Innovation, Propriété Intellectuelle, Concurrence et Régulation : Essais en Economie Industrielle

Yassine Lefouili

Université Paris I Panthéon-Sorbonne

U.F.R de Sciences Economiques

Année 2008-2009 Numéro attribué par la bibliothèque

|_|_|_|_|_|_|_|_|_|_|_|_|

Thèse pour le doctorat de Sciences Economiques

soutenue publiquement par

Yassine Lefouili

le 9 Décembre 2008

Innovation, Propriété Intellectuelle, Concurrence et Régulation: Essais en Economie Industrielle

Directeur de thèse

M. David Encaoua - Professeur à l'Université Paris I Panthéon Sorbonne

---

Jury

M. Rabah Amir (Rapporteur) - Professeur à l'Université d'Arizona

M. Claude d'Aspremont - Professeur à l'Université Catholique de Louvain

M. Claude Crampes (Rapporteur) - Professeur à l'Université Toulouse I

M. Patrick Rey - Professeur à l'Université Toulouse I et à l'Ecole Polytechnique

M. Bertrand Wigniolle - Professeur à l'Université Paris I Panthéon Sorbonne

Remerciements

Je ne remercierai jamais assez David Encaoua pour son implication sans bornes dans l'encadrement dont j'ai bénéficié durant ces trois dernières années. Il a été un directeur de thèse parfait. Toujours à l'écoute et d'une disponibilité rare, David a su trouver l'équilibre subtil entre m'encourager à poursuivre mes propres idées et m'orienter vers les voies de recherche qui lui semblaient les plus prometteuses, et qui se sont très souvent révélées fructueuses. Il m'a également fait l'honneur de co-écrire deux articles repris dans cette thèse. Cette collaboration qui, je l'espère, ne fait que commencer, a été pour moi l'occasion de plonger dans les méandres de l'économie de l'innovation avec un guide exceptionnel. J'aimerais également exprimer à David ma profonde gratitude pour tout le temps et l'énergie qu'il a consacrés à me soutenir dans différents projets professionnels et personnels, bien au-delà du rôle d'un directeur de thèse.

Je remercie vivement les membres de mon jury de thèse. A l'occasion d'une pré-soutenance organisée fin septembre, ils ont fait de nombreux commentaires sur mon travail, que j'ai essayé de mettre à profit de mon mieux. Rabah Amir et Claude Crampes ont accepté d'être les rapporteurs de cette thèse. C'est un grand honneur pour moi et je leur en suis très reconnaissant. Je voudrais profiter de l'occasion pour remercier Rabah Amir pour les conseils qu'il m'a prodigués et les encouragements qu'ils m'a adressés à chacune de nos rencontres. Je tiens également à exprimer ma reconnaissance à Bertrand Wigniolle pour avoir bien voulu faire partie de mon jury et pour m'avoir toujours prêté une oreille attentive lorsque j'en ai eu besoin. Enfin, la présence de Claude d'Aspremont et de Patrick Rey dans le jury est une source de grande fierté pour moi. Qu'ils en soient sincèrement remerciés.

Je tiens à exprimer ma profonde reconnaissance à Jean-Marc Robin. En m'accueillant au sein du laboratoire de microéconométrie du CREST pour un stage pendant l'été 2004, il m'a permis de découvrir le monde de la recherche académique. Il m'a par la suite été d'une aide précieuse à plusieurs reprises pendant les trois années qu'aura duré la préparation de cette thèse.

Je voudrais remercier Catherine Roux pour son amitié qui dure depuis notre rencontre sur les bancs du DEA EIME à Paris-I et sa précieuse collaboration dans la co-écriture d'un des articles repris dans cette thèse. Je remercie également Nikos Ebel dont le séjour à Paris-I a été l'occasion d'écrire un article sur lequel je m'appuie dans le dernier chapitre.

Les travaux qui constituent cette thèse ont bénéficié des commentaires de nombreux participants aux colloques et séminaires où ils ont été présentés. Qu'ils en soient tous remerciés. J'aimerais en particulier exprimer ma gratitude à Christine Halmenschlager et Patrick Waelbroeck pour leur organisation remarquable du séminaire parisien d'économie et économétrie de l'innovation.

Je remercie évidemment Tonia Lastapis et Vivianne Makougni pour leur aide au quodien et Elda André pour sa prise en charge efficace des tracasseries de fin de thèse.

Au-delà de l'aventure intellectuelle qu'elle a représentée, cette thèse a été l'occasion de vivre une véritable aventure humaine grâce aux rencontres faites au Centre d'Economie de la Sorbonne. Je pense notamment à mes collègues thésards d'EUREQua. Pour tous les bons moments que nous avons partagés merci donc à: Ekrame Boubtane, Dramane Coulibaly, Mohamad Khaled, Magali Recoules, Günes Kamber, Fabienne Llense, Nicolas Roys, Natacha Raffin, Victor Hiller, Thomas Baudin, Marie-Pierre Dargnies, Christophe Hachon, Francesco Pappadà, Jeanne Hagenbach, Clémence Berson, Lin Guo, Olfa Jaballi, Sumudu Kankanamge, Morgane Tanvé, Sébastien Le Coent,...et à tous ceux que j'oublie et auprès desquels je m'excuse.

Je tiens à exprimer toute mon amitié et ma profonde reconnaissance à Si Mohamed Sraidi. Ses conseils, fondés sur sa propre expérience doctorale, se sont avérés cruciaux lorsque j'ai dû faire des choix importants concernant ma thèse.

Last but not least, je voudrais remercier mes parents, ma soeur et mon frère pour le soutien inconditionnel qu'ils m'apportent depuis toujours.

A mes parents

Contents

Introduction

Cette thèse est constituée de quatre essais en économie industrielle. Les deux premiers relèvent de l'économie de l'innovation et de la propriété intellectuelle. Ils exploitent l'idée qu'un brevet est un titre incertain et cherchent à en tirer quelques conséquences aussi bien sur le choix du régime de protection que sur le prix d'une licence. Le troisième essai analyse un aspect particulier des programmes de clémence utilisés depuis quelques années par les autorités de la concurrence dans la lutte contre les cartels. Le quatrième essai, enfin, étudie certaines propriétés d'un mode spécifique de régulation des entreprises, à savoir la régulation à partir des performances des autres entreprises du secteur considéré.

Avant de présenter ces différents travaux, les développements qui suivent ont pour objet de rappeler brièvement le cadre conceptuel général dans lequel ils s'inscrivent. Nous présentons successivement dans cette introduction les cadres d'analyse de la notion de brevet probabiliste, des programmes de clémence comme instrument de politique de la concurrence, et du mécanisme de la régulation relative.

Brevet probabiliste : qualité, protection et licences

Un brevet d'invention n'est pas un droit de propriété ordinaire, comme celui lié à un bien physique. En tant que droit de propriété intellectuelle, un brevet porte sur un bien intangible beaucoup plus difficilement identifiable: il protège la connaissance incorporée dans un objet ou un procédé et exprimée à travers les revendications (*claims*) formulées par le demandeur du brevet. De ce fait, un certain nombre d'incertitudes pèsent sur les brevets.

Le premier type d'incertitude se situe au niveau de l'office des brevets lui-même. Un examinateur doit estimer la *nouveauté* du projet par rapport à l'état de l'art (*prior art*) au moment du dépôt, son *inventivité* par rapport à ce même état, selon le jugement qu'il s'en fait, et son *utilité* en tant qu'application industrielle. Ce sont là des critères d'appréciation difficiles pour lesquels il n'existe pas en général de métrique reconnue. De plus, ce jugement doit porter sur chacune des revendications du brevet, dont le nombre peut être très élevé. Comme le nombre de dépôts n'a cessé de croître durant ces vingt dernières années dans les trois principaux offices (USPTO aux Etats-Unis, JPO au Japon et EPO en Europe) et que le temps consacré par un ou deux examinateurs à l'appréciation de ces critères difficiles dépasse rarement trente ou quarante heures, on ne s'étonnera pas que les offices de brevets soient souvent amenés à accorder un brevet à une demande qui ne satisfait pas pleinement ces critères. On est alors en présence de brevets de mauvaise qualité ou de brevets "faibles". Certains auteurs ont ainsi avancé l'argument que la qualité d'une innovation étant une information privée, les détenteurs de "mauvais" projets sont incités à demander un brevet sachant que s'ils le font tous, l'office des brevets sera surchargé et la qualité de son examen s'en ressentira (Caillaud & Duchêne, 2005).

Le deuxième type d'incertitude concerne le détenteur du brevet. Un brevet ne devient exécutoire que si son détenteur parvient à identifier et poursuivre en procès un tiers dont il présume qu'il est un infracteur. Ce n'est là ni une activité aisée ni peu onéreuse. Par ailleurs, le détenteur n'est pas non plus à l'abri d'une contestation par un tiers qui remettrait en cause le caractère novateur ou inventif du brevet. En tout état de cause, le détenteur du brevet est en possession d'un droit que certains auteurs ont qualifié de "droit probabiliste" (Lemley & Shapiro, 2005).

Le troisième type d'incertitude se situe au niveau du tribunal. Au cours d'un éventuel procès en infraction, l'accusé peut contester la validité du brevet sur la base des critères de brevetabilité mentionnés plus haut. Le tribunal doit se prononcer alors sur cette validité, et ceci au niveau de chacune des revendications protégées. Non seulement c'est là une tâche dont le résultat peut être incertain, dans la mesure où la notion d'étendue du brevet est elle-même relativement ambiguë, mais en plus, les doctrines juridiques utilisées peuvent elles-mêmes être différentes. Par exemple, selon la doctrine juridique dite des équivalents, l'infraction peut être présumée dès qu'elle porte sur un item qui remplit la même fonction que celui qui est breveté alors qu'une doctrine plus stricte qui ne protègerait que ce qui est parfaitement décrit pourrait ne pas retenir l'infraction.

Enfin, le quatrième type d'incertitude se situe au niveau de la société. Premièrement, il s'agit de prendre en compte le fait que la mise en œuvre du système de protection par le brevet est longue et coûteuse. On peut alors faire une analyse coûts-bénéfices de ce système. Deuxièmement, l'intérêt du brevet en tant que mécanisme d'incitation à l'innovation, diffère sensiblement entre secteurs. Dans les secteurs où la connaissance est suffisamment bien codifiée pour que sa reproduction puisse être considérée comme étant peu coûteuse et facilement accessible à un tiers,[1] il semble que le brevet joue un rôle crucial en tant qu'instrument d'incitation à l'innovation. Par contre, dans les secteurs où l'innovation procède de manière continue et où une innovation présente ne doit son existence qu'à une succession d'innovations qui la précèdent (secteur des logiciels par exemple), la justification du brevet est beaucoup plus problématique.

La qualité des brevets

Même si la notion de qualité d'un brevet n'est pas appréhendée de façon identique par tous les économistes qui s'y sont intéressés, il ressort de la littérature qu'elle renvoie en général à la difficulté qu'a l'office des brevets à sélectionner parmi l'ensemble des demandes déposées celles qui satisfont les critères de brevetabilité. La qualité d'un brevet est ainsi définie par la probabilité qu'il survive à une contestation juridique de sa validité (Shapiro 2003, NAS, 2004). Si on conçoit le verdict d'un tribunal chargé de juger un litige portant sur un brevet comme le résutat d'un ré-examen approfondi des caractéristiques de l'innovation et de leur adéquation aux critères de brevetabilité, un brevet de bonne qualité a ainsi peu de chances d'être invalidé alors qu'un brevet de mauvaise qualité a de fortes chances d'être invalidé. Hall & Harhoff *(*2004) avancent une autre raison pour justifier cet indicateur de qualité du brevet. Ces auteurs défendent l'idée que du point de vue du bien-être social, un brevet de bonne qualité doit revêtir peu d'incertitude quant à sa validité et à la largeur de son champ de protection, dans la mesure où une incertitude de ce genre pourrait amener le détenteur du brevet à sous-investir dans la technologie qu'il développe, pousser les concurrents potentiels à réduire leurs investissements dans les technologies concurrentes et entraîner des litiges coûteux une fois que le détenteur du brevet

[1] C'est le cas par exemple dans le secteur pharmaceutique.

et ses concurrents ont réalisé leurs investissements.

