Oxford Economic Papers ?? (2002), 1–20 **1**

Sequential innovation and the patent-antitrust conflict

By Vincenzo Denicolò

Department of Economics, University of Bologna, Piazza Scaravilli, 2, Bologna, Italy; email: denicolo@economia.unibo.it

I examine antitrust policy in a model of cumulative innovation, arguing that collusion between successive patentees (e.g. through patent pools or cross-licensing agreements) may be socially beneficial under certain circumstances, even if the patents involved are competing rather than complementary or blocking. Collusion stimulates investment in second-generation innovations, which is welfare-improving if their social returns exceed private returns. However, it discourages investment in first-generation innovations. Thus, for the pooling of subsequent patents to be beneficial, the non-appropriable returns from the second innovation must be large and it must be costly to achieve by comparison with the first.

1. Introduction

When technical progress is cumulative, successive vintages of innovations are potentially competitors in the product market. In this case, the incentives to invest in R&D and hence the speed of technical progress depend on whether or not collusion between the patentees is permitted. The intrinsic tension between the patent and antitrust laws here surfaces.

Antitrust policy tends, indeed, to prohibit the pooling of competing patents obtained by separate innovators.[1] For example, in the United States the 1995 Intellectual Property Guidelines hold that antitrust concerns arise whenever a patent combination 'harms competition among entities that would have been actual or likely competitors in a relevant market'. However, antitrust authorities 'will consider whether the restraint is reasonably necessary to achieve procompetitive efficiencies'. In practice, combination of patents is tolerated when patents are complementary,[2] or when a patentable improvement infringes on the original

[1] Combination of patents may take different forms. Strictly speaking, a patent pool is created when different patents are acquired by an agency that then licenses back the patentees (and possibly other firms as well). But the same result may be achieved if one of the patentees obtains exclusive licenses from the others. Often groups of patents are put together as the result of acquisitions or mergers. Finally, each party may retain his own patent and cross-license the others.

[2] Patents are complementary if the value of each is greater when other patents are also available; they are competing if the value decreases in the presence of the others (because the different technologies would compete in the market). See Shapiro (2001) for a discussion of complementary patents.

patent so that the first patent blocks the second.[3] But Hall (1986) finds that combinations of competing and non-blocking patents have been systematically regulated in a more restrictive way than single owners of several patents.

Does economic analysis lend support to such a policy? In a seminal paper, Green and Scotchmer (1995) develop a two-stage R&D model where one firm has the basic idea and another firm is uniquely capable of obtaining the improvement. They posit that innovators can enter *ex ante* agreements, i.e. before investment in the second innovation is sunk, and argue that if there is infringement collusion between first- and second-generation patentees is indeed desirable.

A more controversial issue is whether collusion should be permitted when, given patent rights, licensing is not required. In this case, in Green and Scotchmer antitrust policy is neutral. Chang (1995) rules out *ex ante* agreements and finds that collusion is beneficial only in limited circumstances so difficult to identify that an outright prohibition is preferable. However, Chang posits that the social and private returns from the second innovation coincide, whereas the social returns to R&D investment are actually estimated to be typically twice the private returns (Jones and Williams, 1998).

In this paper, I re-examine the issue using a two-stage patent race model. The model is an adapted version of Denicolò (2000); the main difference is that here I allow for the possibility that competition among the patentees may lead to profit erosion. The model posits free entry in both patent races (it is then natural to rule out *ex ante* agreements, as it would be too costly for the first innovator to bargain with a myriad of potential entrants); the first innovator can engage in the second race on the same footing as any outsider (there is abundant evidence that leading firms make sizable investments in R&D and can innovate repeatedly);[4] and second-generation patentees may be unable to capture the social value of innovations fully. I argue that in such a model collusion may be socially beneficial under circumstances much less limited than in Chang.

In particular, collusion serves to stimulate investment in second-generation innovations, which is valuable when their social returns exceed the private returns. Permitting collusion induces increased investment in the second innovation because second-generation innovators can gain a share of the profit from the original invention by the threat of competing away the original patentee's profit. But this may reduce investment in the first innovation, because the original innovator's competitors in the second patent race will be tougher. Consequently, if the second-generation innovation contributes significantly to consumers' surplus by

[3] In such cases, neither innovator can use the improvement without the other's consent. Then, if both innovations are to be exploited licensing agreements are indispensable; antitrust authorities and the courts generally allow the combination of such patents.

[4] See e.g. Malerba (1997). A stylized fact that emerges from many empirical studies is that incremental innovations are likely to be dominated by incumbents, while more radical technical changes tend to be associated with the entry of new firms (see Henderson, 1993). In this paper, I focus on incremental second-generation innovations.

comparison with the first, then collusion ought to be permitted.[5] The reverse may hold if the second innovation contributes little to consumers' surplus, but in this case ruling that the second innovation infringes on the first patent may be a more effective way of stimulating initial investment. A policy regime in which the second innovation does not infringe and collusion is banned may then be optimal only for intermediate values of the model's parameters.

To isolate the effects outlined above, I first postulate, as in Gilbert and Newbery (1982), that the original innovator is the first mover in the second patent race. When both patent races are treated as simultaneous-move games, another, perhaps more familiar effect comes into play: prohibiting collusion reduces joint profits but benefits consumers, creating a trade-off between static and dynamic efficiency.[6] The welfare analysis here is more complicated, but it appears that permitting collusion may still be the optimal policy for a range of parameter values.

The rest of the paper is organized as follows. The basic model of sequential patent races is developed in Section 2. Section 3 analyzes the Stackelberg equilibrium where the original innovator is the first mover in the second patent race. Section 4 models the second patent race as a simultaneous-move game. Section 5 concludes the paper.

2. The model

The cornerstone of the model is Dasgupta and Stiglitz (1980), an analysis of a single patent race. Consider an innovation that guarantees a flow of profit π to the patentee; assuming for simplicity an infinite patent life, the reward to the winner of the race is $V = \pi/r$, where r is the interest rate. The winner takes all, and the losers forfeit their R&D investment. The timing of the innovation is a probabilistic function of the amount invested in R&D by research firms. At the outset each participating firm i decides its R&D effort x_i and pays a lump sum cost αx_i, where α is the constant marginal cost of R&D effort. The R&D effort determines the expected time of successful completion of the R&D project according to a Poisson discovery process with a hazard rate equal to x_i. The projects of different firms are independent, so that the aggregate instantaneous probability of success is simply the sum of the individual probabilities. Thus the payoff function of firm i (i.e. the present value of expected profits, net of R&D costs) is

$$\int_0^\infty [\mathrm{e}^{-(X+r)t} x_i V]\, \mathrm{d}t - \alpha x_i = \frac{x_i V}{X + r} - \alpha x_i \tag{1}$$

[5] Because expected profits are driven to zero by free entry in the patent races, the contribution to consumers' surplus is all that matters to the welfare analysis.}

[6] This trade-off does not emerge when the original innovator is the first mover in the second patent race, because he then obtains the second innovation with a probability of one if collusion is prohibited and therefore first- and second-generation innovators never compete in the product market. In the simultaneous-move game there is no such pre-emption equilibrium.

where $X = \sum_i x_i$ is the instantaneous probability that one of the firms innovates, and e^{-Xt} is the probability that no firm has yet innovated by time t. With free entry into the R&D industry, the zero profit condition determines the aggregate R&D effort

$$X = \max\left(\frac{V}{\alpha} - r, 0\right) \tag{2}$$

At equilibrium, there will be a large number of R&D firms, each conducting a negligible amount of research (Denicolò, 1999).

