we know that consumer welfare under \( \tilde{t} \) is described by

\[
\text{CW}^\text{set}(\tilde{t}) = \int_0^{\gamma^\text{set}_g(\tilde{t})} \left[(\gamma + \tilde{t})\text{CW}^m + (1 - \gamma - \tilde{t})\text{CW}^d\right] h(\gamma) d\gamma + \int_{\gamma^\text{set}_g(\tilde{t})}^1 \text{CW}^m h(\gamma) d\gamma
\]

\[
= \left(1 - \gamma^\text{set}_g(\tilde{t}) - \tilde{t}\right)H\left(\gamma^\text{set}_g(\tilde{t})\right) + \int_0^{\gamma^\text{set}_g(\tilde{t})} H(\gamma) d\gamma \left(\text{CW}^d - \text{CW}^m\right) + \text{CW}^m
\]

(9)
Maximizing Equation (9) with respect to \( \tilde{t} \) yields

\[
\frac{\partial CW^{set}}{\partial \tilde{t}} = \frac{\partial y_g^{set}}{\partial \tilde{t}} \left( 1 - y_g^{set} - \tilde{t} \right) (CW^d - CW^m) h(y_g^{set}) - H(y_g^{set}) (CW^d - CW^m)
\]

Incentive Effect

Entry Delay Effect

(10)

The interpretation of (10) is straightforward: If competition authorities are more generous with patent settlements, the joint settlement profit increases, and hence, the generic company challenges more patents. This is reflected by the incentive effect, which captures the positive externality that arises from \( G \)’s market entry because \( G \) ignores the consumer surplus benefits in its private entry calculation. In particular, \( G \) additionally challenges \( \frac{\partial y_g^{set}}{\partial \tilde{t}} \) patents, for which we have an increase in consumer welfare of \( CW^d - CW^m \) for the period after market entry under the marginal settlement: for \( 1 - \tau_g^{set} = 1 - y_g^{set} - \tilde{t} \). This effect is more pronounced if the realization of \( y_g^{set} \) is more likely to occur: if \( h(y_g^{set}) \) takes higher values.

However, due to \( G \)’s delayed market entry, consumer welfare decreases by \( CW^d - CW^m \) for all of the patents that would have been challenged anyway. Here, \( H(y_g^{set}) \) indicates the probability that \( y \in \left[ 0, y_g^{set} \right] \). This entry delay effect basically reflects a negative externality that accompanies each settlement: When settling a case, \( O \) and \( G \) ignore the negative impact on consumer welfare that results from \( G \)'s delayed entry. Instead, in equilibrium, they fully exploit the allowed time period of collusion as this maximizes their private joint settlement profit.

Thus, the optimal policy \( \tilde{t}_{opt} \) is implicitly determined by the point where these two opposing effects marginally compensate for each other: \( \partial CW^{set} / \partial \tilde{t} = 0 \). If we consider a specific example where \( \gamma \sim U[0, 1] \), we can show that the optimal policy is explicitly given by

\[
\tilde{t}_{opt} = \frac{4 \left( \pi^d_g \right)^2 - 2f_g \left( \pi^m + \pi^d_g - \pi^d_o \right)}{\left( \pi^m + 3\pi^d_g - \pi^d_o \right) \left( -\pi^m + \pi^d_g + \pi^d_o \right)}
\]

(11)

Note that in our model the optimal policy is independent of consumer welfare. To explain this result, we need to recall that a marginal change in \( \tilde{t} \) induces incentive and entry delay effects. In particular, we can see from (10) that additional challenges increase consumer welfare by \( CW^d - CW^m \). However, these additional challenges result from the allowed market entry delay, and any generic entry delay affects consumer welfare by \( CW^m - CW^d = -(CW^d - CW^m) \). Hence, in terms of consumer welfare, these proportional effects have the same magnitude with opposite signs.
Therefore, \( CW^d - CW^m \) equally impacts the incentive and entry delay effects, so it cancels out in the optimum. 9

With respect to the second-order condition, we find that \( \partial^2 (CW^{set})/\partial (\tilde{t})^2 < 0 \) if

\[
\begin{align*}
&h'(\gamma_g^{set}) < \left( 2 + \partial \gamma_g^{set} / \partial \tilde{t} \right) \left( h(\gamma_g^{set}) \right)^2 / H(\gamma_g^{set}).
\end{align*}
\]

Hence, we can conclude that \( \tilde{t}_{opt} \) is a maximum solution if \( H(\gamma) \) is not too convex. 10

5 Welfare Analysis

To show that the implementation of \( \tilde{t}_{opt} \) is beneficial for consumers, we have to compare our results to the benchmark case of the litigation solution. Using Equation (9), we can compute consumer surplus under litigation by evaluating \( CW^{set}(\tilde{t}) \) at \( \tilde{t} = 0 \). Hence, consumer welfare under litigation is given by \( CW^{set}(0) \), whereas consumer welfare under the optimal policy is determined by \( CW^{set}(\tilde{t}_{opt}) \).

Our results are summarized in Proposition 1:

**Proposition 1:** The implementation of the optimal policy weakly increases consumer welfare: \( CW^{set}(\tilde{t}_{opt}) \geq CW^{set}(0) \). If challenge costs are sufficiently high—if \( f_g \geq f_g^* = H(\gamma_g^{set}) / h(\gamma_g^{set}) \cdot 2 (\pi^d - \pi^m - \pi^g) > 0 \)—then \( \tilde{t}_{opt} \geq 0 \). Otherwise, \( \tilde{t}_{opt} < 0 \).

**Proof:** See Appendix 1.

Proposition 1 shows that the implementation of \( \tilde{t}_{opt} \) is indeed welfare-increasing for consumers. Except for the case of \( \tilde{t}_{opt} = 0 \), this result holds in general: for \( \tilde{t}_{opt} > 0 \) as well as for \( \tilde{t}_{opt} < 0 \). Note that our results require no additional assumptions with respect to the mode or the intensity of competition.

Given the paper’s focus, it is a very important result that \( \tilde{t}_{opt} \) can be positive or negative. As we can show that \( f_g > 0 \), we know that \( \tilde{t}_{opt} > 0 \) requires challenge costs to be strictly positive and sufficiently high. The main intuition behind this result is as follows: If challenge costs are small—if \( 0 < f_g < f_g^* \)—the generic challenges a large number of patents under litigation anyway, so additional monetary incentives in terms of \( \tilde{t}_{opt} > 0 \) are not necessary. If, however, challenge costs are sufficiently high, the generic’s challenging incentives are too small from a consumer welfare perspective. Hence, we need \( \tilde{t}_{opt} > 0 \) to induce more challenges, which, as we stated before, 9 However, this result is an artefact from our simple setting with one generic entrant. In an earlier working paper version, we considered a framework with two generic entrants. As a result, the welfare expressions did not cancel in the optimality condition.