Scotchmer (2004) insiste sur le fait que la notion de qualité d'un brevet est intimement liée à sa configuration (*design*), en termes d'étendue (*scope*) ou de largeur (*breadth*) des revendications accordées. Sachant qu'une des difficultés inhérentes à la mission de l'office des brevets est de devoir interpréter les standards de brevetabilité pour chaque nouvelle technologie, le demandeur d'un brevet insistera naturellement sur l'interprétation qui lui est la plus favorable, tandis que l'examinateur essayera de circonscrire l'étendue de la protection en fonction de son appréciation de l'état de l'art, appréciation difficile notamment dans les nouveaux domaines ouverts à la brevetabilité. De ce fait l'arbitre ultime concernant l'adéquation aux standards de brevetabilité reste le tribunal. L'office des brevets doit donc essayer d'anticiper l'issue d'un litige ou d'une contestation du brevet, et prendre lui-même les décisions concernant les conditions de brevetabilité lorsque la loi n'est pas claire. A titre d'exemple, l'office des brevets américain (USPTO) a publié en 2001 des lignes directrices (*Utility Examination Guidelines*) sur la brevetabilité des séquences génétiques. L'office des brevets a donc, selon Scotchmer, aussi bien une responsabilité dans la configuration des brevets que dans l'application de la loi. Ainsi, la mauvaise qualité d'un brevet peut être due à la pratique de l'office qui ne sélectionne pas correctement les innovations selon les standards définis par la loi, ou causée par le fait que les standards en question sont trop bas.

Causes et manifestations de la baisse de la qualité des brevets

Plusieurs considérations suggèrent que la qualité des brevets délivrés a diminué lors des dernières années. L'élargissement du domaine du brevetable, notamment aux Etats-Unis, aux logiciels, aux "business methods" et aux "research tools" a été souvent cité comme une des causes et manifestations du déclin de la qualité des brevets. Le fait qu'une grande partie de la connnaissance dans ces domaines existait sous la forme de secrets commerciaux et de produits et services non protégés, rend plus difficile la découverte d'antériorités et donc l'appréciation de la nouveauté et du degré d'inventivité des innovations relevant de ces domaines.

Par aileurs, comme cela a été dit plus haut, le nombre de demandes adressées aux principaux offices de brevets (les offices américain, européen et japonais) n'a cessé de croître[2], tandis que les ressources affectées à l'examen de ces demandes a plutôt eu tendance à stagner ou à décliner. A titre d'exemple, le rapport de la NAS (2004) décèle une baisse de 20% dans le nombre d'examinateurs par millier de demandes entre 1985 et 1997. Par ailleurs, le délai d'attente à partir de la date de dépôt et le stock de brevets en attente n'ont cessé de croître durant les deux dernières décennies.[3]

Une autre raison pouvant expliquer la baisse de la qualité des brevets est le relâchement des standards de brevetabilité, notamment en ce qui concerne le seuil d'inventivité (Barton, 2000). Pour expliquer ce relâchement, Merges (1999) met en cause la structure d'incitation financière des examinateurs de l'office des brevets américain.[4] D'une part, les niveaux des salaires sont insuffisants pour retenir les examinateurs les plus expérimentés, conduisant ainsi à un taux de turnover important. D'autre part, le système de bonus accordés aux examinateurs favorise la délivrance par rapport au rejet. Enfin, il n'existe aucune mesure de sanction des examinateurs en cas d'erreur avérée de leur part (erreur révélée par un jugement ultérieur rejetant la validité totale ou partielle du brevet).

[2]Le nombre de demandes de brevets auprès de l'office américain en 2006 a été de 425966 (World Patent Report 2008).

[3]L'office des brevets américain comptait en 2006, 1.051.502 dossiers qui n'avaient pas encore été examinés (World Patent Report 2008)

[4]La structure d'incitation financière des examinateurs de l'office européen des brevets est examinée dans un rapport récent de l'Institut d'Economie Industrielle (IDEI, 2006).

La perception générale d'une baisse de la qualité des brevets, notamment aux Etats-Unis, est en général illustrée par quelques exemples. Le cas probablement le plus célèbre est celui du "one-click" d'Amazon mais d'autres brevets tels que celui accordé à un système de file d'attente pour accéder aux toilettes dans un avion, ou celui couvrant le sandwich sans croûte, renforcent l'idée qu'une innovation peut obtenir un brevet sans vraiment satisfaire le critère d'inventivité. Par ailleurs, les litiges sur brevets, dont le volume ne cesse d'augmenter depuis les années 1980, avec une accélération significative depuis le début des annés 1990,[5] conduisent souvent à des invalidations.[6] Au cours d'un litige, une entreprise confrontée à l'accusation du détenteur d'un brevet, peut, au regard d'une antériorité trouvée dans l'état de l'art qui remet en cause la nouveauté ou la non-évidence de l'invention, demander l'invalidation du brevet en question. Un procès peut également être initié par un tiers qui conteste la validité d'un brevet et c'est alors le détenteur du brevet qui se retrouve dans le rôle de la défense. Dans les deux cas, une invalidation remet en cause la décision de l'office des brevets et conduit à une appréciation différente de celle donnée par les examinateurs. De ce point de vue, l'accroissement du nombre de brevets invalidés par les tribunaux constitue un indice de la baisse des standards de brevetabilité appliqués par les examinateurs, et par conséquent de la baisse de la qualité des brevets accordés par l'office.

Il faut néanmoins noter que certaines études empiriques portant, entre autres sur la qualité des brevets délivrés, ne détectent pas de dégradation significative de cette qualité.

[5] Le nombre de procès en infraction et/ou invalidation d'un brevet est passé de moins de 1200 procès par an avant 1990 à 2896 procès en 2007. Deux études empiriques récentes montrent que l'augmentation du nombre de litiges sur brevets n'est pas due uniquement à l'augmentation du nombre de brevets délivrés. Lanjouw & Schankerman (2004) trouvent que le nombre moyen de procès portant sur un brevet a augmenté de 21% entre la période 1985-1990 et la période 1991-1995. Bessen & Meurer (2005) montrent que le nombre de litiges par brevet détenu par une entreprise publique a augmenté de 11% entre 1987 et 1999.

[6] Allison & Lemley (1998) font état de 46% d'invalidations à la suite de litiges portant sur l'infraction ou la validité d'un brevet.

Selon les audits de qualité de l'office des brevets américain, les taux d'erreurs des examinateurs ont fluctué entre 3,6% et 7% depuis 1980, connaissant une tendance à la hausse dans les années 1990 mais une tendance à la baisse depuis (NAS 2004). Utilisant des données sur 182 brevets ayant fait l'objet de litiges entre 1997 et 2000, Cockburn et Henderson (2003) ne trouvent pas de preuve soutenant que l'expérience des examinateurs et leur charge de travail permettent de prédire les décisions d'invalidité.

Conséquences de l'existence de brevets de mauvaise qualité

La baisse de la qualité des brevets s'accompagne d'une augmentation de l'incertitude aussi bien chez les inventeurs que chez ceux qui se chargent de la commercialisation des inventions. Elle peut de ce fait créer une baisse du rythme de l'innovation ou du niveau d'investissement dans la commercialisation des nouvelles technologies.

Par ailleurs, la crainte d'un litige peut pousser les entrants potentiels de faible taille à éviter les secteurs où les firmes en place détiennent un grand nombre de brevets, même s'il y a une forte présomption que ces brevets soient de faible qualité (Lerner ,1995). Les coûts d'un litige juridique éventuel jouent dans ce cas le rôle de barrières à l'entrée entravant l'émergence de solutions technologiques alternatives.

De plus, l'absence de procédures rapides pour résoudre les problèmes de validité des brevets peut freiner le rythme de l'innovation dans des secteurs caractérisés par des innovations cumulatives (Hall & Harhoff ,2004). Dans de tels secteurs, les innovations dépendent des avancées techniques précédentes ou des avancées dans des technologies complémentaires. Si ces avancées sont couvertes par des brevets douteux ou d'une largeur excessive, les coûts des innovations qui en dépendent peuvent être plus élevés qu'ils ne devraient l'être, et de ce fait ces innovations peuvent être découragées. Par ailleurs, l'existence d'un grand nombre de brevets de mauvaise qualité peut entraîner une augmentation considérable du niveau de "fragmentation" des droits de propriété intellectuelle couvrant des technologies antérieures ou complémentaires, donnant lieu à une augmentation des coûts de transaction des innovateurs voulant avoir accès à ces technologies (à travers des licences par exemple). On retrouve là la thèse des "anti-communs" développée par Heller & Eisenberg (1998). Enfin, Hall & Harhoff (2004) affirment que la délivrance d'un grand nombre de brevets de mauvaise qualité augmente l'incertitude du détenteur sur le niveau de protection que lui confère son brevet, freinant ainsi ses propres améliorations de la technologie brevetée.

Caillaud & Duchêne (2005) soulignent un autre aspect négatif de l'existence de brevets de mauvaise qualité. Ils suggèrent que face à un processus d'examen superficiel dû à l'engorgement des dossiers, les entreprises sont encouragées à déposer des demandes peu fondées en misant sur l'attention moindre portée aux dossiers par les examinateurs. En effet, la délivrance de brevets de mauvaise qualité peut augmenter le volume des demandes: les "innovateurs" sont d'autant plus incités à déposer une demande qu'ils savent que leurs chances d'obtenir un brevet sont importantes. En conséquence, la charge de travail des examinateurs augmente, ce qui les contraint à effectuer un effort d'examen plus faible pour chaque demande, et à étudier les antériorités de manière moins approfondie, accroissant ainsi le risque de commettre des erreurs de jugement er d'accorder plus de "mauvais" brevets. Se crée ainsi un cercle vicieux qui se nourrit de la mauvaise qualité des brevets pour en produire d'autres de pire qualité.

Ford *et al.* (2007) essaient de quantifier le coût social de l'existence de brevets ne satisfaisant pas aux standards de brevetabilité. Les auteurs tentent de mesurer l'impact de l'existence de tels brevets sur les innovateurs, en partant de l'idée qu'à l'instar de la loi de Gresham, les "mauvais" brevets chassent les "bons" brevets. Ils mettent en évidence le mécanisme suivant: les mauvais brevets permettent à leurs détenteurs d'imposer des coûts de litige ou des coûts de licences suffisamment élevés pour que des innovateurs ultérieurs se découragent au point de réduire leurs investissements en R&D, ce qui donne lieu à moins de "bons" brevets et donc à une augmentation de la proportion des "mauvais" brevets. Pour mesurer le coût social de ces "mauvais" brevets, Ford *et al.* (2007) utilisent des données comparatives sur les brevets accordés aux Etats-Unis, en Europe et au Japon, en faisant l'hypothèse que les offices européen et japonais accordent moins de mauvais brevets que l'office américain. Ils aboutissent à une évaluation de la perte sociale due à la prolifération de mauvais brevets aux Etats-Unis de l'ordre de 25,5 milliards de dollars. Ils trouvent également que le coût social des litiges sur brevets est d'environ 4,5 milliards de dollars.

Bessen et Meurer (2008), quant à eux, estiment à 16 milliards de dollars par an le coût privé des litiges, niveau plus élévé que le coût social calculé par Ford *et al.* (2007). Ils trouvent que la valeur boursière des firmes poursuivies pour infraction décroît en moyenne d'un demi point de pourcentage à la suite du déclenchement d'un litige sur brevet, ce qui correspond en moyenne à 28,7 millions de dollars. Ce montant est largement plus élevé que celui des frais juridiques qui sont d'un demi-million de dollars en moyenne. Ils expliquent ces coûts indirects élévés, avant même que l'issue du litige ne soit connue, par plusieurs raisons. D'abord les chercheurs et managers de l'entreprise doivent consacrer une part importante de leur temps à produire des documents, à faire des dépositions et à monter des stratégies avec leurs avocats. Ensuite, un litige a en général des conséquences néfastes sur les relations entre le détenteur du brevet et l'infracteur potentiel. En particulier, il réduit considérablement la possibilité de coopération sur d'autres fronts. Enfin, une entreprise en mauvaise posture financière peut voir son coût du crédit augmenter significativement à cause du risque de faillite auquel elle peut être exposée suite au litige.

Le choix du régime de protection de la propriété intellectuelle

Le brevet n'est en aucun cas le seul moyen de protéger une innovation. En fait, dans la majorité des industries, il n'est pas l'instrument de protection de la propriété intellectuelle favori des entreprises. Un grand nombre d'études empiriques, dont celles de Mansfield (1986), Levin *et al.* (1987), Harabi (1995), Veugelers (1998), Arundel et Kabla (1999) et Cohen *et al.* (2000), montrent que le pourcentage des innovations brevetées varie d'une industrie à l'autre à cause notamment d'une différence inter-sectorielle dans la capacité des brevets à empêcher l'imitation. Ces études suggèrent également que le secret, l'avance acquise par l'innovateur par rapport à ses concurrents et les efforts de promotion et de vente fournissent souvent une meilleure protection que les brevets.[7] A titre d'exemple, les 600 entreprises européennes sur lesquelles porte l'étude d'Arundel et Kabla (1998) n'ont breveté en moyenne que 36% de leurs innovations de produit et 25% de leurs innovations de procédé. L'industrie pharmaceutique fait figure d'exception avec un taux de dépôt de brevets de 79% pour les produits pharmaceutiques.