I next extend this simple model to cases where there are two patentable innovations in prospect. At each stage there is a patent race with free entry; all firms are symmetric at the start of the first race (the symmetry breaks down after the first race). Unit R&D costs are α_t $(t = 1, 2)$. The parameter α_t is an index of the difficulty of achieving the tth innovation. The two innovations are in sequence, in that only after the first innovation is achieved can the race for the second begin. To keep the analysis meaningful, I assume that the second innovation is non-drastic.[7]

I will not make any particular assumption on the nature of innovations and the structure of the product market. Let π_1 denote the profit flow accruing to the original innovator before the second innovation, and let CS_1 denote the corresponding increase in consumers' surplus over the pre-innovation equilibrium. If the second innovation is obtained by an outsider and the two patentees compete in the product market, the first innovator's profit decreases (for simplicity, I assume that it is driven to zero)[8] and the second innovator reaps π_B; the increase in consumers' surplus over the pre-innovation equilibrium is CS_B. Finally, if the leader innovates repeatedly, or two separate patentees collude, the innovators' joint profits are π_2 and the increase in consumers' surplus is CS_2 (the division of profits between the first and the second innovator depends on patent and antitrust policy and will be specified below).

I make the following assumptions:

Assumption 1 (a) $\pi_2 > \pi_1$; (b) $\pi_2 > \pi_B$; (c) $\pi_B > \pi_2 - \pi_1$.

Assumption 2 $CS_B \geqslant CS_2 \geqslant CS_1$.

The first two parts of Assumption 1 are obvious; the third captures Arrow's replacement effect (Arrow, 1962). Assumption 2 says that the second innovation does not hurt consumers and that their gain from it is greater under competition than under collusion. Assumptions 1 and 2 will hold in most oligopoly models with suitable regularity conditions.

[7] An innovation is drastic if the patentee is unconstrained by outside competition and can therefore engage in monopoly pricing. If the second innovation were drastic, the post-innovation equilibrium would be independent of whether collusion is prohibited or allowed.

[8] This simplifies the analysis but is not essential to the results. All that matters is that competition from the second innovation lowers the first innovator's profit.

Example A homogeneous good, whose demand function is $Q(p)$, is produced under constant returns to scale. A large number of firms can operate at the pre-innovation cost, c_0, so the pre-innovation price is c_0. The first innovation lowers the unit cost to $c_1 < c_0$, and the second lowers it still further to $c_2 < c_1$.

If the first innovation is drastic, π_1 is the monopoly profit associated with the new cost level c_1; if it is non-drastic and the innovator engages in limit pricing, $p = c_0$, we have $\pi_1 = (c_0 - c_1)Q(c_0)$. Similarly, π_2 may be equal to the monopoly profit associated with the new cost level c_2, or to $(c_0 - c_2)Q(c_0)$, depending on whether the monopoly price associated with c_2 is lower or higher than c_0. Assuming Bertrand competition in the product market implies that if two separate patentees compete, the price falls to c_1 upon arrival of the second (non-drastic) innovation, the first innovator's profit falls to zero, and the second innovator gets $\pi_B = (c_1 - c_2)Q(c_1)$. Clearly, in this example the price falls or is constant after each innovation, and Assumptions 1 and 2 are always satisfied.[9] A vertical differentiation model with indentical consumers and quality-improving innovations would be iso-morphic to this example.[10]

To avoid proliferation of cases, I make an assumption that guarantees that R&D investment at each stage is strictly positive in all patent-antitrust regimes:[11]

Assumption 3 (a) $\pi_2 - \pi_1 > 4\alpha_2 r^2$; (b) $\pi_2 - \pi_B > 2\alpha_1 r^2$

To allow for the repeated innovations, I assume that the winner of the first race can participate in the second on the same footing as any outsider. Initially, I suppose that the original innovator has a first-mover advantage in the second race. This assumption is natural provided that outsiders learn of the first innovation only with some delay, and investment in R&D is irreversible so that early investment by the leader may allow him to pre-commit (see Gilbert and Newbery, 1982; Gilbert and Shapiro, 1997). However, I also develop the model under the alternative assumption of simultaneous moves in the second race, which may be appropriate in other circumstances (see Reinganum, 1983).

3. Stackelberg equilibrium

In this section I determine the model equilibrium when the original innovator is the first mover in the second patent race. In this case, if collusion is prohibited

[9] To prove that inequality $\pi_B > \pi_2 - \pi_1$ holds, note that if the first innovation is drastic, $\pi_2 - \pi_1$ is given by the area below the marginal revenue curve between c_1 and c_2, while $\pi_B = (c_1 - c_2)Q(c_1)$. Since $Q(c_1)$ is greater than marginal revenue both at $p = c_2$ and $p = c_1$, the inequality follows. A similar argument applies when the first innovation is non-drastic.

[10] With heterogeneous consumers, competition would not drive the first innovator's profit to zero after the second innovation (see footnote 8 above), but Assumptions 1 and 2 will generally continue to hold.}

[11] These conditions are sufficient, not necessary. In particular, the first part of Assumption 3 also guarantees that outsiders have the incentive to enter the second race in all regimes: see footnote 12 below.

the first innovator enforces a pre-emption equilibrium, obtaining the second innovation with certainty. If collusion is permitted, the pre-emption equilibrium disappears, and it is a matter of indifference who conducts the research at equilibrium.

The pre-emption equilibrium, and its disappearance when collusion is permitted were first analyzed by Gilbert and Newbery (1982) and Salant (1984), respectively. These papers, however, focus on the last patent race. In this section, I embed these results in a two-stage patent race model, and I focus on social welfare from an *ex ante* viewpoint.

3.1 Equilibrium

I search for the model's subgame perfect Nash equilibria. Thus, the solution is worked out backwards, starting from the second race. Assuming that collusion is always permitted if there is infringement, one can distinguish between three cases, depending on whether the second patent infringes and whether collusion is prohibited in the absence of infringement.

When there is no infringement and collusion is prohibited, the value of the second innovation to an outsider is π_B/r. Free entry by outsiders implies the zero profit condition

$$\frac{\frac{\pi_B}{r}}{\left(x_2^L + x_2^O + r\right)} - \alpha_2 \leqslant 0 \tag{3}$$

where x_2^L denotes the second-period investment by the first innovator (leader), and x_2^O denotes the aggregate second-period investment by outsiders. Clearly, $x_2^O > 0$ only if (3) holds as an equality.