10 We discuss the uniqueness of this solution in the proof of Proposition 2.
allows for the interpretation of $\tilde{t}_{opt} > 0$ as the equivalent of a subsidy on patent challenges.

Our finding that $\tilde{t}_{opt}$ can be positive or negative implies that we have to distinguish two cases. In Case 1 ($f_g < f_g$), consumer welfare would be maximized for $\tilde{t}_{opt} < 0$. Here, due to small challenging costs, the challenging incentives provided under the litigation solution are already relatively high, making $G$ challenge a correspondingly large number of patents without additional incentives. This leads to a situation in which the negative externality in terms of the entry delay effect dominates the marginal benefits from the incentive effect for all $\tilde{t} \geq 0$. Therefore, inducing additional challenges by allowing generic entry delay beyond the litigation benchmark in the settlements—$\tilde{t} > 0$—would not be beneficial for consumers and clearly be anticompetitive.

Instead, consumer welfare could be increased by (hypothetically) enforcing earlier generic entry (that is, an entry date prior to the entry date under litigation) in the settlements, which would require $\tilde{t}_{opt} < 0$. Recall that, as mentioned above, in this case the entry delay effect dominates the incentive effect for all $\tilde{t} \geq 0$. By introducing $\tilde{t}_{opt} < 0$, we would, on the margin, create a stronger incentive effect because consumers would benefit for a longer period of time from generic entry resulting from additional challenges. At the same time—and with the assumption that all cases would still be settled—the resulting reduction in the settlement profits would make challenges less attractive, thereby mitigating the entry delay effect. However, as long as the law does not limit the parties’ rights to litigate, competition authorities cannot enforce earlier market entry than the date under litigation, since firms would no longer settle a case but instead prefer to litigate. Hence, the case of $\tilde{t}_{opt} < 0$ remains a hypothetical consideration because it cannot be implemented as an equilibrium solution.

In Case 2 ($f_g > f_g$), the situation is entirely different: If $\tilde{t}_{opt} > 0$, challenging incentives under litigation are inefficiently small, which creates a situation in which the incentive effect dominates the entry delay effect for some $\tilde{t} > 0$. Therefore, inducing additional patent challenges by allowing a longer period of collusion benefits consumers. Since generic entry delay is only possible if the generic obtains a share of the settlement surplus, competition authorities have to allow reverse payments to a certain extent. Hence, our analysis shows that allowing more collusion in patent settlements (with reverse payments) can be procompetitive. However, this holds only for Case 2: for $\tilde{t}_{opt} > 0$. In addition, note that market entry cannot be delayed beyond the original patent duration. This requires that $t_{set}^g \leq 1$, which holds in any case if $\tilde{t}_{opt} \leq \tilde{t} = 2f_g / (\pi^m + \pi^d - \pi^o)$.

To explain the intuition behind the results of Proposition 1, we first have to analyse the impact of $\tilde{t}_{opt}$ on the range of challenged patents. Since we focus on $\tilde{t}_{opt} > 0$, we know that the optimal policy allows more collusion: delayed market entry of the generic. Indeed, we can show that the market entry date $t_{set}^g = t_{lit}^g + \tilde{t}$ is extended beyond its counterpart under litigation: $t_{set}^g > t_{lit}^g$. Therefore, the settlement surplus is strictly positive, which induces the generic to challenge more patents. Hence, we have that $\gamma_{set}^g > \gamma_{lit}^g$, which allows us precisely to identify incentive and entry delay effects.
In particular, for all \( \gamma \in \left[ 0, \gamma^l_g \right] \), consumer welfare is negatively affected by the entry delay effect, as all of these patents would have been challenged anyway under litigation. However, at the same time, consumers benefit from the generic’s incentive effect: the additional patents that are now challenged under the optimal policy. These patents are characterized by \( \gamma \in \left( \gamma^l_g, \gamma^s_g \right] \). For the case of very strong patents—for \( \gamma \in \left( \gamma^s_g, 1 \right] \)—we do not observe any effect because these patents are not challenged under litigation nor under \( \tilde{t}_{opt} \).

Thus far, our analysis has focused on consumer welfare maximization. However, it could be argued that the policy-maker should aim instead to maximize social welfare. The social welfare function includes expected consumer welfare as well as the expected joint profits of the firms. Hence, social welfare is given by

\[
SW^s(\tilde{t}) = \int_0^{\gamma^s_g} [(\gamma + \hat{d})CW^m + (1 - \gamma - \hat{d})CW^d] h(\gamma) d\gamma + \int_{\gamma^s_g}^1 CW^m h(\gamma) d\gamma
\]

\[
+ \int_0^{\gamma^s_g} [\gamma \pi^m + (1 - \gamma) \pi^d_o + s/2] h(\gamma) d\gamma + \int_{\gamma^s_g}^1 \pi^m h(\gamma) d\gamma + \int_0^{\gamma^s_g} (1 - \gamma) \pi^d_g + s/2 - f_g \] h(\gamma) d\gamma,
\]

\[
= \left[ (1 - \gamma^s_g(\tilde{t}) - \tilde{t}) H(\gamma^s_g(\tilde{t})) + \int_0^{\gamma^s_g(\tilde{t})} H(\gamma) d\gamma \right] \varphi + CW^d + \pi^m + (s - f_g) H(\gamma)
\]

where \( \varphi = CW^d - CW^m + \pi^d_o + \pi^d_g - \pi^m \) denotes the difference between the gain in consumer welfare and the reduction in the firms’ joint profits that result from market entry. We assume that \( \varphi > 0 \).