Etant donné l'importance de la protection contre l'imitation dans la stratégie de protection intellectuelle des entreprises, il n'est pas étonnant que les articles théoriques ayant traité à ce sujet aient essentiellement porté sur les aspects stratégiques de la divulgation d'information lors du dépôt d'un brevet.

Crampes (1986) montre dans un modèle de duopole que le secret peut être préféré au brevet lorsque l'innovation est peu profitable ou lorsqu'elle est très profitable. Dans ce contexte d'une innovation isolée, c'est la limitation de la durée de protection par le brevet qui peut pousser une entreprise à vouloir garder une innovation (très profitable) secrète

[7] A titre d'exemple, le responsable de la propriété intellectuelle de Nokia a déclaré en 1998 que Nokia ne dépose un brevet sur une innovation que lorsqu'elle ne peut pas la garder secrète (*Teknikka & Talous* 2.4.1998).

quitte à subir un coût de protection privée pour empêcher sa diffusion et son imitation. Scotchmer & Green (1995) se placent quant à eux dans un cadre d'innovation cumulative et étudient la décision de déposer un brevet ou de garder le secret lors d'une étape de la course à l'innovation. Les auteurs montrent qu'il est possible qu'un innovateur ne dépose pas de brevet pour éviter de divulguer de l'information et conserver ainsi son avantage concurrentiel dans la course.

Horstman *et al.* (1985) considèrent que l'innovateur détient de l'information privée quant à la rentabilité d'une imitation et étudient les conséquences stratégiques de cette asymétrie d'information sur la décision de breveter dans un contexte où la protection conférée par le brevet est limitée. Ils établissent que les innovations ne sont pas toutes brevetées mais que celles qui le sont sont imitées alors que celles qui ne sont pas brevetées ne sont pas copiées. Le brevet joue ainsi, à l'équilibre, le rôle d'un signal de non-rentabilité d'une imitation.

Une autre raison pour laquelle l'option du brevet peut être écartée par un innovateur découle de la nature probabiliste du brevet. En effet, l'incertitude quant à la validité du brevet peut rendre le secret plus attractif et inciter l'innovateur à ne pas breveter ou à ne breveter qu'une partie de la connaissance incorporée dans son innovation. Il s'agit là du contexte dans lequel se placent Anton & Yao (2004) qui conçoivent le brevet comme un mécanisme de transmission de l'information: la quantité d'information brevetée et donc divulguée par l'innovateur donne la possibilité aux concurrents de réviser leurs croyances concernant l'importance de l'innovation, et d'adapter leur agressivité sur le marché et leur comportement d'imitation en conséquence.

Contrastant avec cette littérature, Saarenheimo (1994) s'attelle à expliquer la régularité empirique relevée par Schmookler (1966) selon laquelle la propension à breveter est plus importante au sein des petites entreprises que chez les grandes entreprises. L'explication qu'il propose est la suivante: les petites firmes préfèrent breveter des innovations intermédiaires et avoir ainsi la possibilité de les licencier, au lieu de chercher à se maintenir en tête durant toutes les étapes de la course à l'innovation car la probabilité d'y parvenir est plus faible pour elles que pour leurs concurrentes de grande taille. Cependant, il faut noter que l'étude d'Arundel et Kabla (1998) remet en cause l'observation de Schmookler (1966) et son explication par Saarenheimo (1994) . Utilisant un échantillon de 600 entreprises européennes, ces auteurs montrent que la propension à breveter a tendance à croître avec la taille de l'entreprise.

Les licences: une source de tension entre politique de l'innovation et politique de la concurrence

Un grand nombre de détenteurs de brevets autorisent l'exploitation de leurs innovations par des tiers à travers des accords de licences. Cockburn & Henderson (2003) trouvent que 17.6% des brevets détenus par les entreprises questionnées dans un sondage de la Intellectual Property Owners Association sont licenciés à d'autres entreprises. De plus, un huitième des firmes sondées déclarent avoir déjà développé une technologie uniquement dans la perspective de l'exploiter à travers des licences.

Les licences permettent aux firmes détentrices de brevet de s'approprier, au moins partiellement, les rentes liées à leurs innovations même lorsqu'elles ne les exploitent pas elles-mêmes à des fins productives ou commerciales. Les licences constituent de ce fait un puissant instrument incitatif en faveur de l'innovation. Par ailleurs, elles permettent la diffusion de la connaissance, et surtout la diffusion de *l'utilisation* de la connaissance, contribuant ainsi à la réalisation d'un autre objectif important de la politique de l'innovation.

En particulier, dans un contexte d'innovations séquentielles ou cumulatives, les licences permettent d'intégrer la connaissance incorporée dans une innovation en tant qu'input pour la mise au point d'une autre innovation. Par ailleurs, les accords de licences permettent souvent d'éviter ou de mettre fin à des litiges juridiques coûteux portant sur l'infraction potentielle du brevet d'une entreprise par un tiers.

Enfin, les licences sont des accords de transfert de technologie qui affectent la production et le prix des biens impliqués. De ce fait, au-delà de l'effet bénéfique sur le détenteur du brevet et le licencié, ces accords peuvent aussi affecter positivement les consommateurs. C'est le cas par exemple lorsqu'une cession de licence permet la production d'un bien par une firme plus efficace que l'entreprise détentrice du brevet.

Mais les accords de licence peuvent également contenir des clauses anti-concurrentielles. On peut citer par exemple des clauses de vente liée, de royalties qui continuent au-delà de la date d'expiration du brevet, de restrictions sur les réparations par les licenciés des machines incorporant la technologie licenciée, ou encore des clauses d'épuisement de droit ou de restrictions géographiques. Dès 1920, en déclarant qu'une clause de vente liée dans un accord de licence violait les lois antitrust, la Cour Suprême américaine mit fin à l'idée que le détenteur d'un brevet pouvait imposer toutes les restrictions qu'il voulait dans un accord de licence. Se posa alors la question de savoir quelles restrictions devaient être autorisées. Le procès *U.S. vs General Electric Co.* fournit un début de réponse: pour être légales, les restrictions contenues dans une licence doivent être "raisonnablement adaptées à la récompense pécuniaire du détenteur du brevet". Dans les *Antitrust Guidelines* de 1995, le Département de Justice américain et la Federal Trade Commission proposent une règle de raison stipulant qu'une licence ne doit être autorisée que si elle permet au détenteur du brevet de tirer des bénéfices de "l'exploitation la plus efficace possible de sa propriété intellectuelle", justifiant le préjudice éventuel infligé aux consommateurs. Le problème que pose ce principe, qui revient essentiellement à comparer l'effet *ex post* d'une licence sur la perte sèche et l'effet *ex ante* sur l'incitation à innover, est qu'il ne donne pas de règle directement utilisable par les tribunaux. L'étude de Maurer & Scotchmer (2006) apporte une réponse à la façon dont ce problème est résolu en pratique. Ils affirment que la jurisprudence américaine dans le domaine de la propriété intellectuelle repose essentiellement sur trois principes: le principe de neutralité du profit (*profit neutrality principle*), selon lequel le bénéfice tiré par un innovateur du brevet qu'il détient ne doit pas dépendre de sa capacité à exploiter lui-même son innovation, le principe d'une rémunération dérivée (*derived reward principle*) qui stipule que dans la mesure où une licence augmente la valeur sociale d'une innovation, le détenteur du brevet a droit à une part de ce surplus social, et enfin le principe du minimalisme (*minimalism principle*) selon lequel une licence doit contenir le minimum de clauses restrictives afin que les deux premiers principes soient satisfaits. Par ailleurs, Maurer et Scotchmer (2006) proposent d'adopter un certain nombre de règles *per se* parmi lesquelles la présomption de légalité de toute licence ne faisant intervenir qu'une redevance fixe et des redevances variables unitaires (royalties) constantes.

Or dans un contexte de brevets dont la validité est incertaine, ce type de licences peut avoir un effet anti-concurrentiel comme le montrent Farrell & Shapiro (2007). En effet, ils permettent au détenteur d'un brevet faible d'éviter un litige tout en imposant un taux de royalties relativement élevé que les firmes licenciées répercuteront, au moins partiellement, sur leur prix de vente. Les consommateurs paient ainsi un prix élevé à cause de l'existence d'un brevet qui n'aurait pas dû être accordé par l'office des brevets. Dans ce cas, la dégradation *ex post* du surplus des consommateurs n'est pas compensée par une augmentation de l'incitation *ex ante* à innover. Au contraire, Farrell & Shapiro (2007) montrent que la possibilité pour le détenteur d'un brevet d'imposer un niveau de royalties élévé même lorsque le brevet est faible peut diminuer les incitations à investir dans les innovations satisfaisant aux critères de brevetabilité.

Programmes de clémence: procédures et effets

> "People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices."
>
> Adam Smith. An Inquiry into the Nature and Causes of Wealth of Nations, 1776, chapter X part II.

Mesures de clémence aux Etats-Unis et en Europe

La lutte contre les cartels a connu un changement majeur dans les années 90 avec l'instauration de nouveaux programmes de clémence aux Etats-Unis. Bien que le Département de Justice américain (DoJ) disposait depuis 1978 d'une politique de clémence à l'égard d'une entreprise dénonçant un cartel auquel elle participe (ou a participé), les programmes de 1993 (*Corporate Leniency Policy*) et de 1994 (*Individual Leniency Policy*) ont fondamentalement bouleversé le contexte de la lutte contre les cartels. En effet, ces nouveaux programmes ont été plus généreux que les précédents aussi bien en termes de réduction de sanctions (allant juqu'à l'exemption complète) qu'en termes de conditions d'éligibilité. Ils ont également été plus transparents puisque le pouvoir discrétionnaire qu'avaient les autorités de la concurrence dans la décision d'accorder ou pas une réduction d'amende à un candidat au programme de clémence a été grandement réduit. Plus précisément, la section A du *Corporate Leniency Policy* accorde automatiquement l'immunité complète contre des poursuites de l'Etat au premier dénonciateur d'un cartel lorsque celui-ci ne fait pas encore l'objet d'une enquête. De plus, même si un cartel est déjà sous le coup d'une enquête, un membre du cartel peut obtenir une exemption d'amende s'il est le premier à le dénoncer. Ceci est le cas si l'autorité de la concurrence n'a pas déjà en sa possession des preuves suffisantes pour entraîner la condamnation des membres du cartel.

Ces changements opérés par le DoJ eurent un impact sans précédent sur le nombre de dénonciations de cartels et sur le montant des amendes imposées. Ainsi, le nombre d'entreprises postulant aux nouveaux programmes de clémence américains entre 1995 et 2000 a augmenté plus de dix fois par rapport à la période 1988-1993 précédant leur mise en oeuvre et le montant des amendes payées par les participants aux cartels dénoncés a atteint des sommets.[8] Spagnolo (2006) souligne que cet effet positif est probablement dû en partie à la création du cercle vertueux suivant: grâce aux dénonciations de cartels par certains de leurs membres et leur coopération durant le procès intenté par les autorités de la concurrence américaines, ces dernières ont été capables de faire condamner les partenaires "trahis" à payer des amendes plus élevées dont le montant, évidemment public, a encouragé d'autres membres de cartels à dénoncer leurs partenaires.