The first innovator's profit as of the start of the second patent race (i.e. the prize to the winner of the first race) is given by his current profit π_1 plus the expected profit in case of repeated success, $x_2^L(\pi_2)/r$, both discounted at the adjusted discount rate $x_2^L + x_2^O + r$, less the R&D cost

$$V = \frac{\pi_1 + x_2^L \frac{\pi_2}{r}}{x_2^L + x_2^O + r} - \alpha_2 x_2^L \tag{4}$$

Consider next the case where there is no infringement and collusion is permitted. If the second innovation is obtained by an outsider, then a collusive agreement will be reached and the monopoly profit from the second innovation, π_2, will be shared between the patent holders. The disagreement point in the bargaining process is $(0, \pi_B)$, where the numbers in brackets are the flow of profits to the first and the second innovator. There is a surplus, given by $(\pi_2 - \pi_B)$; and assuming for simplicity a 50-50 split, the bargaining solution will be $(\pi_2 - \pi_B/2, \pi_2 + \pi_B/2)$.

Since $\pi_2 > \pi_B$, outsiders have now a greater incentive to enter the race. The free-entry condition for outsiders in the second patent race becomes

$$\frac{\frac{\pi_2 + \pi_B}{2r}}{x_2^L + x_2^O + r} - \alpha_2 \leqslant 0 \tag{5}$$

The first innovator's profit is

$$V = \frac{\pi_1 + x_2^L \dfrac{\pi_2}{r} + x_2^O \dfrac{\pi_2 - \pi_B}{2r}}{x_2^L + x_2^O + r} - \alpha_2 x_2^L \qquad (6)$$

The added term in the numerator of (6) captures the fact that the first innovator now gains a positive profit even if the second innovation is obtained by an outsider.

Finally, consider the case in which the second innovation infringes. In this case, the second patentee can prevent all others (including the first patentee) from using his improvement, but cannot himself exploit it without a license from the first patentee. And while the first patentee can obviously continue to use the original device, he cannot use the improvement without the consent of the second. There must therefore be bargaining between the two patentees (combinations of such blocking patents are lawful). The disagreement point in the bargaining process is now $(\pi_1, 0)$. Assuming again a fifty-fifty split, post-innovation profits are $(\pi_2 + \pi_1/2, \pi_2 - \pi_1/2)$.

In the second patent race, the payoff to a successful outsider is now $(\pi_2 - \pi_1)/2r$. Thus, free entry by outsiders implies the zero profit condition

$$\frac{\dfrac{\pi_2 - \pi_1}{2r}}{(x_2^L + x_2^O + r)} - \alpha_2 \leqslant 0 \qquad (7)$$

The leader's profit is

$$V = \frac{\pi_1 + x_2^L \dfrac{\pi_2}{r} + x_2^O \dfrac{\pi_2 + \pi_1}{2r}}{x_2^L + x_2^O + r} - \alpha_2 x_2^L \qquad (8)$$

The leader now maintains his entire profit after the second innovation, and because it cannot be used without his consent, he gets a share of the second-period rents even if the winner of the second race is an outsider.

It can be shown that outsiders always have the incentive to enter the second patent race, so at equilibrium each of the outsiders' zero-profit conditions (3), (5), and (7) must hold as an equality.[12] These conditions determine the outsiders' aggregate best-response functions, which have a slope of -1 and therefore fully

[12] To see this, it suffices to show that outsiders' profit is positive when the leader chooses his stand-alone optimal R&D investment. Suppose that $x_2^O = 0$. Then, the first order condition to the leader's problem gives

$$x_2^L = \sqrt{\frac{\pi_2 - \pi_1}{\alpha_2}} - r > 0$$

in all policy regimes (the positivity of x_2^L follows from Assumption 3). Since the outsiders' payoff is lowest under infringement (this follows from Assumption 1c), it suffices to show that outsiders have the incentive to enter under infringement. Plugging x_2^L into (7), it follows that outsiders' profit is positive at $x_2^O = 0$ iff

$$\frac{(\pi_2 - \pi_1)/2r}{\sqrt{(\pi_2 - \pi_1)/\alpha_2}} > \alpha_2$$

Assumption 3 implies that this inequality always holds.

determine aggregate equilibrium R&D expenditure in the second race, X_2. From (3), (5) , and (7) one gets

$$X_2^{NC} = \frac{\pi_B}{\alpha_2 r} - r \tag{9}$$

under 'no collusion'

$$X_2^{C} = \frac{\pi_2 + \pi_B}{2\alpha_2 r} - r \tag{10}$$

under 'collusion,' and

$$X_2^{I} = \frac{\pi_2 - \pi_1}{2\alpha_2 r} - r \tag{11}$$

under 'infringement.'

In these three cases the leader's profit reduces respectively to

$$V^{NC} = \alpha_2 r \frac{\pi_1}{\pi_B} + \alpha_2 x_2^L \left(\frac{\pi_2}{\pi_B} - 1 \right) \tag{12}$$

$$V^{C} = \frac{\pi_2 - \pi_B}{2r} + \alpha_2 r \frac{2\pi_1 + \pi_B - \pi_2}{\pi_2 + \pi_B} \tag{13}$$

and

$$V^{I} = \frac{\pi_2 + \pi_1}{2r} + \alpha_2 r \tag{14}$$

Under both collusion and infringement, V does not depend on x_2^L, which means that it is a matter of indifference who conducts the second-stage research (Salant notes this result for the case of collusion). Under no collusion, V is increasing in x_2^L, implying that the first innovator does all the second-stage research (this is Gilbert and Newbery's pre-emption result). Plugging $x_2^L = X_2^{NC}$ into (12) we get

$$V^{NC} = \alpha_2 r \frac{\pi_1}{\pi_B} + \left(\frac{\pi_B}{r} - \alpha_2 r \right) \left(\frac{\pi_2}{\pi_B} - 1 \right) \tag{15}$$

Working back to the first patent race and using (2), the free entry equilibrium is

$$X_1^{NC} = \frac{\pi_2 - \pi_B}{\alpha_1 r} + r \frac{\alpha_2}{\alpha_1} \frac{\pi_B - \pi_2 + \pi_1}{\pi_B} - r \tag{16}$$

$$X_1^{C} = \frac{\pi_2 - \pi_B}{2\alpha_1 r} + r \frac{\alpha_2}{\alpha_1} \frac{2\pi_1 + \pi_B - \pi_2}{\pi_2 + \pi_B} - r \tag{17}$$

and

$$X_1^{I} = \frac{\pi_2 + \pi_1}{2\alpha_1 r} - r \left(1 + \frac{\alpha_2}{\alpha_1} \right) \tag{18}$$

respectively.

I can now compare the R&D effort in the first and the second race under the three patent-antitrust regimes.