Maximizing \( SW^s \) with respect to \( \tilde{t} \) yields

\[
\frac{\partial SW^s}{\partial \tilde{t}} = \frac{\partial \gamma^s_g}{\partial \tilde{t}} \left[ (1 - \gamma^s_g - \tilde{t}) \varphi - f_g \right] h(\gamma^s_g) - \varphi H(\gamma^s_g)
\]

(12)

which can be interpreted as follows: If \( \tilde{t} \) marginally increases, \( G \) additionally challenges \( \frac{\partial \gamma^s_g}{\partial \tilde{t}} \) patents, from which consumers benefit by \( CW^d - CW^m \) as a consequence of market entry; at the same time, the firms are hurt by the reduction in their joint profits: \( \pi^d_o + \pi^d_g - \pi^m \). In addition, the generic has to pay \( f_g \) for all additionally challenged patents. This describes the net incentive effect. On the other hand, a marginal increase in \( \tilde{t} \) allows more collusion, so consumers face a reduction in their welfare by \( CW^d - CW^m \) for all patents that would have been challenged anyway. However, here, the firms benefit from the higher settlement surplus, which increases by \( \pi^m - \pi^d_o - \pi^d_g \). This reflects the net entry delay effect.

By rewriting (12), we find that the optimal policy is now determined by

\[
\frac{\partial SW^s}{\partial \tilde{t}} = \frac{\partial \gamma^s_g}{\partial \tilde{t}} \left[ (1 - \gamma^s_g - \tilde{t}) \varphi - f_g \right] h(\gamma^s_g) - H(\gamma^s_g) = 0,
\]

whereas under consumer welfare maximization, we know from Equation (10) that

\[
\frac{\partial CW^s}{\partial \tilde{t}} = \frac{\partial \gamma^s_g}{\partial \tilde{t}} (1 - \gamma^s_g - \tilde{t}) h(\gamma^s_g) - H(\gamma^s_g) = 0.
\]
A comparison of these two optimality conditions shows that both equations differ only in the expression $-f_g/\varphi < 0$. This expression represents the negative impact from the challenge costs, which constitute a welfare loss from an overall welfare perspective. In addition, note that $-f_g/\varphi$ still includes welfare and profit expressions, so the socially optimal policy depends on consumer surplus benefits, which is in contrast to our findings from the first part of our analysis but in line with economic intuition. In particular, larger welfare benefits for consumers would reduce $-f_g/\varphi$ in absolute terms, which thereby increases $\tilde{t}_{opt}$.

As the optimal policy now accounts for the negative impact of challenge costs, the policy parameter takes lower values under social welfare maximization. However, apart from this quantitative impact, the implications of our model do not change. Therefore, we focus on consumer welfare maximization for the remainder of this paper.

6 Comparative Statics

In this section, we analyse the comparative statics of the optimal policy parameter $\tilde{t}_{opt}$ under consumer welfare maximization.\textsuperscript{11} We will ask what the effects on the size of the optimal policy parameter are if a) the quality of patents (operationalized as a change in the distribution of patent strength) changes, b) the challenge costs change, and c) the generic’s profit under duopoly increases.\textsuperscript{12} For steps b) and c), we use the specific example where $\gamma \sim U[0, 1]$.

6.1 The Effects of a Change in the Distribution of Patent Strength

It is an important objective of many patent offices to increase the quality of the granted patents. Since the quality of patents can be operationalized as the probability of a patent not being invalidated in patent litigation, the distribution of the strength of the patents in our model is a good proxy for the quality of the patents granted by a patent office. Successes of patent offices in increasing their patent quality would therefore translate into a systematic change of the distribution function. Would such a change in patent quality lead to a higher or lower optimal policy parameter for the permitted collusion period in patent settlements?

Let $H_1(\gamma)$ and $H_2(\gamma)$ be two different cumulative distribution functions of the patent strength $\gamma$. Then, we can analyse the impact of a shift from $H_1(\gamma)$ to $H_2(\gamma)$ on the optimal policy parameter.

Our findings are presented in Proposition 2.

\textsuperscript{11} As we have found that Case 1, where the optimal policy parameter is smaller than zero, strictly results in litigation, we restrict our analysis to Case 2: the case of a positive optimal policy parameter.

\textsuperscript{12} Note that in principle, one could also analyse comparative static changes in the originator’s profit under monopoly as well as under duopoly. However, the corresponding results are ambiguous, and we feel that our analysis would not benefit from including them. Hence, we omit these results.
Proposition 2: The optimal policy weakly increases if \( H_2(\gamma) \) stochastically dominates \( H_1(\gamma) \) according to the reverse hazard rate order: if \( h_2(\gamma)/H_2(\gamma) \geq h_1(\gamma)/H_1(\gamma) \) for all \( \gamma \in [0, 1] \). If the opposite holds, the optimal policy weakly decreases.

Proof: See Appendix 1.

To understand the logic behind the results of Proposition 2, we have to recall from our model setup that the generic’s challenge decision takes place after \( \gamma \) has been drawn from the distribution of patent strength. Hence, the generic’s challenge calculus is not affected by changes in the distribution, which implies that the critical level of patent strength, \( \gamma_{g}^{\text{set}} \), remains unchanged if \( H(\gamma) \) shifts. Instead, a change in \( H(\gamma) \) affects only the ex ante rationale of the policy-maker when deciding on \( \tilde{t}_{\text{opt}} \) at stage 1 of our model.

Here, we specifically see from Eq. (10) that \( h\left(\gamma_{g}^{\text{set}}\right) \) and \( H\left(\gamma_{g}^{\text{set}}\right) \) have a crucial impact on the optimal policy. On the left-hand side of this equation, \( h\left(\gamma_{g}^{\text{set}}\right) \) indicates the likelihood of drawing a patent from the distribution that is among those patents that would be additionally challenged if \( \tilde{t} \) were marginally increased; on the right-hand side of (10), \( H\left(\gamma_{g}^{\text{set}}\right) \) determines the overall probability that \( \gamma \leq \gamma_{g}^{\text{set}} \). The patent strength that is drawn is one that would have led \( G \) to challenge anyway. Hence, if \( h\left(\gamma_{g}^{\text{set}}\right) \) takes higher values, it becomes more likely that a case for which the generic ignores the positive externality from market entry occurs. Therefore, ceteris paribus, \( \tilde{t} \) needs to be larger. However, for higher values of \( H\left(\gamma_{g}^{\text{set}}\right) \), the entry delay effect would become more relevant, which would result in \( \tilde{t} \) being smaller.