L'Union Européenne a été parmi les premiers à emboîter le pas aux Etats-Unis en mettant en place dès 1996 un programme de clémence. Ce premier programme se révéla inefficace dans l'incitation des membres d'un cartel à dénoncer leurs partenaires, probable-

[8]Parmi les cartels dénoncés par l'un de leurs membres, on peut citer le cartel de la vente d'art aux enchères dénoncé par Christie et celui des électrodes en graphite dénoncé par Carbide/Graphite. Mais le cas le plus célèbre reste celui du cartel mondial des vitamines. Suite à la dénonciation de ce dernier par Rhône-Poulenc, Hoffman-la-Roche a été condamné à payer 500 millions de dollars et Basf à payer 225 millions de dollars aux Etats-Unis.

ment à cause du pouvoir discrétionnaire important laissé aux autorités de la concurrrence quant à la réduction d'amende accordée. En 2002, l'Union Européenne opéra une révision profonde de son programme dans l'esprit de la révision des programmes de clémence opérée aux Etats-Unis en 1993. Depuis cette révision, le premier membre d'un cartel à dénoncer ce dernier avant l'ouverture d'une enquête obtient automatiquement une exemption d'amende. Si la dénonciation intervient après le début d'une enquête sur le marché concerné, le dénonciateur obtient une réduction d'amende allant de 30% à 100% selon l'impact des informations fournies sur le déroulement de l'enquête. Entre février 2002, date d'entrée en vigueur des nouveaux programmes de clémence européens, et décembre 2005, la Commission Européenne a reçu 167 demandes de clémence, dont 87 intervenant avant l'ouverture d'une enquête et 80 après l'ouverture d'une enquête. Le nombre de ces demandes a ainsi dépassé celles adressées au Département de Justice américain et les amendes imposées aux cartels dénoncés ont atteint des records.[9]

Une des différences fondamentales entre les programmes de clémence américain et européen concerne la possibilité laissée à une entreprise impliquée dans un cartel d'obtenir une réduction d'amende même lorsqu'elle n'est pas la première à dénoncer le cartel. En Europe, cette possibilité existe alors qu'elle est absente aux Etats-Unis: un second informateur peut obtenir une réduction d'amende de 30% à 50% si aucune enquête n'a été ouverte par la Commission Européenne et de 20% à 30% si une enquête a déjà été ouverte alors qu'il ne peut pas prétendre à une réduction d'amende aux Etats-Unis, où prévaut une

[9]En février 2007, la Commission Européenne imposa une amende totale de 990 millions d'euros à 5 membres du cartel des ascenseurs et escaliers mécaniques suite à la dénonciation de ce dernier par Kone. Le 12 novembre 2008, ce record a été battu par l'amende totale de 1,4 milliard d'euros imposée aux entreprises Saint-Gobain, Pilkington, Asahi et Soliver. Saint-Gobain a écopé de la plus forte amende jamais prononcée par Bruxelles contre une entreprise reconnue coupable d'une entente, à 896 millions d'euros. Asahi a obtenu une réduction d'amende de 50% pour sa collaboration avec la Commission *après* qu'une enquête ait été ouverte.

approche de type "winner-take-all".[10]

Une seconde différence entre les mesures de clémence américaines et européennes réside dans le traitement de l'instigateur d'un cartel. Aux Etats-Unis, une telle entreprise ne peut pas prétendre au programme de clémence alors qu'en Europe ce n'est pas systématique. En effet, l'instigateur d'un cartel n'est écarté de la procédure de clémence européenne que s'il a usé de coercition vis-à-vis des autres membres du cartel.

Une autre différence entre les deux procédures concerne la possibilité pour une entreprise d'obtenir une réduction d'amende pour un cartel qu'elle n'a pas dénoncé si elle fournit des informations sur un autre cartel dans lequel elle est, ou a été, impliquée. Le programme Amnesty Plus aux Etats-Unis prévoit cette possibilité alors qu'aucune mesure similaire n'existe en Europe.[11] Par ailleurs, la dénonciation d'autres cartels est encouragée aux Etats-Unis par l'*omnibus question.* Toute entreprise impliquée dans le procès d'un cartel se voit poser la question suivante: "Détenez-vous des informations, qu'elles soient directes ou indirectes, relatives à l'existence d'accords de fixation de prix, d'ententes lors d'enchères, etc... portant sur d'autres produits dans cette industrie ou dans une autre industrie?". Si une firme postulant au programme de clémence ne répond pas correctement à cette question ou refuse d'y répondre, elle perd l'immunité qui lui a été accordée. Cet outil semble avoir joué un rôle important dans la détection de certains cartels aux Etats-Unis tels que celui du gluconate de sodium qui a été révélé aux autorités lors de l'enquête sur les cartels de l'acide citrique et de la lysine (First, 2001). Harrington (2006) soutient que l'efficacité de cet instrument de détection des cartels provient essentiellement

[10]Il est déjà arrivé plusieurs fois aux Etats-Unis que le membre d'un cartel se présente au DoJ pour dénoncer un cartel qui l'a été à peine quelques heures plus tôt par un de ses partenaires (Hammond, 2004, 2006). En tant que second à la "course vers la dénonciation", ce membre n'a droit à aucune réduction d'amende.

[11]L'analyse de l'effet du programme Amnesty Plus sur la formation de cartels sur plusieurs marchés fait l'objet du troisième chapitre de cette thèse.

des incitations individuelles qu'ont les managers à répondre corrrectement à cette question. Ceci nous mène à une autre différence de taille entre les programmes de clémence américain et européen à savoir l'introduction aux Etats-Unis d'un programme de clémence pour individus à cause de la responsabilité pénale des managers d'entreprises participant à un cartel.[12]

Effets potentiels des programmes de clémence

Les programmes de clémence ont essentiellement quatre objectifs. Le premier est de dissuader la formation de cartels, en les rendant moins profitables. Le second est de pousser des entreprises à interrompre leur participation à un cartel. Le troisième est de détecter des cartels grâce aux informations fournies par les entreprises ayant recours aux programmes de clémence. Et le quatrième est de faciliter les poursuites judiciaires d'entreprises participant à un cartel.[13]

Une littérature récente s'est intéressée au fonctionnement des programmes de clémence, à leurs effets potentiels, aussi bien positifs que négatifs, et à la configuration (*design*) optimale de tels programmes.[14] S'inscrivant dans la lignée de la littérature sur les négociations de plaidoirie, l'article pionnier de Motta & Polo (2003) porte sur l'effet d'inclure dans les mesures de clémence les entreprises fournissant des informations sur un cartel après qu'une enquête sur le cartel en question ait été ouverte. Les auteurs montrent qu'accorder l'éligibilité aux programmes de clémence à ce stade peut avoir un premier effet, négatif, sur les incitations *ex ante* des entreprises à former le cartel en réduisant les amendes qu'elles auront à payer. Mais un second effet, positif, peut également se produire: celui de dissuader

[12]Un tel programme n'existe évidemment pas en Europe puisque les managers impliqués dans un cartel ne subissent pas de sanctions pénales.

[13]Ce dernier objectif est particulièrement important aux Etats-Unis où le DoJ doit poursuivre les membres d'un cartel devant un tribunal et convaincre un jury du bien-fondé de son accusation.

[14]Pour un survey détaillé de cette littérature, voir Spagnolo (2006).

la formation d'un cartel du fait que les poursuites par l'autorité de la concurrence deviennent moins onéreuses, plus rapides et plus efficaces. Motta et Polo (2003) soutiennent que le second effet domine le premier et en concluent que l'inclusion de firmes collaborant avec les autorités de la concurrence après l'ouverture d'une enquête est justifiée. Rey (2003) et Spagnolo (2004) mettent en lumière un autre effet déstabilisateur des programmes de clémence: ils peuvent augmenter le gain d'une entreprise qui dévie d'un accord collusif et dénonce le cartel par rapport à une firme qui dévie du cartel mais ne le dénonce pas. Cet effet, appelé par Spagnolo (2004), "*protection from fines effect*" et par Harrington (2008), "*deviator amnesty effect*" rend l'option de dévier plus profitable et peut ainsi dissuader les entreprises à former un cartel. Spagnolo (2004) identifie un autre effet positif potentiel des programmes de clémence en montrant qu'ils peuvent rendre la création d'un cartel plus "risquée" au sens de la dominance par rapport au risque d'Harsanyi & Selten (1988). Ellis & Wilson (2003), quant à eux, montrent que les programmes de clémence peuvent être utilisés par les firmes pour causer des dommages à leurs partenaires dans le cartel, qui n'en restent pas moins leurs concurrents, en augmentant leurs futurs coûts à travers les amendes payées et l'éventuel emprisonnement de leurs dirigeants. Si les partenaires en question anticipent ces incitations créées par la procédure de clémence, il est clair que cette dernière peut avoir un effet de dissuasion concernant la décision de former un cartel. Par ailleurs, Aubert *et al.* (2006) montrent que les programmes de clémence comprenant une clause de récompense monétaire aux personnes physiques dénonçant un cartel, diminuent la valeur du cartel (et donc sa stabilité) car ils obligent les entreprises à "indemniser" leurs employés pour leur "manque à gagner" s'ils ne dénoncent pas le cartel.

Mais les programmes de clémence n'ont pas que des effets positifs. Nous avons mentionné précédemment que Motta & Polo (2003) identifient un effet négatif potentiel des programmes de clémence sur la dissuasion à former un cartel à cause de la réduction de l'amende espérée en cas de création d'un cartel. Un second effet négatif provient de la possibilité pour des firmes formant un cartel d'utiliser le recours à un programme de clémence comme menace crédible vis-à-vis d'un membre du cartel qui ne respecterait pas l'accord collusif. Ce renforcement de la sanction en cas de déviation a un effet stabilisateur et peut faire qu'un cartel qui n'était pas stable en l'absence de programmes de clémence le devienne en leur présence (Bucirossi & Spagnolo, 2001, 2006, Ellis & Wilson, 2003 et Brisset & Thomas, 2004). En ce qui concerne les programmes de clémence comprenant la possibilité de récompenses monétaires pour les individus, Aubert *et al.* (2006) montrent qu'ils peuvent dissuader des formes de coopération socialement désirables. En effet, ces dernières pourraient être dénoncées par des employés aux autorités de la concurrence et leur être présentées comme étant de la collusion.

Régulation relative : avantages et limites

Dans certaines industries, un régulateur fait face à plusieurs monopoles locaux. On peut trouver cette structure de marché dans des secteurs tels que celui de la distribution de l'eau ou du gaz, de la collecte des déchets ménagers ou encore dans le secteur hospitalier. Dans une telle situation, le régulateur peut réguler un monopole local sur ses propres performances mais il peut également choisir de tenir compte d'informations sur les autres monopoles locaux pour réguler le monopole local en question.

Shleifer (1985) montre que la mise en place d'un schéma de régulation où le prix régulé pour chaque monopole local est égal à la moyenne des coûts réalisés des autres monopoles locaux, permet de mettre en oeuvre la solution de premier rang lorsque des transferts de la part du régulateur sont possibles. Un tel schéma crée en effet une "concurrence artificielle" entre les firmes (d'où le nom de *yardstick competition* donné à ce type de régulation) qui entraîne une réduction de coûts par les firmes régulées jusqu'au niveau socialement optimal. L'intuition derrière ce résultat tient au fait que le prix fixé pour chaque firme et la compensation qu'elle reçoit ne dépendent pas de sa propre performance en termes de réduction de coût, ce qui aligne ses incitations privées à réduire ses coûts avec les incitations sociales. Contrairement à la littérature récente sur la *yardstick competition* (Dalen, 1998, Sobel, 1999, Tangerås, 2002, 2003), Shleifer (1985) ne modélise pas explicitement l'incertitude qui peut exister du côté du régulateur. Ce n'est pas pour autant qu'il s'agit d'un modèle de régulation en information symétrique. En effet, pour mettre en oeuvre le schéma de régulation décrit par Shleifer, le régulateur a besoin d'observer uniquement les coûts réalisés et les dépenses des entreprises pour réduire leurs coûts. En particulier, il n'a pas besoin d'informations sur le coût initial ou sur la fonction de coût de l'investissement (excepté le fait qu'ils sont identiques pour toutes les entreprises) alors qu'il aurait besoin de ces informations pour déterminer l'optimum social et y aboutir via un schéma de régulation individuelle.

Dans un contexte de régulation par comparaison avec anti-sélection, Auriol & Laffont (1992) montrent comment les externalités informationnelles entre firmes peuvent être exploitées par un régulateur pour inciter les entreprises régulées à révéler leur type. Utilisant la structure stochastique développée par Auriol & Laffont (1992), Dalen (1998) et Tangerås (2002) montrent que l'utilisation de la *yardstick competition* peut permettre au régulateur de réduire l'asymétrie d'information dont il fait l'objet en exploitant les informations révélées pour réguler chaque firme. Exploitant la corrélation entre les performances des firmes, il peut réduire les rentes informationnelles (socialement coûteuses) concédées aux firmes régulées et donc instaurer une régulation plus efficace.