Proposition 1 Investment in the second innovation is greatest under collusion and least under infringement; the reverse is true of investment in the first innovation: $X_2^C > X_2^{NC} > X_2^I$ and $X_1^C < X_1^{NC} < X_1^I$.

Proof See the Appendix.

When collusion is permitted, a successful outsider can obtain a share of the profit from the original invention threatening the original patentee with competition. Consequently, the outsiders' incentive to obtain the second innovation is greatest under collusion, and since aggregate investment in the second innovation is determined solely by this incentive, it follows that $X_2^C > X_2^{NC}$. Under infringement, in contrast, the successful outsider must share the new rents with the incumbent, and the incentive to invest is therefore reduced; this implies that $X_2^{NC} > X_2^I$.

Note that aggregate profits after the second innovation are independent of antitrust and patent policy (when collusion is prohibited, there is a pre-emption equilibrium in the second race, implying that there is never competition between successive patentees in the product market). Turning to the first innovation, this means that the greater the share accruing to the successful outsider, the lower the value of the first innovation. Thus, there is a trade-off between investment in the first and in the second innovation: the regime that maximizes second-period investment generates the lowest investment in the first innovation, and vice versa.

It is interesting to contrast Proposition 1 with Green and Scotchmer's and Chang's results. In Green and Scotchmer, the fact that *ex ante* agreements are feasible means that investment in the second innovation is always at the optimal level (from the viewpoint of joint profit maximization). If the second innovation infringes, permitting collusion stimulates investment in the first innovation because it increases the second innovator's profit in case of disagreement, thus weakening the credibility of the threat not to invest in the second innovation. This allows the first innovator to increase his share of second-period rents in the *ex ante* bargaining. (If the second innovation does not infringe, however, antitrust policy does not affect the disagreement point and thus does not affect the model's equilibrium.)

In Chang, collusion is good for the original innovator *ex post*, because it prevents profit erosion; however, it also speeds up the second innovation, on average, which reduces the first innovator's expected profit. In the present model, these conflicting effects disappear as the leader pre-empts outsiders in the second race, and the sole effect at work is that collusion makes outsiders' tougher in the second patent race.

3.2 Welfare

Now let us compare the three policy regimes in terms of welfare. Social welfare is evaluated at the start of the first patent race. Since there is free entry in both races, all profits are dissipated; as a result, social welfare coincides with the discounted expected value of consumers' surplus.[13]

[13] More generally, it may include any non-appropriable returns from the innovations; for example, increased profits for firms that do not participate in the patent races but may invent around the patents or may enjoy positive spillovers, or the decrease in the profits of firms displaced by the innovations. To keep these into account, it suffices to re-interpret $\%CS$ appropriately.

We know that at equilibrium there is never competition between first and second-generation innovators in the product market. Therefore, the expected increase in social welfare over the pre-innovation equilibrium is

$$\Delta W = \int_0^\infty e^{-(X_1+r)t} X_1 \left[\int_0^\infty e^{-(x_2^L+r)s} \left(CS_1 + X_2 \frac{CS_2}{r} \right) ds \right] dt$$
$$= \frac{X_1}{X_1 + r} \quad \frac{CS_1 + X_2 \dfrac{CS_2}{r}}{X_2 + r} \tag{19}$$

This formula holds under all regimes, but X_1 and X_2 vary across regimes.

One simple special special case is when $CS_2 = CS_1 > 0$. In this case, the increase in social welfare reduces to $\Delta W = CS_1(X_1/X_1 + r)$ and therefore the social problem boils down to stimulating the first innovation. By Proposition 1, this is achieved by finding an infringement.

The analysis is more difficult when $CS_2 > CS_1$. One can show that for certain parameter values permitting collusion—with no infringement—is optimal.

Proposition 2 If $CS_2 > CS_1$, then for any given set of parameter values (i) there exists $\bar{\alpha}_1 > 0$ such that permitting collusion is optimal if $\alpha_1 < \bar{\alpha}_1$; (ii) there exists $\bar{\pi}_1 > 0$ such that permitting collusion is optimal if $\pi_1 > \bar{\pi}_1$.

Proof See the Appendix

In fact, permitting collusion may be optimal only if α_1 is sufficiently small, or if π_1 is sufficiently large. To understand why, consider the case in which all the social surplus is associated with the second innovation, i.e. $CS_1 = 0$ and $CS_2 > 0$. In this case, the overall welfare problem is to reach the second innovation as quickly as possible. Permitting collusion increases investment in the second race, which however cannot start until after the first innovation. Thus, even if all of the social surplus is associated with the second innovation, for collusion to be optimal the first innovation must be sufficiently profitable or easy to achieve.

The next Proposition shows that the welfare ranking of the regimes is monotonic in the R&D unit costs, α_1 and α_2.

Proposition 3 Ceteris paribus, the ratios $\Delta W^C/\Delta W^{NC}$ and $\Delta W^{NC}/\Delta W^I$ (and hence the ratio $\Delta W^C/\Delta W^I$) are increasing in α_2 and decreasing in α_1.

Proof See the Appendix.

Intuitively, when α_1 is small the first innovation occurs fairly soon in all regimes, so the timing of the second innovation has greater weight in the welfare comparison and permitting collusion becomes more desirable. Similarly, low values of α_2 imply that the second innovation is not much delayed anyway; therefore, the policymaker can focus on the first innovation which means that permitting collusion is now a bad policy. However, the most effective way of stimulating initial investment is to rule that the second innovation infringes on the first patent, which is therefore the best policy regime when α_2 is sufficiently low. Hence, a policy regime in which the second innovation does not infringe and collusion is banned may be optimal only for intermediate values of α_1 and α_2 (see Figs 1 and 2).

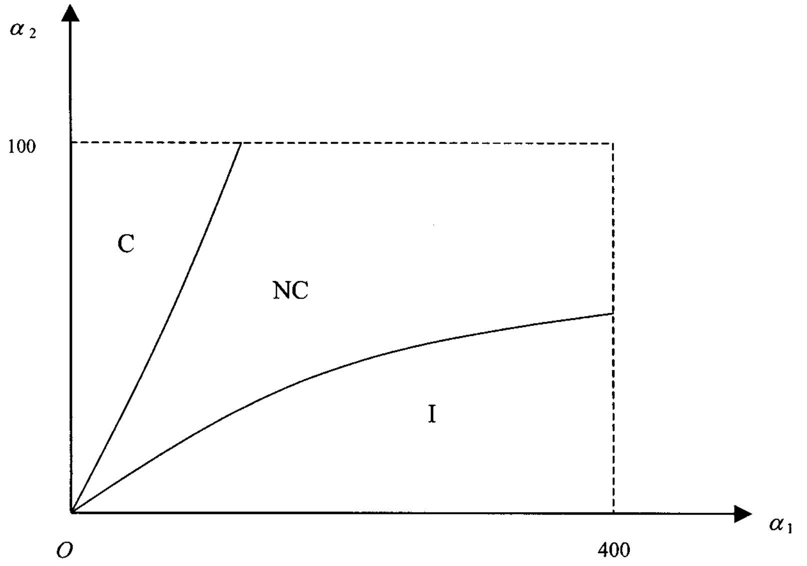

Fig. 1. Optimal policy regimes (C, collusion; NC, no collusion; I, infringement) when $CS_1 = 0$ and $CS_2 > 0$ in the Stackelberg model.