If we now consider a shift in the distribution, so that we have reverse hazard rate dominance (RHRD) of \( H_2(\gamma) \) over \( H_1(\gamma) \), we immediately know that RHRD implies first-order dominance: \( H_2(\gamma) \leq H_1(\gamma) \) for all \( \gamma \in [0, 1] \) and therefore \( H_2\left(\gamma_{g}^{\text{set}}\right) \leq H_1\left(\gamma_{g}^{\text{set}}\right) \). Hence, patent validity must have increased: The average patent has become more valid. This reduces, ceteris paribus, the entry delay effect, which positively affects the optimal policy as it decreases the marginal costs of more collusion.

On the other hand, we could have \( h_2\left(\gamma_{g}^{\text{set}}\right) \geq h_1\left(\gamma_{g}^{\text{set}}\right) \) or \( h_2\left(\gamma_{g}^{\text{set}}\right) < h_1\left(\gamma_{g}^{\text{set}}\right) \): The realization of an additionally challenged patent may have become more or less likely under \( H_2(\gamma) \). However, as we know that \( H_2\left(\gamma_{g}^{\text{set}}\right) \leq H_1\left(\gamma_{g}^{\text{set}}\right) \), RHRD requires that \( h_2\left(\gamma_{g}^{\text{set}}\right) \) has either increased or at least decreased less than \( h_1\left(\gamma_{g}^{\text{set}}\right) \). Hence, even if the incentive effect becomes less beneficial because of a decreasing \( h_2\left(\gamma_{g}^{\text{set}}\right) \), this decrease is smaller than the benefit from the reduction of the entry delay effect. Therefore, \( \tilde{t}_{\text{opt}} \) increases.
6.2 The Effects of an Increase in Challenge Costs

As we already know that challenge costs have a crucial impact on the optimal policy, we now want to examine how $\tilde{t}_{opt}$ changes in $f_g$. The corresponding result is given in Proposition 3:

**Proposition 3:** The optimal policy parameter $\tilde{t}_{opt}$ is strictly increasing in $f_g$. 

**Proof:** See Appendix 1.

The intuition behind Proposition 3 is as follows: A marginal increase in $f_g$ makes patent challenges more costly. Hence, ceteris paribus, the generic would challenge only relatively weaker patents: patents with a smaller $\gamma$. Therefore, we have that the critical level of patent strength $\gamma_{set}^g$ is strictly decreasing in $f_g$, which in turn has an impact on the incentive effect: Since $\gamma_{set}^g$ decreases, $t_{set}^g(\gamma_{set}^g) = \gamma_{set}^g + \tilde{t}$ decreases as well: Firms would, ceteris paribus, agree on earlier market entry under the marginal settlement. Hence, consumers benefit from $CW_d$ for a longer period of time: The positive externality from market entry increases. As this increase in the positive externality is ignored by the generic, $\tilde{t}$ needs to be larger to induce more challenges.

In contrast to the incentive effect, we find that the entry delay effect strictly decreases because the range of patents that would be challenged anyway, ceteris paribus, diminishes. This outcome is intuitive: The entry delay effect for consumer welfare is based on a larger extent of collusion for all $\gamma \in [0, \gamma_{set}^g]$. Since $\gamma_{set}^g$ decreases in $f_g$, the entry delay effect also decreases, making additional collusion less costly in terms of consumer welfare. Consequently, with respect to the marginal impact of $\tilde{t}$ on $CW_{set}$, a marginal increase in $f_g$ makes additional collusion more beneficial and less costly. Hence, we find that the optimal policy in terms of allowed collusion is strictly increasing in challenge costs.

6.3 The Effects of an Increase in the Generic’s Profit under Duopoly

In addition to $f_g$, the generic’s challenge decision is crucially determined by the profit after market entry. This profit may change, ceteris paribus, for several reasons: For instance, a profit shift may result from changes in the generic’s production costs or from changes in consumers’ perception of the substitutability between the generic’s and the originator’s product.

Proposition 4 summarizes the impact of a change in $\pi_d^g$ on the optimal policy:

**Proposition 4:** The optimal policy strictly decreases if the generic’s profit under duopoly marginally increases.

**Proof:** See Appendix 1.
Based on our explanation for Proposition 3, the argument behind Proposition 4’s result is intuitive: On the one hand, the increase in $\pi^d_g$ makes patent challenges, cet-eris paribus, more attractive. Hence, we can show that the generic initially challenges more patents: $\gamma^\text{set}_g$ increases if $\pi^d_g$ grows, which reduces the incentive effect. On the other hand, the entry delay effect from allowing more collusion is increasing because the range of initially challenged patents: $\gamma^\text{set}_g$ has increased. As a result, we find that at the margin additional collusion becomes less beneficial and costlier, so that $\tilde{t}_{\text{opt}}$strictly decreases.

The results of our comparative static analysis are summarized in Table 1:

### 7 The Policy Parameter’s Impact on Innovation Incentives

Thus far, our analysis has exclusively focused on the trade-off between price effects and challenging incentive effects. However, another important question is the extent to which the implementation of $\tilde{t}_{\text{opt}}$ has an impact on the originator’s innovation incentives. While we do not model optimal innovation incentives, our model still allows for an analysis of the originator firms’ profits. We assume that a higher profit for the originator leads to higher innovation incentives. Our approach is to use the example where $\gamma \sim U[0, 1]$ and to compare the originator’s profit under $\tilde{t}_{\text{opt}}$ to the corresponding profit under litigation, which is given by $\Pi^\text{set}_o(0)$. This allows us to determine the changes in innovation incentives.

With respect to the critical level of patent strength, we can derive the originator’s profit in the case of collusion, which is described by

---

13 An alternative argument would be that higher profits conditioned on having a weak patent would imply that firms have greater incentives to develop weak patents. Following this argument, our analysis would show that the implementation of the optimal policy would reduce these incentives to invest in weak patents.
\[ \Pi^\text{set}_o = \begin{cases} \gamma \pi^m + (1 - \gamma) \pi^d_o + s/2 & \text{for } \gamma \in \left[0, \gamma^\text{set}_g\right], \\ \pi^m & \text{for } \gamma \in \left(\gamma^\text{set}_g, 1\right]. \end{cases} \] (13)

At the point in time when the originator decides on innovation-related investments, the patent strength, and hence the realization of \( \gamma \), are unknown. Therefore, we have to determine the originator’s ex ante expected profit. Given Eq. (13), this profit can be expressed as.