Sobel (1999) vient nuancer ces résultats en établissant que la *yardstick competition* ne permet pas d'obtenir un meilleur résultat que la régulation individuelle lorsque les fonds publics ne sont pas coûteux (i.e. pas de *shadow costs*).[15] Par ailleurs, il montre que la *yardstick competition* peut aggraver le problème de sous-investissement qui apparaît lorqu'un régulateur ne peut pas s'engager à ne pas utiliser l'information révélée par une firme pour la réguler dans le futur. Vu que la régulation relative permet au régulateur d'obtenir plus facilement de l'information, ce problème de *hold-up* est accentué lorsque la *yardstick competition* est utilisée. Néanmoins, Dalen (1998) montre que ce résultat n'est pas systématique. En effet, il met en lumière un mécanisme où n'apparaît pas l'arbitrage classique entre extraction de la rente et incitations à l'investissement (Riordan & Sappington, 1989) et montre que la *yardstick competition* peut générer des investissements plus élévés que ceux réalisés sous un régime de régulation individuelle lorsque les investissements sont spécifiques à la firme qui les entreprend. Par contre, il souligne que l'incitation individuelle à entreprendre des investissements améliorant l'efficacité de toutes les firmes régulées est nulle sous le régime de régulation relative alors que ce n'est pas le cas lorsqu'un schéma individuel de régulation est utilisé.

Une autre limite potentielle de la *yardstick competition* réside dans le fait qu'elle peut inciter les entreprises à adopter des stratégies collusives. Dans le cadre d'un modèle d'anti-sélection, Laffont & Martimort (2000) montrent que le régulateur doit introduire des nouvelles contraintes pour dissuader les firmes de coopérer. Ces contraintes induisent une distorsion supplémentaire au niveau de l'efficacité productive mais les auteurs montrent que le *third best* ainsi réalisé par le régulateur peut encore dominer le résultat qu'il obtiendrait sous un régime de régulation individuelle. Par ailleurs, Tangerås (2002) considère

[15] Dans ce cas, les rentes informationelles que les firmes obtiennent ne sont pas socialement coûteuses.

un modèle de risque moral et montre que l'incorporation de contraintes dissuadant la collusion pousse le régulateur à distordre plus l'effort requis de la part des firmes les moins productives mais permet néanmoins de mettre en oeuvre des niveaux d'effort plus efficaces pour les firmes les plus productives.

Structure et contenu de la thèse

Les quatre essais qui composent cette thèse prennent appui sur les travaux qui viennent d'être évoqués. Les deux premiers chapitres prennent pleinement en compte le caractère incertain du brevet en tant que titre de propriété intellectuelle, soit pour déterminer le choix du régime de protection (chapitre 1), soit pour examiner les propriétés des deux mécanismes de licence les plus usuels que sont la redevance unitaire et la redevance fixe (chapitre 2). Le troisième chapitre, plus proche d'une préoccupation de politique de la concurrence, étudie les propriétés du programme de clémence Amnesty Plus qui permet à une firme dénonçant un cartel sur le marché B de bénéficier d'une réduction d'amende pour le cartel du marché A découvert par l'autorité de la concurrence. Le quatrième chapitre, qui relève de la théorie de la régulation économique, étudie les propriétés d'un mécanisme de régulation relative où le prix régulé d'une entreprise dépend des coûts des autres entreprises du même secteur.

Dans le premier chapitre, nous nous intéressons au choix du régime de protection de la propriété intellectuelle par un innovateur. Plus précisément, nous étudions la décision d'un innovateur de breveter ou garder le secret sur une innovation de procédé. Pour cela, nous adoptons un modèle à plusieurs étapes dans lequel le choix entre brevet et secret industriel est affecté par trois paramètres: la force du brevet définie comme la probabilité qu'un tribunal confirme la validité du brevet et juge que l'imitation est une infraction, le coût d'imiter une innovation brevetée relativement à celui d'imiter une innovation gardée secrète, et enfin la taille de l'innovation définie comme l'importance de la réduction de coût permise par la nouvelle technologie.

Nous établissons que le choix du régime de protection est le résultat de deux effets: l'*effet dommages* et l'*effet concurrence*. Le premier effet implique que toutes choses égales par ailleurs, notamment à degré d'imitation donné, le brevet est préféré au secret. Le second effet peut jouer en sens inverse dans la mesure où l'imitation, ayant un coût moindre sous le régime du brevet que sous le régime du secret, peut conduire à un degré d'imitation plus élevé sous le premier régime.

L'un des résultats les plus importants auxquels nous parvenons est que les innovations majeures (i.e. celles qui permettent une réduction de coût substantielle) ont tendance à être tenues secrètes alors que les innovations mineures sont toujours brevetées. Les innovations intermédiaires, quant à elles, ne sont brevetées que si les brevets sont suffisamment forts. Dans ce même chapitre, nous nous intéressons à une classe de licences incorporant des royalties et une redevance fixe dans le cadre d'accords à l'amiable entre les détenteurs de brevets et leurs concurrents lorqu'un litige concernant le brevet survient. Nous montrons ainsi qu'une redevance fixe négative, c'est à dire un transfert du détenteur du brevet vers le licencié, peut soutenir un niveau élevé de royalties.

Le deuxième chapitre porte sur l'effet de l'incertitude quant à la validité d'un brevet sur les accords de licences proposés par le détenteur d'un brevet. Nous examinons deux types de licences: celles faisant intervenir uniquement une redevance fixe et celles impliquant uniquement une redevance unitaire constante. Il ressort de notre analyse que si le détenteur du brevet utilise un schéma de licence avec redevance unitaire, il peut obtenir un surprofit par rapport au profit espéré qu'il aurait eu si l'accord de licence n'était intervenu qu'une fois la question de la validité tranchée. Néanmoins, ce résultat n'est plus vrai si la licence proposée stipule uniquement un paiement fixe. Il peut ainsi arriver que le détenteur du brevet préfère avoir recours à une licence avec redevance unitaire dans le cas d'un brevet de validité incertaine alors qu'il aurait préféré proposer une licence avec redevance fixe si la validité du brevet avait été certaine. L'incertitude sur la validité d'un brevet peut donc être vue comme une explication alternative au fait que le schéma avec redevance unitaire est utilisé par certains détenteurs de brevets. Par ailleurs, nous montrons que le pouvoir de marché excessif dont jouit le détenteur du brevet lorsqu'il utilise un schéma de licence avec redevance unitaire peut être diminué si les entreprises prennent ensemble la décision d'accepter la licence ou d'attaquer la validité du brevet. Enfin nous dégageons deux recommandations de politique économique: d'une part, les autorités de la concurrence devraient être particulièrement vigilantes lorsqu'un accord de licence fait intervenir une redevance unitaire et d'autre part, les incitations de licenciés potentiels à attaquer collectivement la validité d'un brevet devraient être renforcées.

Dans le troisième chapitre, nous nous intéressons au programme de clémence mis en place par les autorités américaines pour encourager des entreprises participant à un cartel découvert à dénoncer d'autres cartels auxquels elles participent (ou ont participé). Ce programme, qui porte le nom d'Amnesty Plus, permet au membre d'un cartel découvert d'obtenir une réduction de l'amende qui lui est imposée pour sa participation à ce cartel, s'il dénonce un autre cartel auquel il participe (ou a participé). De façon plus précise, nous nous intéressons à l'effet de ce programme sur les incitations des firmes à former un ou plusieurs cartels. Il ressort qu'Amnesty Plus n'a d'effet pro-concurrentiel que lorsque certaines conditions sont réunies. Lorsque ce n'est pas le cas, ce programme peut avoir un effet stabilisateur et encourager la formation d'un plus grand nombre de cartels qu'en son absence. Nous identifions les circonstances exactes sous lesquelles cet effet anti-concurrentiel a lieu et proposons une règle simple permettant de l'éviter. Enfin, nous montrons comment les entreprises peuvent exploiter leur contact sur plusieurs marchés pour diminuer l'effet pro-concurrentiel ou renforcer l'effet anti-concurrentiel d'Amnesty Plus.

Le quatrième chapitre étudie l'effet du degré de contrainte sur les prix d'une politique de régulation sur les incitations des entreprises régulées à réduire leurs coûts de production. Nous nous intéressons à deux modes de régulation: une régulation individuelle à travers laquelle une firme se voit fixer un prix plafond dépendant de son propre coût réalisé, et une régulation relative où le prix plafond fixé par le régulateur est fonction du coût réalisé d'autres firmes, elles-mêmes soumises au même schéma de régulation. Nous montrons qu'il existe une différence fondamentale entre ces deux types de régulation quant à l'effet d'une politique de régulation plus contraignante sur l'investissement en réduction de coût. Lorsqu'un schéma de régulation individuelle est utilisé, une contrainte plus forte (en termes de prix plafond) sur les firmes régulées diminue leurs incitations à réduire leurs coûts alors que sous un schéma de régulation relative c'est le contraire qui se produit. En ce qui concerne les prix, nous montrons qu'une régulation plus contraignante induit un prix plus faible dans le cas de la régulation relative et que l'effet est ambigu lorsque la régulation est individuelle. Il s'avère ainsi qu'il n'y a pas de tension entre l'amélioration de l'efficacité productive et celle de l'efficacité allocative sous le régime de régulation relative alors que cette tension peut apparaître sous le régime de régulation individuelle. Ceci est un avantage de la régulation relative puiqu'elle permet d'aligner les deux objectifs de diminution des prix et de réduction des coûts. Néanmoins, nous montrons qu'il existe deux forces qui pourraient dissuader le régulateur d'imposer une régulation trop contraignante sous un régime de régulation relative. La première est due au fait que, sous ce régime, la contrainte de participation des firmes n'est pas satisfaite si la régulation est trop stricte, ce qui n'est pas le cas sous le régime de régulation individuelle. La seconde concerne l'effet négatif d'une régulation plus contraignante sur l'investissement en amélioration de la qualité, effet qui pourrait pousser un régulateur à relâcher la contrainte sur les firmes pour ne pas pénaliser ce type d'investissement.

Références

Allison, J.R and M.A. Lemley, 1998, "Empirical evidence on the validity of litigated patents, *AIPLA Quarterly Journal* 26, 3, 185-275.

Anton, J.J. and D. A. Yao, 2004, "Little patents and big secrets: managing intellectual property", *Rand Journal of Economics* 35, 1-22.

Arundel, A. and I. Kabla, 1998, "What percentage of innovations are patented? Emprirical estimates for European firms", *Research Policy* 27, 127-141.

Aubert, C., Kovacic, W and P. Rey, 2006, "The impact of leniency and whistle-blowing programs on cartels", *International Journal of Industrial Organization* 24, 1241-1266.

Auriol, E. and J.J. Laffont, 1992, "Regulation by duopoly", *Journal of Economics and Management Strategy* 1, 507-533.

Barton, J., 2000, "Intellectual property rights: reforming the patent system", *Science* 287, 1933-1934.

Bessen, J. and M. Meurer, 2005, "The patent litigation explosion", Boston University School of Law Working Paper No. 05-18.

Bessen, J. and M. Meurer, 2008, "The private costs of patent litigation", Boston University School of Law Working Paper No. 07-08.

Brisset, K., and T. Lionel, 2004, "Leniency program : A new tool in competition colicy to deter cartel activity in procurement auctions", *European Journal of Law and Economics* 17, 5-19.

Buccirossi, P., and G. Spagnolo, 2001, "Leniency programs and illegal exchange: how (not) to fight corruption", Working Papers in Economics and Finance No. 451, Stockholm School of Economics.

Buccirossi, P., and G. Spagnolo, 2006, "Leniency programs and illegal transactions", *Journal of Public Economics* 90,1281-1297.

Caillaud, B., A. Duchêne, 2005, "Patent office in innovation policy: Nobody's Perfect", mimeo.

Chong, E., 2004, "Yardstick competition vs. individual incentive regulation: What has the theoretical literature to say?", *mimeo.*

Cockburn, I. and R. Henderson, 2003, "Survey results from the 2003 Intellectual Property Owners Association Survey on Strategic management of Intellectual Property", *mimeo.*

Cohen, W.M, R.R. Nelson and J.P. Walsh, 2000, "Protecting their intellectual assets: appropriability conditions and why US manufacturing firms patent or not", NBER W.P. 7552

Crampes, C, 1986, "Les inconvénients d'un dépôt de brevet pour une entreprise innovatrice", *L'Actualité Economique, Revue d'Analyse Economique*, 62, 4, 521-534

Dalen, D. M., 1998, "Yardstick competition and investment incentives", *Journal of Economics & Management Strategy* 7, 105-126.

Ellis, C. J. and W. Wilson, 2003, "Cartels, price-fixing, and corporate leniency policy: what doesn't kill us makes us stronger." Manuscript, University of Oregon.

Farrell, J. and C. Shapiro, 2007, "How strong are weak patents", forthcoming, *American Economic Review.*

First, H., 2001, "The vitamins case: cartel prosecution and the coming of international competition law", *Antitrust Law Journal* 68, 711-734.

Ford, G., Koutsky, T. and L. Spivak, 2007, "Quantifying the cost of substandard patents: some preliminary evidence", Phoenix Center Policy Paper Number 30.