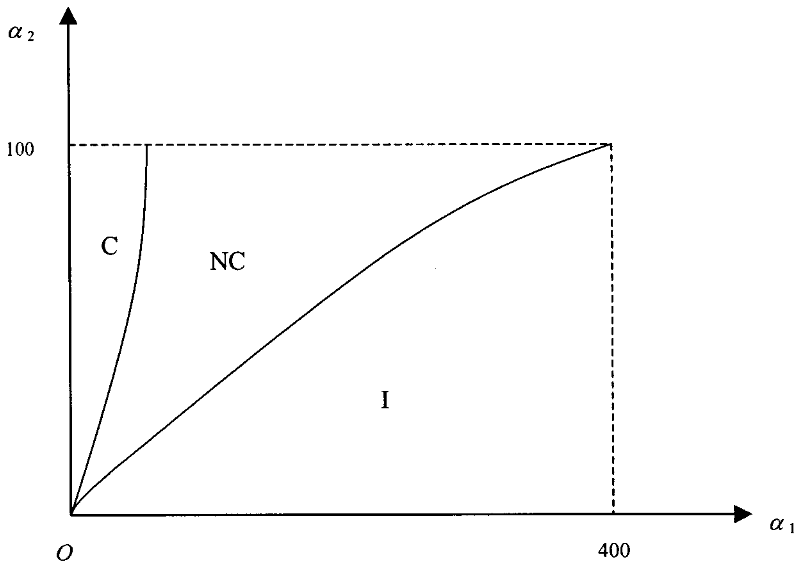

Fig. 2. Optimal policy regimes when $CS_1 = CS_2$ in the Stackelberg model.

The last issue we address is how the optimal policy depends on the magnitude of innovations. It is useful to distinguish between two questions, namely the effect of a change in the appropriable and the non-appropriable returns from the innovations.

Let us first analyze the welfare effect of a change in the profitability of the innovations. Concerning the first innovation, proceeding like in the proof of Proposition 3 it is possible to show the following (the intuition is similar to that underlying Proposition 3).

Proposition 4 Ceteris paribus, an increase in π_1 leads to an increase in both $\Delta W^C/\Delta W^{NC}$ and $\Delta W^{NC}/\Delta W^I$.

The analysis of the effect of a change in the profitability of the second innovation is more involved. There are two different indices of the profitability of the second innovation, namely π_B and $(\pi_2 - \pi_1)$, which generally do not coincide. Moreover, a change in π_B or in $(\pi_2 - \pi_1)$ affects R&D investment at both stages. Although it seems natural to conjecture that an increase in the profitability of the second innovation should have the same effects as a decrease in α_2, I have not been able to obtain any neat result.

Turning to the second question, i.e. the welfare effect of a change in the non-appropriable returns from the innovations, it is clear from (19) that the increase in social welfare is homogeneous of degree 1 in CS_1 and CS_2. This means that the only thing that matters is the ratio between the increases in consumers' surplus associated with the first and the second innovation, $CS_2/CS_1 \geqslant 1$. Because collusion stimulates the second innovation but discourages the first, high values of the ratio CS_2/CS_1 tend to favor a lenient antitrust policy.

Proposition 5 Ceteris paribus, an increase in CS_2/CS_1 leads to an increase in both $\Delta W^C/\Delta W^{NC}$ and $\Delta W^{NC}/\Delta W^I$.

Proof See the Appendix.

To summarize, permitting collusion is more likely to be socially desirable when the second innovation contributes most to consumers' surplus, the first innovation is very profitable and/or easy to achieve, and the second innovation is difficult to achieve.

The main message that emerges from the analysis is that collusion stimulates investment in the second innovation, not in the first; consequently, it is optimal when the second innovation is socially valuable. This contrasts with earlier findings in the literature, which stress the role of collusion in stimulating the first innovation. In particular, Green and Scotchmer (1995) argue that permitting collusion (in the presence of infringement) is always desirable in that it stimulates investment in the first innovation with no countervailing effects on the second. Similarly, in Chang (1995) collusion can be socially desirable only if it increases investment in the first innovation. This is because in his model the private and the social returns from investment in the second innovation coincide, so permitting collusion will lead to overinvestment in it. Creating such a distortion can be part of the optimal policy only if there is a countervailing gain in R&D investment in the first innovation (the private value of which falls short of its social value).[14]

4. Simultaneous-move equilibrium

A property of the Stackelberg equilibrium is that there is never competition in the product market along the equilibrium path, because when collusion is prohibited the winner of the first race enforces a pre-emption equilibrium in the second. This considerably simplifies the welfare analysis. In particular, it means that there is no trade-off between static and dynamic efficiency; the social trade-off is between first and second innovation.[15] In this section, I model the second patent race as a simultaneous-move game. The pre-emption result disappears, and the model may display competition among successive patentees. A trade-off between static and dynamic efficiency emerges.

4.1 Equilibrium

To proceed, I first determine the new aggregate R&D effort at each stage. Being fully determined by the outsiders' zero-profit condition, aggregate second-period investment is independent of the order of moves. Equations (13) and (14) tell us that the leader's second-period profit V depends only on aggregate second-period R&D investment if collusion is permitted or if the second innovation infringes. This means that in these cases the incentive to invest in the first innovation is independent of the order of moves; hence, $\hat{X}_t^C = X_t^C$ and $\hat{X}_t^I = X_t^I$, where a hat denotes variables pertaining to the simultaneous-move equilibrium.

When collusion is prohibited and there is no infringement, we have $\hat{X}_2^{NC} = X_2^{NC}$ but the equilibrium in the first patent race changes, because V now depends also on who conducts the research. To determine the new equilibrium, it suffices to solve the system comprising the outsiders' zero profit condition (3), taken as an equality, and the first order condition to the leader's problem

$$\frac{\left(x_2^O + r\right)\pi_2/r - \pi_1}{\alpha_2} = \left(x_2^L + x_2^O + r\right)^2 \tag{20}$$

This yields

[14] This countervailing effect can arise only if the second innovation is very profitable. In this case, it will occur quite soon whether or not collusion is permitted, and the positive effect of permitting collusion on first period investment (namely, that it prevents profit erosion) prevails (see the discussion following Proposition 1). Consequently, in Chang collusion should be permitted when the second innovation is highly profitable, whereas in the present model collusion is desirable when the non-appropriable value of the second innovation is large.