To determine the impact of the policy parameter on the originator’s profit, we take the derivative of \( \Pi^\text{set}_o(\tilde{t}) \) with respect to \( \tilde{t} \), which yields

\[ \frac{\partial \Pi^\text{set}_o(\tilde{t})}{\partial \tilde{t}} = \left( \pi^d_o - \pi^m \right) \left( 1 - \gamma^\text{set}_g \right) \frac{\partial \gamma^\text{set}_g}{\partial \tilde{t}} + \frac{1}{2} \frac{\partial s}{\partial \tilde{t}} \gamma^\text{set}_g + \frac{1}{2} \frac{\partial \gamma^\text{set}_g}{\partial \tilde{t}}. \]

Interpreting this derivative shows that a marginal increase in \( \tilde{t} \) basically affects \( \Pi^\text{set}_o(\tilde{t}) \) in two different ways: First, the incentive effect leads to \( \frac{\partial \Pi^\text{set}_o(\tilde{t})}{\partial \tilde{t}} \) additionally challenged patents for which the originator realizes only duopoly profits instead of monopoly profits for the time period after market entry. Second, however, this effect is compensated for to some extent by a stronger entry delay effect because a marginal increase in \( \tilde{t} \) leads to an increase in the surplus and hence to stronger benefits for the originator resulting from more collusion. As these two effects have opposite signs, the overall impact on \( \Pi^\text{set}_o(\tilde{t}) \) is ambiguous and depends on the value of \( \tilde{t} \).

As we want to inquire about the impact of the optimal policy parameter on the originator’s innovation incentives, we assume in the following that \( \tilde{t}^\text{opt} > 0 \) has been implemented.

We summarize our findings in Proposition 5:

**Proposition 5:** For \( f_g > 0 \) and \( \tilde{t}^\text{opt} \in [0, \tilde{t}] \), it strictly holds that \( \Pi^\text{set}_o(0) > \Pi^\text{set}_o(\tilde{t}^\text{opt}) \).

**Proof:** See Appendix 1.

Proposition 5 implies that the originator’s expected profit under the optimal policy parameter strictly falls below the benchmark level under litigation. Since this profit is supposed to reflect indirectly the level of the firm’s innovation incentives, we conclude that these incentives are always negatively affected by the implementation of \( \tilde{t}^\text{opt} > 0 \). However, as our model does not account for the social benefits.
of innovation, we do not know anything about the optimality of the initial innovation incentives. Hence, in terms of the trade-off between the social gains from marginal innovation versus the welfare benefits for consumers, we cannot determine whether this reduction is detrimental, neutral, or beneficial from an overall welfare perspective.\textsuperscript{14}

Nevertheless, our model allows for a comparison of the reduction in the originator’s profit and the gain in consumer welfare from the implementation of the optimal degree of collusion. To analyse this relationship, we have to determine the gain in consumer welfare, which is expressed by \( CW_{set}(\tilde{t}_{opt}) - CW_{set}(0) > 0 \), as well as the absolute value of the reduction in the originator’s expected profit, which is given by \( \Pi_o^{set}(0) - \Pi_o^{set}(\tilde{t}_{opt}) > 0 \).

We compare these two expressions and summarize the corresponding results in Proposition 6:

Proposition 6: For \( \pi^m - (\pi_o^d + \pi_g^d) > 0, \pi^m, \pi_o^d, \pi_g^d > 0 \) and \( \tilde{t}_{opt} \in [0, \tilde{t}] \), it holds that \( CW_{set}(\tilde{t}_{opt}) - CW_{set}(0) > \Pi_o^{set}(0) - \Pi_o^{set}(\tilde{t}_{opt}) \) if \( CW_d > CW^{d*} \).

Proof: See Appendix 1.

Proposition 6 contains an important theoretical implication: For specific parameter constellations—in particular for those cases where duopolistic competition after market entry is sufficiently strong—the gain in consumer welfare that results from the implementation of \( \tilde{t}_{opt} > 0 \) strictly exceeds the reduction in the originator’s expected profit. This allows us to conclude that, in principle, it would be possible to re-establish the initial level of the originator’s innovation incentives by redistributing part of the consumer welfare gain to the originator firm. This means that the decline in innovation incentives is more than compensated for by the increase in consumer welfare that results from more patent challenges and lower prices. In such a case, consumers are still strictly better off, while the originator is at least indifferent between the litigation benchmark and the implementation of the optimal policy parameter.

8 Discussion and Conclusion

In our model, we have analysed the trade-off between entry delay and incentive effects in patent settlements between originators and generics. The entry delay effect can be interpreted as a negative externality from later generic market entry, whereas the incentive effect reflects a positive externality that results from additional challenged patents. By introducing a policy parameter, we explicitly allow a longer period of collusion in antitrust law, which can have a positive impact on consumer welfare. Our key result is that under very general

\textsuperscript{14} This trade-off has, for instance, been analysed in the seminal work of Nordhaus (1969).
conditions, there exists an optimal specification of the policy parameter—which can be positive or negative—where consumer welfare is higher than its level under the benchmark case of litigation.

A positive optimal policy parameter means that challenging incentives are inefficiently small and can be compensated for by allowing later generic market entry and a longer collusion period (see Case 2 in Sect. 5 where the challenge costs are sufficiently high). This provides an additional reason why some patent settlements with reverse payments might not harm consumers and would therefore not be anticompetitive.

A negative optimal policy parameter implies that the settlement-related date of generic market entry is inefficiently late and consumer welfare can be increased only by stipulating earlier generic market entry than would be expected under litigation (see Case 1 in Sect. 5 where the challenge costs are sufficiently low). In this case, the interplay of incentive and entry delay effect leads to a situation where forcing earlier market entry—and thereby discouraging some patent challenges—proves more efficient than encouraging additional patent challenges by allowing later generic entry. As in reality and in our model, parties can always choose to litigate, this result is purely theoretical. However, it clearly shows that taking into account challenge incentives can lead to the conclusion that patent settlements without reverse payments can also harm consumers and be anticompetitive.\(^{15}\)

We additionally show—when the optimal policy parameter is positive—that the parameter increases with the size of challenge costs and when patents are on average more valid, and decreases with an increase in the generic’s profit under duopoly.

Our model also allows us to analyse how the innovation incentives of an originator are affected by the optimal policy parameter. By analysing the originator’s profit in the case of \(t_{\text{opt}} > 0\) and by using the example of the uniform distribution, we show that this profit and therefore the originator’s innovation incentives decline. Our result is a previously undescribed counterargument for using innovation incentives as a defence for longer collusion. We can also show for specific parameter constellations that this reduction in the originator’s expected profit is more than compensated for by the gain in consumer welfare that results from \(t_{\text{opt}} > 0\). Under these assumptions, it would therefore be possible to maintain the level of innovation incentives for the originator—by redistributing rents to the originator—while making consumers strictly better off.