Hall, B. and D. Harhoff, "“Post-grant patent reviews in the U.S. patent system - design choices and expected impact,” *Berkeley Law and Technology Review* 19, 989-1015.

Hammond, S. D. 2004, "Cornerstones of an effective leniency program", presented at the ICN Workshop on Leniency Programs, Sydney

Hammond, S. D., 2006, "Measuring the value of second-in cooperation in corporate plea negotiations", presented at the 54th Annual American Bar Association Section of Antitrust Law Spring Meeting.

Harabi, N., 1995, "Appropriability of technical innovation: an empirical analysis", *Research Policy*, 24, 981–992.

Harsanyi, J., and R. Selten, 1988, *A General Theory of Equilibrium Selection in Games*, Cambridge MA: MIT Press.

Harrington, J., 2006, "Corporate leniency programs and the role of the antitrust authority in detecting collusion", Competition Policy Research Center Discussion Paper CPDP-18-E.

Harrington, J., 2008, "Optimal corporate leniency programs", Journal of Industrial Economics 56, 215–246.

Heller, M. and R. Eisenberg, 1998, "Can patents deter innovation? The anticommons in biomedical research", *Science* 280,698-701

Horstman, I. M., MacDonald, G. and A. Slivinski, 1985, "Patents as information transfer mechanism: to patent or (maybe) not to patent", *Journal of Political Economy* 93, 837–858.

IDEI report, 2006, *Objectives and Incentives at the European Patent System*, Institut d'Economie Industrielle, Toulouse.

Laffont, J.-J., and Martimort, D., 2000, "Mechanism design with collusion and correlation", *Econometrica* 68, 309-342.

Laffont, J.-J., and Tirole, J., 1993, *A Theory of incentives in procurement and regulation*, MIT Press, Cambridge.

Lanjouw, J.O. and M. Schankerman, 2004, "Protecting intellectual property rights: are small firms handicaped?", J*ournal of Law and Economics* 114, 441-465.

Lemley, M.A. and C. Shapiro, 2005, "Probabilistic patents", *Journal of Economic Perspectives* 19, 75-98

Lerner, J., 1995, "Patenting in the shadow of competitors", *Journal of Law and Economics* 38, 563-95.

Levin, R.C., A.K. Klevorick, R.R. Nelson and S.G. Winter, 1987, "Appropriating the returns from industrial R&D", *Brooking Papers on Economic Activity*, 783-820

Mansfield, E., 1986, "Patents and innovation: an empirical study", Management Science, 32, 217–223.

Maurer, S.M. and S. Scotchmer, 2006, Profit neutrality in licensing: the boundary between antitrust law and patent law", *American Law and Economics Review* 8,476-522.

Merges, R., 1999, "As many as six impossible patents before breakfast: property rights for business concepts and patent system reform", *Berkeley Technology Law Journal* 14, 577-615.

Motta, M, and M. Polo., 2003, "Leniency programs and cartel crosecution," *International Journal of Industrial Organization* 21, 347-379.

NAS (National Academy of Sciences), 2004, *A Patent system for the 21*$^{\text{st}}$ *century*, Washington DC: National Academy od Sciences.

Rey, P., 2003, "Towards a theory of competition policy", Ch. 3 of *Advances in Economics and Econometrics: Theory and Applications*, Eight World Congress,M. Dewatripont, L.P. Hansen S.J. Turnovsky eds., Cambridge University Press.

Riordan, M., and D.E.M., Sappington, 1989, "Second sourcing", *Rand Journal of Economics* 20, 41-58.

Saarenheimo, T., 1994, "Studies on Market Structure and Technological Innovation". Bank of Finland Publications, Series B:49.

Schmookler, J., 1966, *Invention and economic growth*, Harvard University Press.

Shapiro, C., 2003, "Antitrust analysis of patent settlements between rivals", *Rand Journal of Economics* 34, 391-411

Scotchmer, S., 2004, *Innovation and Incentives*, The MIT Press, Cambridge, Ma.

Scotchmer, S. and J. Green, 1995, "On the division of profit between sequential innovators", *Rand Journal of Economics* 26, 20-33.

Shleifer, A., 1985, "A theory of yardstick competition", *Rand Journal of Economics*, 16,319–327.

Smith, A., 1904, *An Inquiry into the nature and causes of the Wealth of Nations.* E. Cannan (ed), vol. 2.London: Methuen & Co. (Originally published in 1776).

Sobel, J. 1999, "A reexamination of yardstick competition," *Journal of Economics & Management Strategy* 8, 33-60.

Spagnolo, G., 2006, " Leniency and whistleblowers in antitrust", CEPR Discussion Paper No. 5794.

Tangerås, T. P., 2002, "Collusion-proof yardstick competition", *Journal of Public Economics*, 83, 231–254.

Veugelers, R., 1998, "Collaboration in R&D: an assessment of theoretical and empirical findings", *De Economist* 146, 419–443.

World Patent Report, 2008, World Intellectual Property Organization, available at: http://www.docstoc.com/docs/1998416/World-Patent-Report-2008.

Chapter 1

Choosing Intellectual Protection: Imitation, Patent Strength and Licensing[1]

1.1. Introduction

The traditional view that firms always prefer patents to other forms of protection for their innovations has been empirically challenged for a long time. It is now well known that secrecy, first mover advantage and exploitation of lead time may be preferred forms of protection in many industries. Even in the same industry, forms of protection may differ according to the nature and importance of the innovation and to the disclosure effect. Early studies by Scherer (1965, 1967, 1983) have shown that the propensity to patent varies significantly accross industries and that inter-industrial variations in patenting activity are not explained by R&D expenditures. Pakes and Griliches (1980) were among the first

[1]This chapter is based on a joint work with David Encaoua.

to find that the degree of randomness in the patenting activity within industries was not explained by R&D variations. They have shown that the residual patenting behaviour was explained by the potential imitation allowed by the disclosed information and by the innovator's capability to appropriate the rents generated by the innovation. Mansfield (1986) obtained similar results based on a survey where US manufacturing firms were asked what fractions of inventions they would not have developed in the absence of patents between 1980 and 1983. These fractions were very low in many industries (less than 10% in electrical equipment, primary metals, instruments, motor vehicles and others) and relatively high in industries like pharmaceuticals (60%) and chemicals (40%). Two more recent surveys (Yale Survey by Levin *et al.,* 1987 and Carnegie Mellon Survey by Cohen *et al.* 2000) confirm these trends: it is only in industries where knowledge is strongly codified that patents appear to be substantially preferred to other forms of protection.

Despite this accumulated empirical evidence, theoretical explanations of why and when an innovator would prefer to keep an innovation secret rather than patent it remain rather scarce. Before turning to the related literature, note that even if patenting is not considered as the best form of protection, innovators have a lot of reasons to apply for patents serving purposes different from protection (Hall and Ziedonis, 2001, Encaoua *et al.* 2004). This feature complicates the problem. Indeed, a theoretical explanation of why and when patenting is not the best form of protection must also be consistent with the more cumbersome issue of why, despite the existence of preferred forms of protection, patents remain so widespread (Scotchmer, 2004).

At least three types of theoretical arguments are required to explain the protection choice.

First, patents must be recognized as not being ironclad property rights but rather *probabilistic rights.* Lemley and Shapiro (2005) qualify the uncertain nature of patents in a suggestive way: "*A patent does not confer upon its owner the right to exclude a potential imitator but rather a right* to try *to exclude by asserting the patent in court. When a patent holder asserts its patent against an alleged infringer, the patent holder is rolling the dice. If the patent is found invalid, the property right will have evaporated*". Thus, patent strength refers to the probability to recover damages, with the consequence that only strong patents give in principle to the patent holder the right to exclude an infringer or to force him to buy a license. But as we argue below, even holders of weak patents may escape the uncertain litigation process. They succeed through their licensing strategies to capture a significant part of the consumers' surplus. This is why the notion of patent quality enters so forcefully in the agenda of antitrust authorities nowadays, especially through the criticism of the examination system in the Patent Office (Merges, 1999, Lemley, 2001).

Second, the traditional view that knowledge is a blueprint has also been challenged. Replicating an existing invention may be costly and time consuming because knowledge may be more embedded in individuals and firms than in physical products or equipment. One has to distinguish innovations according to their *secrecy effectiveness*, which is the main determinant of the imitation cost (Anton *et al.* 2005). Many innovations involve hidden know-how even if the allowed performance is perfectly observable. Consider for instance a process innovation leading to a cost reduction that is reflected back in the market price but for which the technological knowledge is neither perfectly revealed nor easily reverse-engineered. In this case, imitating the process innovation or building around it may be rather costly. Moreover, the imitation cost may depend on whether the invention has been patented or not. As the patent discloses some enabling technological information, it is clear that imitating a patented innovation should be at most as costly as if it was kept secret. Our paper offers a natural framework to analyze the classical trade-off between getting a legal protection involving a compulsory disclosure of enabling information and keeping secrecy by giving up legal protection.

Third, even if patents do not always appear as the best form of protection, innovators may nevertheless prefer to patent their innovations because holding a patent offers the possibility to settle a dispute against an alleged infringer through a licensing agreement (Lemley and Shapiro, 2005, Farrell and Shapiro, 2007). Alleged infringers may also prefer to avoid a litigation process not only because litigation is costly but also because winning the lawsuit against the patent holder involves a free-riding aspect, as other competitors benefit from the asserted patent's invalidity. Therefore, even when they are weak, patents may generate substantial revenues through licensing royalties that are likely to harm consumers. This is why patent settlements raise serious concerns for competition policy authorities (Shapiro, 2003, Encaoua and Hollander, 2004).

The main objective of this paper is to introduce these arguments in a simple model allowing a discussion of the following issues:

i/ What are the different forces that interact in the choice of a protection regime (patent or secrecy)?

ii/ How are these forces affected by the patent strength, imitation cost and innovation size?

iii/ What sort of licensing agreements are likely to emerge in order to avoid patent litigation?

Two main contributions have explicitly explored the decision whether to patent an innovation or not.[2] Hortsman *et al.* (1985) assume that an innovating firm possesses private

[2] Among other papers related to the choice of an intellectual property regime, one can mention Crampes (1986), Gallini (1992) and Scotchmer and Greene (1990). Crampes (1986) examines the trade-off between keeping an invention secret during an indefinite time or obtaining a legal protection over a finite duration (the statutory patent life). Gallini (1992) introduces the idea that breadth governs the cost of inventing around the patent. However, it is entry cost rather than imitation cost that matters. Scotchmer and

information about profits available to competitors and that patent coverage may not exclude profitable imitation. Conceived as an information transfer mechanism, a patent that covers full information is not optimal. The optimal innovator's choice is a mixed strategy between patenting and keeping secrecy while the follower's optimal choice is to stay out of the market when the innovator patents and to imitate when the innovator does not. The peculiarities of this model, in terms of the patent's signaling aspect and the *a priori* restrictions put on the follower's actions, explain why there is no imitation of a patented innovation at equilibrium. Since our paper is close to Anton and Yao (2004), we describe more thoroughly their framework. Starting from the premise that disclosure provides competitors with usable information and focusing on the innovator's decision about how much of an innovation should be disclosed, their model is particularly relevant for a special kind of secrecy effectiveness. They describe a situation where the real innovation performance is not directly observable while the disclosed know-how enables a competitor to costlessly replicate it. Therefore, by choosing the amount to be disclosed, the innovator directly controls the behavior of the potential imitator. Their model is a signaling game where the innovator has private information on the innovation size and decides to reveal, partially or fully, this information, letting the potential imitator infer the leader's advance. The follower chooses either to imitate or not under the risk of infringement. A refined perfect bayesian equilibrium of the signaling game involves a separating strategy in which: i/ small innovations are patented and fully disclosed; ii/ medium innovations are patented and partially disclosed; iii/ large innovations are not patented and partially disclosed through a public announcement. This result is illustrated by their suggestive title: "little patents

Greene (1990) focuses on the impact of patent policy on the incentives to innovate. Their model involves a binary choice as the innovation would not be realized if it were not patented. They also assume full disclosure of technological know-how.

and big secrets".