[15] The same is true of Green and Scotchmer (1995) and Chang (1995). With *ex ante* agreements in place, as in Green and Scotchmer (1995), monopoly pricing will always obtain, independent of antitrust policy. Chang (1995) assumes that innovators face inelastic demand, so that there are no deadweight losses.

$$\hat{x}_2^O = \frac{\pi_B^2}{\alpha_2 r \pi_2} - \frac{r(\pi_2 - \pi_1)}{\pi_2} \tag{21}$$

and

$$\hat{x}_2^L = \frac{\pi_B(\pi_2 - \pi_B)}{\alpha_2 r \pi_2} - \frac{r\pi_1}{\pi_2} \tag{22}$$

Substituting into (12) and working back to the first patent race, we finally get

$$\hat{X}_1^{NC} = \frac{(\pi_2 - \pi_B)^2}{\alpha_1 r \pi_2} + r\frac{\alpha_2}{\alpha_1}\frac{\pi_1}{\pi_2} - r \tag{23}$$

Comparing $\hat{X}_1^{NC}$ and X_1^{NC}, we get

$$X_1^{NC} - \hat{X}_1^{NC} = \frac{\pi_B^2(\pi_2 - \pi_B) + \alpha_2 r^2 \pi_2(\pi_B - \pi_2 + \pi_1)}{\alpha_1 r \pi_2 \pi_B} > 0$$

which means that when collusion is prohibited, R&D investment in the first race is less than in the Stackelberg equilibrium. The reason is that the outsiders now do part of the second-period research, reducing the leader's profit, and hence decreasing first-period R&D investment.

While the first part of Proposition 1 continues to hold (i.e. $\hat{X}_2^C > \hat{X}_2^{NC} > \hat{X}_2^I$), we can no longer be sure that $\hat{X}_1^{NC} > \hat{X}_1^C$: it may now happen that permitting collusion encourages both innovations.

4.2 Welfare

When collusion is permitted or the second innovation infringes, nothing changes. When collusion is prohibited and there is no infringement, with simultaneous moves the expected increase in social welfare becomes

$$W^{NC} = \frac{\hat{X}_1^{NC}}{\hat{X}_1^{NC} + r} \frac{CS_1 + \hat{x}_2^L \dfrac{CS_2}{r} + \hat{x}_2^O \dfrac{CS_B}{r}}{\hat{X}_2^{NC} + r} \tag{24}$$

because when the second innovation is obtained by an outsider, the patent holders now compete, so the increase in consumers' surplus is given by CS_B.

For any fixed patent length, the welfare comparison of the regimes is now more complicated, as two significant novel effects come into play: consumers will gain more if collusion is prohibited than if it is permitted, but the decrease in the patentees' reward means that innovations are delayed. The former effect tends to favour a strict antitrust policy, the latter a lenient.[16] Without

[16] Most of the legal literature on patent combinations focuses on these effects: see e.g. Kaplow (1984), who proposes to prohibit those collusive practices that are associated with a low ratio between patentees' reward and monopoly deadweight loss.

imposing more structure to the model, it is not possible to tell which effect will prevail.[17]

To gain more insight into the strength of these various effects, I have worked out some numerical examples. Assuming $\pi_1 = 10$, $\pi_2 = 14, \pi_B = 6$, and $r = 0.1$, I have calculated the frontier between the areas where the three regimes (infringement, no collusion, collusion) are optimal in the (α_1, α_2)-space.[18] Figures 1 and 2 display the Stackelberg-equilibrium frontiers. In Fig. 1, it is assumed that $CS_1 = 0$ and $CS_2 > 0$.[19] In Fig. 2, the two innovations contribute to consumers' surplus in the same proportion, i.e. $CS_2/CS_1 = 2$.[20] In both cases, prohibiting collusion can be optimal only for intermediate values of the parameters. As is predicted by Proposition 5, the area in which collusion is optimal shrinks when CS_2/CS_1 decreases; however, the area where infringement is optimal expands, so the net effect on the area where prohibiting collusion is optimal is ambiguous.

Figures 3 and 4 illustrate the simultaneous-move model. Assuming $CS_1 = 0$ and $CS_2 > 0$, I have re-calculated the frontiers for two cases: $CS_B = CS_2$ and $CS_B/CS_2 = 1, 1$. In the first case, the no-collusion area shrinks substantially (compare Fig. 3 with Fig. 1). In the second case (Fig. 4), permitting collusion is not optimal even if α_1 is very low, unless α_2 is sufficiently large. When the ratio CS_B/CS_2 is still larger, it turns out that a policy of permitting collusion is not optimal for any value of the R&D costs. However, these examples show that even in the simultaneous-move model collusion may be optimal for values of the parameters that seem far from extreme.

---

[17] Because these two effects are the same as those that would arise if patent length were reduced, one could think that if patent length is adjusted to balance them optimally, antitrust policy could focus on the trade-off between the first innovation and the second. In other words, the welfare analysis of the Stackelberg model would apply to the simultaneous-move model as well. However, consumers' gain when the patents expire is not the same as that entailed by competition between the patentees: patent length, that is to say, is not a perfect substitute for antitrust policy in balancing static and dynamic efficiency.

[18] With these parameter values, Assumption 3 requires that $\alpha_1 < 400$ and $\alpha_2 < 100$.

[19] In the example of cost-reducing innovations discussed in Section 2, this would correspond to the case in which the first innovation is non-drastic (so that the price does not change after the first innovation), but the two innovations cumulatively constitute a drastic cost reduction (thus, the equilibrium price falls after the second innovation). If the two innovations cumulatively are non-drastic, we have $CS_1 = CS_2 = 0$. In this special case, social welfare is independent of patent and antitrust policy in the Stackelberg model, whereas with simultaneous moves it is always desirable to prohibit collusion (because $CS_B > 0$).

[20] In the cost-reducing example of Section 2, assuming a constant elasticity demand function $Q = Ap^{-\varepsilon}$, $c_0 = 1$, $c_1 = 0.5$, and $c_2 = 0.4$, one gets $CS_2/CS_1 = 2.28$ for $\varepsilon = 3$, and $CS_2/CS_1 = 3.2$ for $\varepsilon = 6$. These calculations indicate that values of CS_2/CS_1 around 2 may correspond to cases in which, in terms of cost reduction, the second innovation is in fact much smaller than the first one.

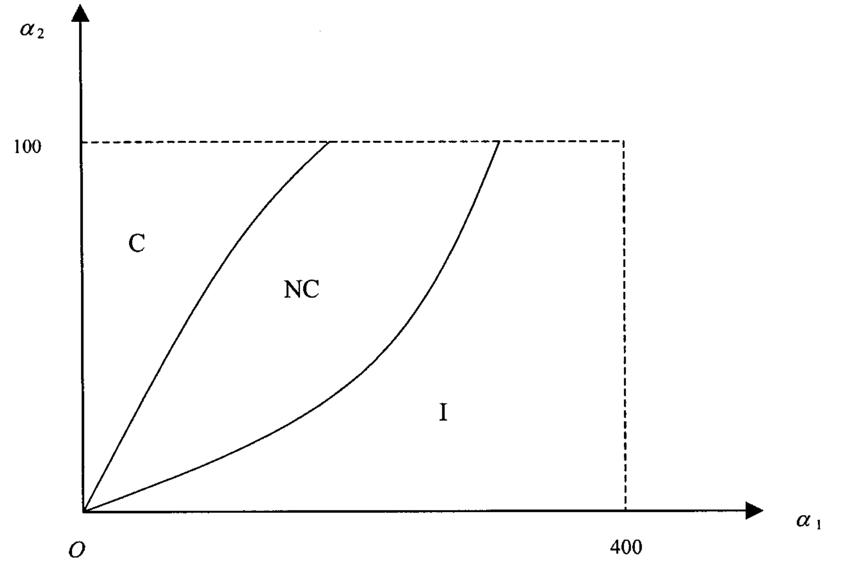

Fig. 3. Optimal policy regimes when $CS = 0$ and $CS_2 = CS_B$ in the simultaneous-move model.