What conclusions can be drawn from our results, and what limitations do we see? We primarily provide an integrated analysis of the trade-off between price effects and challenging incentives but analyse, to some extent, the effects on innovation incentives. However, a comprehensive model that simultaneously analyses all three groups of effects on consumer welfare and derives optimal solutions for antitrust limits on patent settlements is still missing. In addition, we do not model

\(^{15}\) Schankerman and Schuett (2017, p. 19) similarly argue that there can be excessive patent challenges from a consumer welfare perspective since private gains from challenging patents might exceed social gains or challenging might even be wasteful in the case of valid patents.
the multiple challenger problem: There might be several generic entrants that simultaneously decide to challenge patents, with the ensuing public good problem.\textsuperscript{16} To date, only a few models include the phenomenon of multiple generic entries in the economic analysis of the antitrust treatment of patent settlements.\textsuperscript{17}

Another question is whether the results of our model on the trade-off between price and challenging incentive effects also hold under less restrictive assumptions. For example, we assumed that: (i) all patents have the same value; (ii) the true patent strength is common knowledge; and (iii) the challenge costs are the same for all patents. Under more relaxed assumptions, our specific results and the optimal policy parameter would certainly change, but it is not clear why this should change our main results on the existence of an optimal policy parameter and the possibility of a trade-off, which leads to the conclusion that a longer period of collusion might be justified. For the case of sequential entry of two generics and—presumably—more than two entrants, we do not expect our results to change qualitatively as long as a similar structure of sequential entry decisions is assumed.

We show that implications from our model are robust to the inclusion of litigation costs (see Appendix 2). Whereas under litigation, challenging incentives would be smaller due to litigation costs, under settlement litigation costs are saved. In our model, this leads to smaller reverse payments since the outside option (to litigate) becomes less attractive for the entrant. Correspondingly, challenging incentives do not change. We also show that our results are robust to the maximization of social welfare instead of consumer welfare for determining the optimal policy (see Sect. 5). In this case, the optimal policy parameter takes lower values, but all implications of our model remain unchanged.

What conclusions can be drawn for the initial question of how to deal with the weak patent problem? For competition policy, it is a clear result from our model that the consideration of challenging incentives can help to optimize the antitrust assessment of patent settlements. We also provide further understanding of the interdependencies of all channels of patent settlement effects (price, innovation, and patent-challenge effects; see also Frank & Kerber, 2018).

It is still necessary to search for other solutions as part of the general problem of the optimal patent system’s design. Important proposals for improving patent examination procedures in patent offices or by private parties, facilitating and strengthening patent opposition and patent litigation procedures—by joint challenges, and promoting subsidization of patent challenges—can be found in Thomas (2001), Miller (2004), Farrell and Shapiro (2008), and Encaoua and Lefouili (2009). We see from

\textsuperscript{16} In the United States, the Hatch–Waxman Act grants the first successful generic challenger 180 days of marketing exclusivity before other entrants are allowed to enter the market (FTC 2011, p. 138; Hemphill 2006). This regulation addresses the incentive problems that result from the public good problem.

\textsuperscript{17} Some authors have discussed scenarios of multiple generic entries in specific frameworks. However, they have not analysed the public good problem (Edlin et al. 2015; Kobayashi et al. 2015). Padilla and Meunier (2015) analyse a scenario with multiple entrants that includes one credible litigation threat. Palikot and Pietola (2019) analyse settlement externalities in a setting with multiple potential entrants and strategic considerations of the patent holder (entry delay of generics, licensing, and litigation).
our model that these different instruments can be intertwined from a consumer welfare perspective and hence should be part of an integrated balancing approach.

Appendix 1: Proofs

Proof of Proposition 1 The proof for the first part of Proposition 1 simply follows from the fact that $\tilde{t}_{opt}$ maximizes $CW_{set}(\tilde{t})$. In addition, we know that $\tilde{t}_{opt} \geq 0$ if

$$\frac{\partial CW_{set}}{\partial t} \bigg|_{t=0} = \frac{\partial y_g^{set}}{\partial t} \left(1 - y_g^{set}\right) (CW^d - CW^m) h\left(y_g^{set}\right) - H\left(y_g^{set}\right) (CW^d - CW^m) \geq 0,$$

which holds for $f_g \geq f_g = H\left(y_g^{set}\right)/h\left(y_g^{set}\right) \cdot 2\left(\pi^d_g\right)^2 / \left(\pi^m - \pi^d - \pi^d\right) > 0$. For $0 < f_g < f_g$, $\tilde{t}_{opt} < 0$

$$\frac{\partial CW_{set}}{\partial t} = \frac{\partial y_g^{set}}{\partial t} \left(1 - y_g^{set} - \tilde{t}\right) (CW^d - CW^m) h\left(y_g^{set}\right) - H\left(y_g^{set}\right) (CW^d - CW^m) = 0,$$

Proof of Proposition 2 We know from Equation (10) that the optimal policy is implicitly determined by which can be rewritten as

$$\frac{\partial y_g^{set}}{\partial t} \left(1 - y_g^{set} - \tilde{t}\right) = \frac{H\left(y_g^{set}(\tilde{t})\right)}{h\left(y_g^{set}(\tilde{t})\right)}$$

(14)

As $\frac{\partial y_g^{set}}{\partial t}$ does not depend on $\tilde{t}$, we can show that for the LHS of (14),

$$\frac{\partial LHS}{\partial t} = \begin{cases} \frac{\partial y_g^{set}}{\partial t} \left(-1 - \frac{\partial y_g^{set}}{\partial t}\right) < 0, \\ >0 \end{cases}$$

For the RHS of (14), we have

$$\frac{\partial RHS}{\partial t} = \frac{\partial y_g^{set}}{\partial t} \left(h\left(y_g^{set}(\tilde{t})\right)^2 - H\left(y_g^{set}(\tilde{t})\right) h'\left(y_g^{set}(\tilde{t})\right)\right) h\left(y_g^{set}(\tilde{t})\right)^2.$$
Now, it is easy to see from (14) that a shift from $H_1(\gamma)$ to $H_2(\gamma)$ affects only the RHS of this optimality condition, while the LHS remains unchanged. This allows us to conclude that $\bar{t}_{opt}$, weakly increases if the RHS takes weakly smaller values for any given $i$. Hence, $\bar{t}_{opt}$ weakly increases if $H_2(\gamma)/h_2(\gamma) \leq H_1(\gamma)/h_1(\gamma)$ for all $\gamma \in [0, 1]$. This condition is equivalent to $h_2(\gamma)/H_2(\gamma) \geq h_1(\gamma)/H_1(\gamma)$ for all $\gamma \in [0, 1]$, which proves our result. If $H_2(\gamma)/h_2(\gamma) \geq H_1(\gamma)/h_1(\gamma)$, we find the opposite.