In our model, we maintain the general trade-off an innovator is confronted to when choosing its protection regime. However, rather than focusing on the signaling aspect we assume that: i/ the process innovation size, measured by the cost reduction, is directly observable; ii/ a patent reveals technological information that lowers the imitation cost relative to the situation where the know-how embedded in the innovation is kept secret.[3] Choosing to patent the invention may expose the innovator either to an increased imitation level or to a lower one because the imitation level does not only depend on the imitation cost but also on two other crucial parameters: the innovation size and patent strength. It may happen that an innovator benefits from being imitated : this occurs whenever the incurred loss due to imitation is overcompensated by the damages it receives from an imitator if the court upholds the patent validity *and* the patent infringement. If patent protection and secrecy lead to the same imitation level, then patenting is preferred since damages are expected under the patent regime. This corresponds to the *damage effect*. But whenever imitation levels differ according to the protection regime, a conflict arises as long as imitation becomes higher under the patent regime. This corresponds to the *competition effect*. Therefore, when the imitation extent is decided by the follower, different interactions may occur between the competition effect and the damage effect. Our paper aims to clarify these interactions. We propose a complete information multistage game in which three common knowledge parameters are important: the innovation size, the patent strength and the relative cost of imitation. We depart from the assumption limiting *a priori* an imitator's behaviour by letting it choose its own cost reduction in response to

[3]If the reduction of the imitation cost directly depends on the disclosed level of enabling knowledge, using a relative imitation cost parameter is equivalent to using a disclosure level. It appears however that working with the relative imitation cost parameter is more realistic since the imitation extent remains controlled by the imitator, while the choice of a protection regime is made by the innovator.

the process innovation. We also assume that the damages paid by an imitator infringing a valid patent are equal to its profit, which corresponds to the *unjust enrichment* legal doctrine. Our damages specification only presents a slight difference with the infringer's revenue rule assumed in Anton and Yao (2004).[4] Thus, our results may be compared to the "little patents and big secrets" results in Anton and Yao.

Our main findings are as follows. For a given innovation size, patent strength and relative imitation cost generally act as substitutes: A decrease in one of these parameters must be compensated by an increase in the other one in order to keep the same value of the innovator's profit. Inventors of small process innovations always prefer patent protection to secrecy. This reminder of the "little patents" result in Anton and Yao (2004) rests however on a different argument in our model. For large process innovations, our results present some difference with the " big secrets" characterization in Anton and Yao. Our model does not totally discard the possibility of patenting some large process innovations, whenever imitation is too costly. This may happen when information is poorly disclosed in the patent. In this case the innovator is indifferent between secrecy and patent protection. For medium process innovations, our results differ more significantly from those of Anton and Yao. It is not optimal for a firm producing such an innovation to file a patent of bad quality, that is a patent having a low probability to be upheld by the court, unless the disclosed information does not significantly lower the imitation cost. We show that there exists a *safe protection level* that is sufficient to deter any imitation and that this level is lower than a 100% protection. As the innovation size decreases, protecting the innovation by means of secrecy becomes less likely. Finally, the "one size fits all" principle in the

[4]In a more recent paper, Anton and Yao (2005) introduce the "lost" profits of the patentee, defined as the profits that would have occured in the absence of infringement. They show that at equilibrium, infringement may take one of two forms: a "*passive*" form in which lost profits of the patentee are zero and an "*agressive*" form where they are positive. One of the main results in Anton and Yao (2005) is that infringement always occurs when damages equal lost profits. This last result does not hold in our model.

patent design is not validated by our analysis.

These results raise many practical issues. While the model predicts that it is seldom optimal for a firm to file a patent when the probability that it will be upheld by the court is low, bad quality patents (relatively to novelty and non-obviousness requirements) are widespread in the real word. How can this be explained? Moreover, why are low quality patents not litigated more often than we observe? These issues are at the heart of what has been called the *patent paradox* (Hall and Ziedonis, 2001, Scotchmer, 2004). We devote a brief analysis that suggests a possible answer to explain this paradox. Whenever a patent is not conceived only as a protection against imitation but also as a tool to reach private settlements through licensing agreements (Shapiro, 2003), licensing agreements may act as an alternative to patent litigation. A royalty rate independent of the patent strength combined to a specific fixed fee may serve this purpose.

The model is presented in section 1.2. The market competition outcome under the shadow of infringement is described in section 1.3. The imitator's behaviour is analyzed in section 1.4. The core of the paper, which corresponds to the protection regime choice is examined in section 1.5. We devote section 1.6 to licensing agreements. Our conclusions are presented in section 1.7.

1.2. The basic set-up

We examine a process innovation in a framework involving two competing firms. We suppose that firm 1 is an innovating firm and firm 2 is a potential imitator. Each firm is risk-neutral and seeks to maximize its expected profit. Initially, both firms produce at the same marginal cost $c > 0$. Fixed production costs are assumed to be zero. We assume that firm 1 undertakes an R&D investment which reduces its marginal cost to the level

$d_1 < c$. The game we consider below starts once the innovation is realized and involves three stages.

First, in the *protection stage*, the innovator has to choose between two protection regimes. The first regime, which we denote by P, is to patent its innovation, and the second regime, which we denote by S, consists of protecting its innovation by means of secrecy.

Second, in the *imitation stage*, after the observation of the innovator's marginal cost d_1, firm 2 chooses whether, and the extent to which, it wishes to imitate (or "build around") the innovator's technology. It makes an investment to reduce its marginal cost to $d_2 \in [d_1, c]$. Note that we do not allow the imitator to improve the innovator's technology.[5] The difference $c - d_2$ captures the "extent of imitation". When $d_2 = c$, there is no imitation at all and when $d_2 = d_1$, imitation is full. We assume that imitation at a level $d_2 \in [d_1, c[$ induces a fixed imitation cost $I(d_2)$ which depends on whether the innovation is patented or kept secret. More specifically, we assume that the imitation cost under the patent regime, which we denote by $I^P(d_2)$, and the imitation cost under the secrecy regime, which we denote by $I^S(d_2)$, satisfy the following condition :

$$I^P(d_2) = fI^S(d_2)$$

where the parameter $f \geq 0$ measures the relative costs of imitation under the regimes P and S. We assume that $f \leq 1$: since patenting involves a compulsory disclosure, it is likely that imitating a patented innovation will turn out to be less costly than imitating a secret innovation.

[5] Green and Scotchmer (1995) and Chang (1995) examine the issue of technology improvement by a second generation innovator in a cumulative innovation framework.

Third, in the *competition stage*, market outcomes are determined under the shadow of punishment. We assume that, when the innovation is patented, firm 1 sues firm 2 for infringement if firm 2 imitates firm 1 by adopting a follow-up technology that reduces its marginal cost to $d_2 < c$. We also assume that firm 2 systematically contests the validity of the patent covering the innovation.[6] We denote by g the probability that the patent survives the imitator's legal contestation of the patent's validity. We interpret this parameter as the *patent's quality*: low quality patents have higher chances to be invalidated by a court than high quality patents. Thus, a higher quality patent (in terms of novelty and inventiveness) is less uncertain: the probability that a court will uphold its validity is higher. We denote by e the probability that an imitation infringes the innovator's patent conditional on the patent being valid. This probability can be interpreted as an indicator of the lagging patent's *breadth*: the broader the patent's breadth, the higher the probability that a follow-up technology that reduces the marginal cost c to $d_2 \in [d_1, c[$ is an infringement of the patent on the process innovation d_1. Firm 2 is compelled to pay damages, supposed to be equal to its market profit, if and only if the patent is held valid and the imitation infringes the patent. This occurs with probability $\theta = eg$. The parameter $\theta \in [0, 1]$, which we assume to be common knowledge,[7] corresponds to what it is called the *patent strength* (Shapiro 2003). When the innovation is not patented, no damages are paid.

Following Anton and Yao (2004), we model our duopoly market competition as a tra-

[6] The typical defense in real-world patent infringement suits is to contest infringement *and* patent validity.

[7] Assuming that the parameter θ is common knowledge is consistent with our goal of putting aside any signaling effect. However, this assumption might seem to be quite strong: It is reasonable to think that the innovator might have some private information as to the validity of its patent (captured through the parameter g) and the imitator might have some private information as to the probability of infringement e. See Bebchuk (1984) and Meurer (1989) for analysis of patent litigation under asymmetric information.

ditional Cournot competition with linear market demand:

$$p(x_1 + x_2) = a - (x_1 + x_2)$$

where x_1 is the output of firm 1, x_2 is the output of firm 2 and $p(x_1 + x_2)$ is the market clearing price.

We assume that $c < a < 2c$. The first inequality is usual and means that the marginal cost before innovation is below the choke price. The second inequality expresses that the market is small which is a likely scenario for innovative markets, as it allows the possibility that the innovative firm becomes at least twice as efficient as it currently is. In other words, the inequality $a < 2c$ implies that there exist innovations d_1 such that $d_1 < 2c - a$, which can also be written as $a - d_1 > 2(a - c)$.

We choose a convex specification for the imitation technology and, to reach analytical results, we use the following quadratic expression :

$$I(d_2) = \begin{cases} f\frac{(c-d_2)^2}{2} & if \quad Patent \\ \frac{(c-d_2)^2}{2} & if \quad Secret \end{cases}$$

1.3. Competition stage

Competition occurs under the shadow of litigation only if the innovation is patented. Therefore, the outcome of the competition stage depends on whether the innovation is patented or not.

1.3.1. Patented innovation

We separately examine the cases $d_2 < c$ (i.e. the follower imitates the innovator, at least partially) and $d_2 = c$ (i.e. the follower does not imitate the innovator), as the profit functions differ between these two cases.

The follower imitates ($d_2 < c$) : Under regime P, the expected gross profits of firm 1 and firm 2 are given by:

$$\Pi_1^P(x_1, x_2, d_1, d_2, \theta) = \underbrace{(a - (x_1 + x_2) - d_1)\, x_1}_{\text{market profits}} + \underbrace{\theta\, (a - (x_1 + x_2) - d_2)\, x_2}_{\text{expected damages}}$$

and

$$\Pi_2^P(x_1, x_2, d_2, \theta) = (1 - \theta)\,(a - (x_1 + x_2) - d_2)\, x_2$$

From the expected profits, one derives the Cournot-Nash equilibrium outputs $x_1^P(d_1, d_2, \theta)$ and $x_2^P(d_1, d_2, \theta)$. They correspond either to an interior solution where both firms are active : $x_1^P(d_1, d_2, \theta) \geq x_2^P(d_1, d_2, \theta) > 0$ or to a boundary solution where only firm 1 is active : $x_1^P(d_1, d_2, \theta) > x_2^P(d_1, d_2, \theta) = 0$.

Consider first an interior solution. Routine computations lead to:

$$x_1^P = \frac{a(1-\theta) + d_2(1+\theta) - 2d_1}{3-\theta}$$

$$x_2^P = \frac{a - 2d_2 + d_1}{3-\theta}$$

A necessary and sufficient condition for an interior solution to exist is :

$$d_2 < \frac{a + d_1}{2} \tag{1.1}$$

Note that this condition is always satisfied when $d_1 > 2c - a$. Note also that the market price $p^P(\theta)$ is given by $p^P(\theta) = \frac{a+d_2(1-\theta)+d_1}{3-\theta}$ which is increasing in $\theta \in [0,1]$ as long as condition (1) is satisfied.

Consider now a boundary solution. Such a solution arises when condition (1) is not satisfied and is characterized by:

$$x_1^P = \frac{a - d_1}{2} \quad \text{and} \quad x_2^P = 0$$

The follower does not imitate $(d_2 = c)$ **:** The equilibrium outputs in this case can be derived from those of the previous case by taking $\theta = 0$ and $d_2 = c$. Hence:

- If $d_1 > 2c - a$, then the equilibrium outputs are given by:

$$x_1^P = \frac{a + c - 2d_1}{3} \quad \text{and} \quad x_2^P = \frac{a + d_1 - 2c}{3}$$

- If $d_1 \leq 2c - a$, then we have the same boundary solution as in the imitation case:

$$x_1^P = \frac{a - d_1}{2} \quad \text{and} \quad x_2^P = 0$$

Summing up these cases, the expected equilibrium gross profits depend on d_1, d_2 and θ in the following way :

- If $d_1 \leq 2c - a$ then:

$$\Pi_1^P(d_1, d_2, \theta) = \begin{cases} \frac{[a-d_1(2-\theta)+d_2(1-\theta)][a(1-\theta)-2d_1+d_2(1+\theta)]+\theta[a-2d_2+d_1]^2}{(3-\theta)^2} & if \quad d_2 < \frac{a+d_1}{2} \\ \frac{(a-d_1)^2}{4} & if \quad \frac{a+d_1}{2} \leq d_2 \leq c \end{cases} \tag{1.2}$$

$$\Pi_2^P(d_1, d_2, \theta) = \begin{cases} (1-\theta)\frac{(a-2d_2+d_1)^2}{(3-\theta)^2} & if \quad d_2 < \frac{a+d_1}{2} \\ 0 & if \quad d_2 \geq \frac{a+d_1}{2} \end{cases} \tag{1.3}$$

- If $d_1 > 2c - a$ then:

$$\Pi_1^P(d_1, d_2, \theta) = \begin{cases} \frac{[a-d_1(2-\theta)+d_2(1-\theta)][a(1-\theta)-2d_1+d_2(1+\theta)]+\theta[a-2d_2+d_1]^2}{(3-\theta)^2} & if \quad d_2 < c \\ \frac{(a+c-2d_1)^2}{9} & if \quad d_2 = c \end{cases} \tag{1.4}$$

$$\Pi_2^P(d_1, d_2, \theta) = \begin{cases} (1-\theta)\frac{(a-2d_2+d_1)^2}{(3-\theta)^2} & if \quad d_2 < c \\ \frac{(a+d_1-2c)^2}{9} & if \quad d_2 = c \end{cases}$$

Therefore, under Cournot competition, firm 2 is driven out of the market if it keeps its old technology when the innovation is *large enough* ($d_1 < 2c - a$) and remains active on the market (even without imitating firm 1) when the cost reduction innovation is *small enough* ($d_1 > 2c - a$). This result depends on the small market assumption ($a < 2c$). In a large market ($a > 2c$), firm 2 would remain in the market whatever the innovation size. Thus, the small market assumption allows for strategic aspects that would not be captured otherwise.