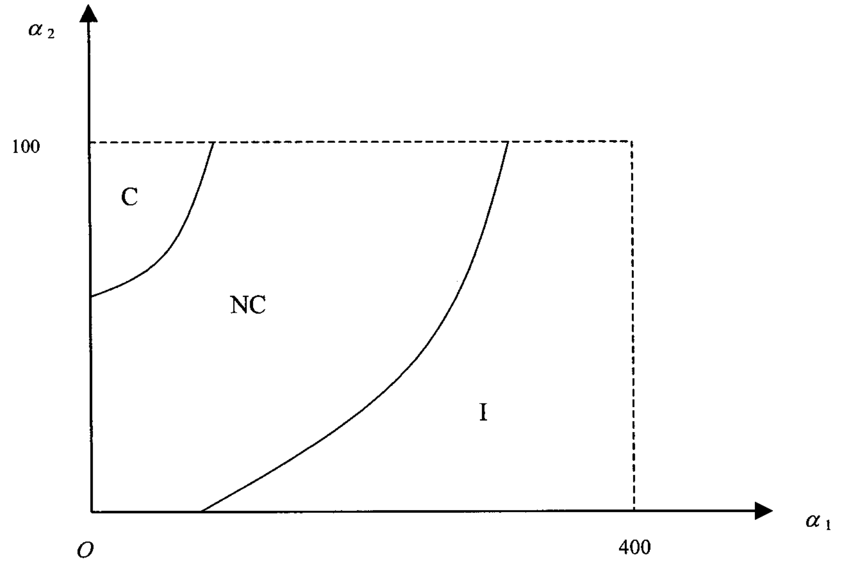

Fig. 4. Optimal policy regimes when $CS_1 = 0$ and $CS_B = 1.1\ CS_2$ in the simultaneous-move model.

5. Conclusion

Combinations of competing patents have long been subject to harsh antitrust treatment. My analysis suggest a more lenient policy in the case in which innovation is cumulative. In the Stackelberg equilibrium, permitting collusion is optimal if the non-appropriable value of the second innovation is large by comparison with the first and the second innovation is difficult to achieve or the first is easy. In the simultaneous-move model the welfare comparison is more complicated, but it may still be optimal to permit collusion in certain circumstances. These results indicate that there may be scope for a dynamic-efficiency defense of patent combinations in the case of cumulative innovation.

However, caution should be used in drawing practical implications. First, the policy conclusions are sensitive to certain critical assumptions, such as whether or not firms can enter *ex ante* licensing agreements and whether repeated innovation may occur, and more research is required to obtain a fuller characterization of the optimal policy in more highly structured models. Second, the implementation of the optimal policy is difficult because of the informational constraints faced by antitrust authorities and the courts. The alternative of permitting collusion outright is unlikely to be feasible, as it could open the gates to opportunistic behaviour: trifling improvements might be patented merely to camouflage overt collusion. Finally, other policy tools need to be taken into account, such as patent life, various requirements for patentability, or even more radical reforms of the patent system (Hopenhayn *et al.* 2000).

References

Arrow, K. (1962). 'Economic welfare and the allocation of resources for invention', in R. Nelson (ed.), *The rate and direction of innovative activity,* Princeton University Press, Princeton.

Chang, H. (1995). 'Patent scope, antitrust policy, and cumulative innovation', *RAND Journal of Economics*, **26,** 34–57.

Dasgupta, P. and Stiglitz, J. (1980). 'Uncertainty, industrial structure and the speed of R&D', *Bell Journal of Economics,* **11**, 1–28.

Denicolò, V. (1999). 'The optimal life of a patent when the timing of innovations is stochastic', *International Journal of Industrial Organization*, **17,** 827–46.

Denicolò, V. (2000). 'Two-stage patent races and patent policy', *RAND Journal of Economics*, **31,** 488–501.

Gilbert, R. and Newbery, D. (1982). 'Preemptive patenting and the persistence of monopoly', *American Economic Review*, **72,** 514–26.

Gilbert, R. and Shapiro, C. (1997). 'Antitrust issues in the licensing of intellectual property: the nine no-no's meet the nineties', *Brooking Papers on Economic Activity* (Microeconomics), 283–336.

Green J. and Scotchmer, S. (1995). 'On the division of profit in sequential innovation', *RAND Journal of Economics*, **26,** 20–33.

Hall, C.D. (1986). 'Patents, licensing, and antitrust', *Research in Law and Economics*, **8,** 59–86.

Henderson, R. (1993). 'Underinvestment and incompetence as responses to radical innovation', *RAND Journal of Economics*, **24,** 248–70.

Hopenhayn, H., Llobet G., and Mitchell, M. (2000). 'Rewarding sequential innovators: prizes, patents and buyouts', mimeo, University of Rochester.

Jones, C.I. and Williams J.C. (1998). 'Measuring the social return to R&D', *Quarterly Journal of Economics*, **113,** 1119–35.

Kaplow, L. (1984). 'The patent-antitrust intersection: a reappraisal', *Harvard Law Review,* **97,** 1813–92.

Malerba, F., Orsenigo, L., and Peretto, P. (1997). 'Persistence of innovative activities, sectoral patterns of innovation and international technological specialization', *International Journal of Industrial Organization,* **15,** 801–26.

Reinganum, J. (1983). 'Uncertain innovations and the persistence of monopoly', *American Economic Review*, **73,** 741–48.

Salant, S.W. (1984). 'Preemptive patenting and the persistence of monopoly: comment', *American Economic Review,* **74,** 247–50.

Shapiro, C. (2001), 'Navigating the patent thicket: cross licenses, patent pools, and standard setting', in A.B. Jaffe, J. Lerner and S. Stern (eds), *Innovation Policy and the Economy*, Vol. 1, MIT Press, Cambridge, MA.