**Proof of Proposition 3** Using Equation (11), we can show that for $\pi^m - (\pi^d + \pi^g) > 0$ and $\pi^m, \pi^d, \pi^g > 0$,

$$\frac{\partial \bar{t}_{opt}}{\partial f_g} = \frac{2(\pi^m + \pi^g - \pi^d)}{(\pi^m + 3\pi^d - \pi^d)(\pi^m - \pi^d - \pi^d)} > 0.$$

**Proof of Proposition 4** By using Equation (11), we find that

$$\frac{\partial \bar{t}_{opt}}{\partial \pi^d} = -\frac{3(f_g + \pi^d)}{(3\pi^d + \pi^m - \pi^d)^2} + \frac{1}{3\pi^d + \pi^m - \pi^d},$$

$$\frac{\partial \bar{t}_{opt}}{\partial \pi^g} = \frac{1}{\pi^d - \pi^m + \pi^d}.$$

Now, we can show that for $\pi^m - (\pi^d + \pi^g) > 0$ as well as for $\pi^m, \pi^d, \pi^g > 0$ and $\bar{t}_{opt} \in [0, 1]$, it always holds that $\partial \bar{t}_{opt}/\partial \pi^d < 0$.

**Proof of Proposition 5** We find that the firm’s profits are

$$\Pi_o^u(0) = 1/2 \left( \pi^m + (f_g)^2(\pi^m - \pi^d) \right)/\left( \pi^d + \pi^d \right)$$

and

$$\Pi_o^u(\bar{t}_{opt}) = 1/2 \left( 2f_g + \pi^m - 4\pi^d f_g + \pi^d \right)^2 / \left( 3\pi^d + \pi^m - \pi^d \right)^2 + \pi^d. \right).$$

Then, we can show that for $f_g > 0$ as well as for $\bar{t}_{opt} \in [0, 1]$, it never holds that $\Pi_o^u(0) \leq \Pi_o^u(\bar{t}_{opt})$.

**Proof of Proposition 6** The proof consists of two steps. First, we determine $CW_o^u(\bar{t}_{opt}) - CW_o^u(0) - \Pi_o^u(0) - \Pi_o^u(\bar{t}_{opt})$. Then, respecting all existence conditions, we show that $CW_o^u(\bar{t}_{opt}) - CW_o^u(0) > \Pi_o^u(0) - \Pi_o^u(\bar{t}_{opt})$ holds for $CW^d > CW^{d*}$. The expression for $CW^{d*}$ has been obtained in Mathematica and is omitted because of complexity.

**Appendix 2: Extension**

**Introducing Litigation Costs**

We now assume that the originator and generic have to pay litigation costs $l > 0$ if the patent challenge is resolved by a court decision. Hence, the firms’ litigation profits become $\pi_o^{lit} = \pi_o^{lit} \pi^m + (1 - \pi_o^{lit}) \pi_o^d - l = \gamma \pi^m + (1 - \gamma) \pi_o^d_1 - l$ and...
\( \pi_{g}^{\text{lit}} = (1 - t_{g}^{\text{lit}}) \pi_{g}^{d} - l = (1 - \gamma) \pi_{g}^{d} - l \), so that the generic challenges patents until
\( \pi_{g}^{\text{lit}} - f_{g} = (1 - \gamma) \pi_{g}^{d} - l - f_{g} = 0 \Leftrightarrow \gamma_{g}^{\text{lit}} = 1 - (f_{g} + l) / \pi_{g}^{d} \).

If the firms agree on a settlement, litigation costs are saved, and the settlement surplus is now given by
\( s = \left( t_{g}^{\text{set}} - t_{g}^{\text{lit}} \right) \left( m^{\text{m}} - \left( \pi_{o}^{d} + \pi_{g}^{d} \right) \right) + 2l \). Again, we assume that the generic is compensated for the net change in its profit compared to that under litigation, which is now \( \left( t_{g}^{\text{set}} - t_{g}^{\text{lit}} \right) \pi_{g}^{d} - l \). In addition, firms equally share the surplus. Therefore, the reverse payment from \( O \) to \( G \) is given by \( r_{g}^{\text{set}} = \left( t_{g}^{\text{set}} - t_{g}^{\text{lit}} \right) \pi_{g}^{d} - l + s / 2 \), and the generic’s settlement profit becomes
\( \pi_{g}^{\text{set}} = \left( 1 - t_{g}^{\text{set}} \right) \pi_{g}^{d} + \left( t_{g}^{\text{set}} - t_{g}^{\text{lit}} \right) \pi_{g}^{d} - l + s / 2 = \pi_{g}^{\text{lit}} + s / 2 \). This corresponds to
\( \left( 1 - t_{g}^{\text{set}} \right) \pi_{g}^{d} + \left( t_{g}^{\text{set}} - t_{g}^{\text{lit}} \right) \pi_{g}^{d} - l + \left( \left( t_{g}^{\text{set}} - t_{g}^{\text{lit}} \right) \left( m^{\text{m}} - \left( \pi_{o}^{d} + \pi_{g}^{d} \right) \right) \right) / 2 + l \),
which is perfectly equivalent to Equation (4), since litigation costs cancel out.

Therefore, the generic’s challenging incentives under the settlement regime remain unchanged compared to our original assumption where \( l = 0 \). This is not surprising because the generic’s savings in litigation costs are one-to-one offset by a lower reverse payment. Hence, the critical level of patent strength \( \gamma_{g}^{\text{set}} \) and \( \tilde{t}_{\text{opt}} \) do not change.