Appendix

Proof of Proposition 1 Let us first compare the no-collusion and infringement equilibria. From (9) and (11) one gets

$$X_2^{NC} - X_2^I = \frac{\pi_B - \pi_2 + \pi_1}{2\alpha_2 r} > 0$$

where the sign follows from Assumption 1(c). Next, from (12) and (14) we get

$$X_1^I - X_1^{NC} = \frac{2\pi_B - \pi_2 + \pi_1}{\alpha_1}\left(\frac{1}{2r} - \frac{\alpha_2 r}{\pi_B}\right)$$

By Assumptions 1 and 3, this expression is positive and therefore $X_1^{NC} < X_1^I$.

Let us turn to the comparison between collusion and no-collusion. From (9) and (10) one gets

$$X_2^{NC} - X_2^C = \frac{\pi_B - \pi_2}{2\alpha_2 r} < 0$$

Next, from (12) and (13)

$$X_1^{NC} - X_1^C = \frac{\pi_2 - \pi_B}{\alpha_1}\left[\frac{1}{2r} - \alpha_2 r \frac{\pi_2 - \pi_1}{\pi_B(\pi_2 + \pi_B)}\right]$$

Since $\pi_2 > \pi_B$ and $\pi_B > \pi_2 - \pi_1$, we have

$$\frac{\pi_2 - \pi_1}{\pi_2 + \pi_B} < \frac{1}{2}$$

and since $\pi_B/\alpha_2 > r^2$, we get

$$\alpha_2 r \frac{\pi_2 - \pi_1}{\pi_B(\pi_2 + \pi_B)} < \frac{1}{2r}$$

and therefore $X_1^{NC} > X_1^C$. □

Proof of Proposition 2 If α_1 is sufficiently small, or if π_1 is sufficiently large, under all regimes X_1 will be so large that the ratio $X_1/(X_1+r)$ is close to 1. Then, the welfare ranking of the regimes will be determined almost exclusively by the term

$$\frac{CS_1 + X_2\dfrac{CS_2}{r}}{X_2+r} = \frac{1}{r}\left[CS_1 + \frac{X_2}{X_2+r}(CS_2 - CS_1)\right]$$

in (19), which by Proposition 1 is largest when collusion is permitted and there is no infringement. The result then follows by continuity. □

Proof of Proposition 3 Consider first the effect of a change in α_1 on W^C/W^{NC}. From (19)

$$\frac{W^C}{W^{NC}} = \frac{X_1^C(X_1^{NC}+r)}{(X_1^C+r)X_1^{NC}}\;\frac{(X_2^{NC}+r)}{(X_2^C+r)}\;\frac{X_2^C(CS_2/CS_1)+r}{X_2^{NC}(CS_2/CS_1)+r} \tag{A.1}$$

Since α_1 does not affect x_2^L, the problem reduces to calculating the effect on $X_1/(X_1+r)$. We have

$$\frac{\mathrm{d}}{\mathrm{d}\alpha_1}\log\left[\frac{X_1^C/(X_1^C+r)}{X_1^{NC}/(X_1^{NC}+r)}\right] = r\left[\frac{\mathrm{d}X_1^C/\mathrm{d}\alpha_1}{X_1^C/(X_1^C+r)} - \frac{\mathrm{d}X_1^{NC}/\mathrm{d}\alpha_1}{X_1^{NC}/(X_1^{NC}+r)}\right] \tag{A.2}$$

and since $\mathrm{d}X_1/d\alpha_1 = -(X_1+r)/\alpha_1$ under both antitrust regimes, the above derivative reduces to

$$\frac{\mathrm{d}}{\mathrm{d}\alpha_1}\log\left[\frac{X_1^C/(X_1^C+r)}{X_1^{NC}/(X_1^{NC}+r)}\right] = -\frac{r}{\alpha_1}\left(\frac{1}{X_1^C} - \frac{1}{X_1^{NC}}\right) > 0$$

where the inequality follows from Proposition 1. Thus, the ratio W^C/W^{NC} is increasing in α_1.

Next, let us consider the effect of a change in α_2. This affects both X_1 and x_2^L. Consider first the effect on W^C/W^{NC} through x_2^L. We have

$$\frac{\mathrm{d}}{\mathrm{d}\alpha_2}\log\left[\frac{(X_2^{NC}+r)}{(X_2^C+r)}\;\frac{X_2^C(CS_2/CS_1)+r}{X_2^{NC}(CS_2/CS_1)+r}\right]$$
$$= \frac{1}{\alpha_2}\left[\frac{X_2^{NC}+r}{X_2^{NC}(CS_2/CS_1)+r} - \frac{X_2^C+r}{X_2^C(CS_2/CS_1)+r}\right]$$

Since $(X_2+r)/(AX_2+r)$ is decreasing in X_2 for $A>1$, and since $X_2^C > X_2^{NC}$ by Proposition 1, the above derivative is positive. Consider next the effect through X_1. Similarly to (A.2) we have

$$\frac{\mathrm{d}}{\mathrm{d}\alpha_2}\log\left[\frac{X_1^C/(X_1^C+r)}{X_1^{NC}/(X_1^{NC}+r)}\right] = r\left[\frac{\mathrm{d}X_1^C/d\alpha_2}{X_1^C/(X_1^C+r)} - \frac{\mathrm{d}X_1^{NC}/d\alpha_2}{X_1^{NC}/(X_1^{NC}+r)}\right]$$

Next, note that

$$\frac{\mathrm{d}X_1^C}{\mathrm{d}\alpha_2} = \frac{r}{\alpha_1}\frac{\pi_B + 2\pi_1 - \pi_2}{\pi_2\pi_B}$$

and

$$\frac{\mathrm{d}X_1^{NC}}{\mathrm{d}\alpha_2} = \frac{r}{\alpha_1}\frac{\pi_B + \pi_1 - \pi_2}{\pi_B}$$

from (16) and (17). A little algebra shows that $\mathrm{d}X_1^C/\mathrm{d}\alpha_2 > \mathrm{d}X_1^{NC}/\mathrm{d}\alpha_2$, and since by Proposition 1 $X_1^C/(X_1^C+r) < X_1^{NC}/(X_1^{NC}+r)$, it follows that the effect through X_1 is also positive.

Concerning the effect of a change in α_1 and α_2 on W^{NC}/W^I, the proof is similar and is not repeated here; just note that $dX_1^I/d\alpha_2 = -r/\alpha_1$ and therefore $dX_1^{NC}/d\alpha_2 > dX_1^I/d\alpha_2$. □

Proof of Proposition 5 As in the proof of Proposition 3, consider first the effect of a change in CS_2/CS_1 on W^C/W^{NC}. Differentiating (A.1) one gets

$$\frac{d(W^{NC}/W^C)}{d(CS_2/CS_1)} = \frac{X_1^C(X_1^{NC}+r)}{(X_1^C+r)X_1^{NC}} \quad \frac{(X_2^{NC}+r)}{(X_2^C+r)} \quad \frac{r(X_2^C-X_2^{NC})}{[X_2^{NC}(CS_2/CS_1)+r]^2}$$

which by Proposition 1 is positive. This implies that W^C/W^{NC} is increasing in CS_2/CS_1. The proof that W^C/W^I is also increasing in CS_2/CS_1 is similar. □