The only aspect that changes is that consumer welfare under litigation no longer corresponds to the case where \( \tilde{t} = 0 \). Instead, we find that for \( \tilde{t} = 0 \), \( \gamma_{g}^{\text{set}} > \gamma_{g}^{\text{lit}} \). The challenge incentives under litigation are strictly smaller than those under the worst settlement (that is, a settlement without entry delay), because of the litigation costs. Consequently, we can show that \( CW_{\text{lit}} = CW(\gamma_{g}^{\text{lit}}) < CW_{\text{set}}(\gamma_{g}^{\text{set}}) = CW_{\text{set}}(\tilde{t} = 0) \), which would be even more pronounced for the implementation of the optimal policy.

**Acknowledgements** The authors thank two anonymous referees, Lawrence J. White, Tim Friese, Elisabeth Schulte, Thomas Gallt, Bernd Hayo, Jan Voßwinkel, Moritz Suppliet, Laura Werner, Rami Woodcock, and Alexandra Zaby as well as the participants of EALE 2016, the 2016 Annual Meeting of the German Economic Association, the Fifth MaCCI Annual Conference 2016, and the 45th Hohenheimer Oberseminar 2016 for valuable comments and suggestions on earlier versions of this paper. All remaining mistakes are our responsibility.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.
References

Allison, J. R., & Lemley, M. A. (1998). Empirical evidence on the validity of litigated patents. *American Intellectual Property Law Association Quarterly Journal, 26*(3), 185–277

Ayres, I., & Klemperer, P. (1999). Limiting patentees’ market power without reducing innovation incentives: The perverse benefits of uncertainty and non-injunctive remedies. *Michigan Law Review, 97*(4), 985–1033

Dickey, B., Orszag, J., & Tyson, L. (2010). An economic assessment of patent settlements in the pharmaceutical industry. *Annals of Health Law, 19*(2), 367–400

Dixit, K. A., & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *The American Economic Review, 67*(3), 297–308

Edlin, A., Hemphill, S., Hovenkamp, H., & Shapiro, C. (2015). The actavis inference: Theory and practice. *Rutgers University Law Review, 67*(3), 1–51

Elhauge, E., & Krüger, A. (2012). Solving the patent settlement puzzle. *Texas Law Review, 91*, 283–330

Encaoua, D., & Lefouili, Y. (2009). Licensing weak patents. *Journal of Industrial Economics, 57*(3), 492–525

European Commission DG Comp (2009): Pharmaceutical Sector Inquiry Final Report, Retrieved April 09, 2021, from https://ec.europa.eu/competition/sectors/pharmaceuticals/inquiry/staff_working_paper_part1.pdf

Farrell, J., & Shapiro, C. (2008). How strong are weak patents? *American Economic Review, 98*(4), 1347–1369

Farrell, J., & Merges, R. P. (2004). Incentives to challenge and defend patents: Why litigation won’t reliably fix patent office errors and why administrative patent review might help. *Berkeley Technology Law Journal, 19*(3), 943–970

Federal Trade Commission (2011): Authorized Generic Drugs: Short-Term Effects and Long-Term Impact. Retrieved April 09, 2021, from https://www.ftc.gov/sites/default/files/documents/reports/authorized-generic-drugs-short-term-effects-and-long-term-impact-report-federal-trade-commission/authorized-generic-drugs-short-term-effects-and-long-term-impact-report-federal-trade-commission.pdf

Federal Trade Commission (2002): Generic Drug Entry Prior to Patent Expiration: An FTC Study. Retrieved April 09, 2021, from https://www.ftc.gov/sites/default/files/documents/reports/generic-drug-entry-prior-patent-expiration-ftc-study/genericdrugstudy_0.pdf

FTC v. Actavis, 570 U.S. 136 (2013)

Frank, S., & Kerber, W. (2018). Patent settlements in the pharmaceutical industry: What can we learn from economic analysis? In P. Nihoul & P. Van Cleynenbreughel (Eds.), *The roles of innovation in competition law analysis.* (pp. 215–245). Edward Elgar.

Gratz, L. (2012): Economic Analysis of Pay-for-delay Settlements and Their Legal Ruling. Munich Discussion Paper, 2012–6. Retrieved April 09, 2021, from https://epub.unimuenchen.de/12734/

Hemphill, C. S. (2006). Paying for delay: Pharmaceutical patent settlement as a regulatory design problem. *New York University Law Review, 81*(5), 1553–1623

Kobayashi, B. H., Wright, J. D., Ginsburg, D. H., & Tsai, J. (2015). Actavis and multiple Anda entrants: Beyond the temporary duopoly. *Antitrust, 29*(2), 89–97

Lemley, M. A., & Shapiro, C. (2005). Probabilistic patents. *Journal of Economic Perspectives, 19*(2), 75–98

Lemley, M. A. (2001). Rational ignorance at the patent office. *Northwestern University Law Review, 95*(4), 1497–1532

Manganelli, A.-G. (2014): Delay Competition to Increase Competition: Should Reverse Payments Be Banned Per Se? Retrieved April 09, 2021, from https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SAEe2014&paper_id=511

Miller, J. S. (2004). Building a better bounty: Litigation-stage rewards for defeating patents. *Berkeley Technology Law Journal, 19*(2), 667–739

Nordhaus, W. (1969). *Invention, growth, and welfare: A theoretical treatment of technological change.* MIT Press.

Padilla, J., & Meunier, V. (2015): Should Reverse Payment Patent Settlements be Prohibited Per Se? Retrieved April 09, 2021, from SSRN
Palikot, E., & Pietola, M. (2019): Pay-for-Delay with Settlement Externalities. Retrieved April 09, 2021, from: http://www.pietola.com/research/documents/patent2016.pdf

Priest, G. L., & Klein, B. (1984). The selection of disputes for litigation. *The Journal of Legal Studies, 13*(1), 1–55

Schanckerman, M., & Schuett, F. (2017): Screening for patent quality: Examination, fees and the courts. TILEC Discussion Paper 2016–036, Retrieved April 09, 2021, from SSRN

Shapiro, C. (2003). Antitrust limits to patent settlements. *RAND Journal of Economics, 34*(2), 391–411

Thomas, J. R. (2001). Collusion and collective action in the patent system: A proposal for patent bounties. *University of Illinois Law Review, 1*, 305–353

Willig, R. D., & Bigelow, J. P. (2004). Antitrust policy toward agreements that settle patent litigation. *The Antitrust Bulletin, 49*(3), 655–698

Woodcock, R. A. (2017). Innovation and reverse payments. *Florida State University Law Review, 44*(2), 773–818

Woodcock, R. A. (2016): Product innovation in a model of settlements of drug patent litigation. Retrieved April 09, 2021, from SSRN

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.