1.3.2. Non-patented innovation

The equilibrium outcomes under the secrecy regime are derived from those under the patent regime by taking $\theta = 0$. This simply means that no damages are paid when imitation occurs under secrecy. One obtains :

$$\Pi_1^S(d_1,d_2)=\begin{cases}\frac{(a-2d_1+d_2)^2}{9} & if \quad d_2<\frac{a+d_1}{2}\\ \frac{(a-d_1)^2}{4} & if \quad d_2\geq\frac{a+d_1}{2}\end{cases} \tag{1.5}$$

$$\Pi_2^S(d_1,d_2)=\begin{cases}\frac{(a-2d_2+d_1)^2}{9} & if \quad d_2<\frac{a+d_1}{2}\\ 0 & if \quad d_2\geq\frac{a+d_1}{2}\end{cases} \tag{1.6}$$

1.4. Imitation stage

Firm 2 aims to maximize its net profit when it chooses its imitation level $d_2 \in [d_1, c]$. Since the follower's gross profit and imitation cost depend on whether the innovation is patented or not, we need to distinguish between these two regimes.

1.4.1. Patented innovation

Under this regime, the imitator's net profit when it adopts a technology allowing to produce at marginal cost $d_2 \in [d_1, c]$ is given by :

$$G_2^P(d_1,d_2,f,\theta)=\Pi_2^P(d_1,d_2,\theta)-\frac{1}{2}f\,(c-d_2)^2$$

The follower's optimal imitation level when the innovation is patented is determined as:

$$d_2^P(d_1,f,\theta)=Arg\max_{d_2\in[d_1,c]}G_2^P(d_1,d_2,f,\theta)$$

Define $A(\theta)=\frac{1-\theta}{(3-\theta)^2}$. It is a decreasing function of $\theta\in[0,1]$ such that $A(0)=\frac{1}{9}$ and $A(1)=0$.

The function $d_2 \rightarrow H(d_1, d_2, f, \theta) = A(\theta)(a - 2d_2 + d_1)^2 - \frac{1}{2} f\,(c - d_2)^2$, which is the expression of $G_2^P(d_1, d_2, f, \theta)$ when $d_2 < \frac{a+d_1}{2}$ and $d_2 < c$, is necessarily either convex or concave over its whole definition domain. The following preliminary results are easy to show:

1- The function $d_2 \rightarrow H(d_1, d_2, f, \theta)$ is stricly convex if $f < 8A(\theta)$ and is strictly concave if $f > 8A(\theta)$.

2- The unconstrained extremum of $d_2 \rightarrow H(d_1, d_2, f, \theta)$ is easily obtained by the the FOC:

$$d_2^{int}(d_1, f, \theta) = c + \frac{4A(\theta)(d_1 - 2c + a)}{8A(\theta) - f} \tag{1.7}$$

In order to obtain the value of $d_2^P(d_1, f, \theta)$, it is necessary to know whether $H(d_1, d_2, f, \theta)$ is convex or concave and to compare $d_2^{int}(d_1, f, \theta)$ to d_1 and c. For instance, when $d_1 < 2c - a$ and $f > 8A(\theta)$ (H strictly concave), equation (8) leads to $d_2^{int}(d_1, f, \theta) > c$. Moreover, as $G_2^P(d_1, d_2, f, \theta)$ is a discontinuous function in d_2 for $d_2 = c$, it is necessary to compare the value of $G_2^P(d_1, c, f, \theta)$ obtained in the absence of imitation to the value of $\underset{d_2 \in [d_1, c[}{Arg\max} G_2^P(d_1, d_2, f, \theta)$ obtained with imitation.

1.4.1.1. Drastic innovations ($d_1 < 2c - a$)

With drastic innovations, partial imitation never occurs. The following proposition (see appendix A1 for the proof) distinguishes two areas in the (θ, f) space according to whether the optimal imitation level is maximum (i.e. $d_2^P(d_1, f, \theta) = d_1$) or minimum (i.e. $d_2^P(d_1, f, \theta) = c$). In the (θ, f) space, the extent of each of these two areas depends on d_1.

Proposition 1 *For drastic innovations ($d_1 < 2c - a$), there exists a threshold function $\rho\,(d_1, \theta) = 2A(\theta)\left(\frac{a-d_1}{c-d_1}\right)^2$ which is decreasing in the patent strength θ and the innovation* *size* $c - d_1$ *such that :*

If $f < \rho(d_1, \theta)$, *then the follower fully imitates:* $d_2^P(d_1, f, \theta) = d_1$.

If $f > \rho(d_1, \theta)$, *then the follower does not imitate* $d_2^P(d_1, f, \theta) = c$ *and is driven out of the market.*

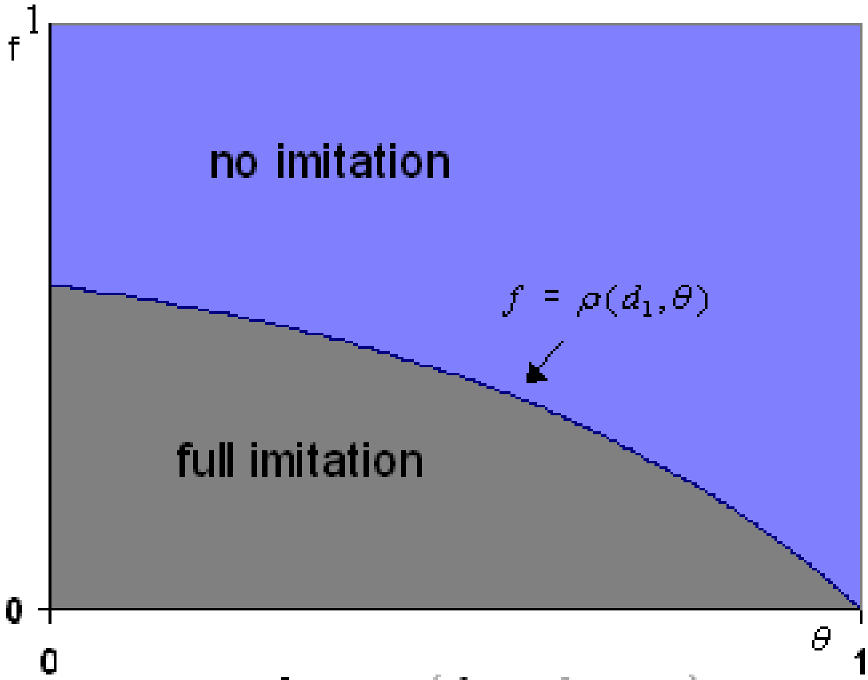

figure **1** $(d_1 < 2c - a)$

The interpretation of this proposition, illustrated in figure 1, is quite straightforward. When the process innovation is large enough ($d_1 < 2c-a$), there exists a threshold imitation cost $\rho(d_1, \theta)$ such that if f is below this threshold it is optimal for a follower to fullly imitate the patented innovation ($d_2^P(d_1, f, \theta) = d_1$), whereas if f is above the threshold, it does not pay to imitate. Note that $\rho(d_1, 0) = \frac{2}{9}\left(\frac{a-d_1}{c-d_1}\right)^2 < \frac{8}{9}$ for any $d_1 < 2c - a$. Therefore, sufficiently large patented innovations, even when protected by a weak patent (θ not far from 0), will not be imitated as long as the imitation cost parameter f is sufficiently high ($f > \frac{8}{9}$). This result means that for a sufficiently high lead advance of the innovator and a sufficiently high imitation cost, imitation of the patented innovation never occurs and the technological follower is driven out of the market. This result, which occurs under a low intensity of competition in the product market (Cournot competition) is likely to hold under a higher intensity of competition.[8] Another important result is that the threshold imitation cost $\rho(d_1, \theta)$ decreases as the patent is stronger (higher θ) and as the innovation is larger (lower d_1). Hence the size of the region where no imitation occurs increases in the innovation size. Furthermore, the patent strength θ and the imitation cost parameter f act as substitutes, because both f and θ include a cost dimension for the imitator, directly via f and indirectly via θ. As θ increases, the expected damages paid by the infringer increase and correspond to a higher cost of infringement. Therefore an increase in one of these cost parameters must be compensated by a decrease in the other one in order for the imitator to keep the same expected profits.

Remark: The fact that partial imitation does not occur in this case rests on two key properties of the Cournot duopoly equilibrium profit under a linear demand (and possibly other demand specifications). These are convexity in the firm's own marginal cost and submodularity with respect to own marginal cost and rival's marginal cost. These properties along with the convexity of the imitation cost makes the region of the (θ, f) space where the imitator's net profit $G_2^P(d_1, d_2, f, \theta)$ is convex expands as d_1 decreases. In this region, the only potential optimal behaviors for the imitator is to imitate fully or not imitate at all. If d_1 is sufficiently low (i.e. the innovation is sufficiently large), this region is

[8] It is different from the result obtained in Anton and Yao (2005) according to which imitation always occurs under the lost profit damage rule

sufficiently wide to encompass (high) values of f (for a fixed θ) for which non-imitation is optimal. Since non-imitation will still be optimal for higher values of f, partial imitation will never occur.

1.4.1.2. Small and medium innovations ($d_1 > 2c - a$)

This case is more complicated to analyze. Partial imitation is no more discarded. Indeed, three situations, namely full imitation, partial imitation and no imitation, may occur according to the values of the parameters (d_1, f, θ). The following proposition (proven in appendix A2) summarizes the follower's optimal strategy according to the values of the parameters f , θ and d_1 when $d_1 > 2c - a$.

Proposition 2 *Consider small and medium innovations ($d_1 > 2c - a$). For each value of d_1, there exist three separating functions in the (θ, f) space defined by $\alpha(d_1, \theta) = 4A(\theta)\frac{a-d_1}{c-d_1}$, $\beta(d_1, \theta) = 2A(\theta)\left(\frac{a-d_1}{c-d_1}\right)^2 - \frac{2}{9}\left(\frac{d_1-2c+a}{c-d_1}\right)^2$ and $\gamma(\theta) = \frac{8A(\theta)}{1-9A(\theta)}$ that delineate three regions:*

- If (f, θ, d_1) satisfy $f < Min(\alpha(d_1, \theta), \beta(d_1, \theta))$, then the follower fully imitates: $d_2^P(f, \theta, d_1) = d_1$

- If (f, θ, d_1) satisfy $\alpha(d_1, \theta) < f < \gamma(\theta)$, then the follower partially imitates: $d_2^P(d_1, f, \theta) = d_2^{int}(d_1, f, \theta)$

- If (f, θ, d_1) satisfy $\beta(d_1, \theta) < f < \alpha(d_1, \theta)$ or $f > Max(\alpha(d_1, \theta), \gamma(\theta))$, then the follower does not imitate: $d_2^P(f, \theta, d_1) = c$.

The functions $\alpha(d_1, \theta)$ and $\beta(d_1, \theta)$ are decreasing in the patent strength θ and in the innovation size $c - d_1$ and $\gamma(\theta)$ is decreasing in the patent strength θ. Moreover the equations in θ given by $\alpha(d_1, \theta) = \beta(d_1, \theta)$ and $\alpha(d_1, \theta) = \gamma(\theta)$ have the same solution $\theta_0(d_1) \in [0, 1[$ which means that the curves $f = \alpha(d_1, \theta)$, $f = \beta(d_1, \theta)$ and $f = \gamma(\theta)$ meet at a same point

$\theta_0(d_1)$ in the (θ, f) space for a given $d_1 > 2c - a$.

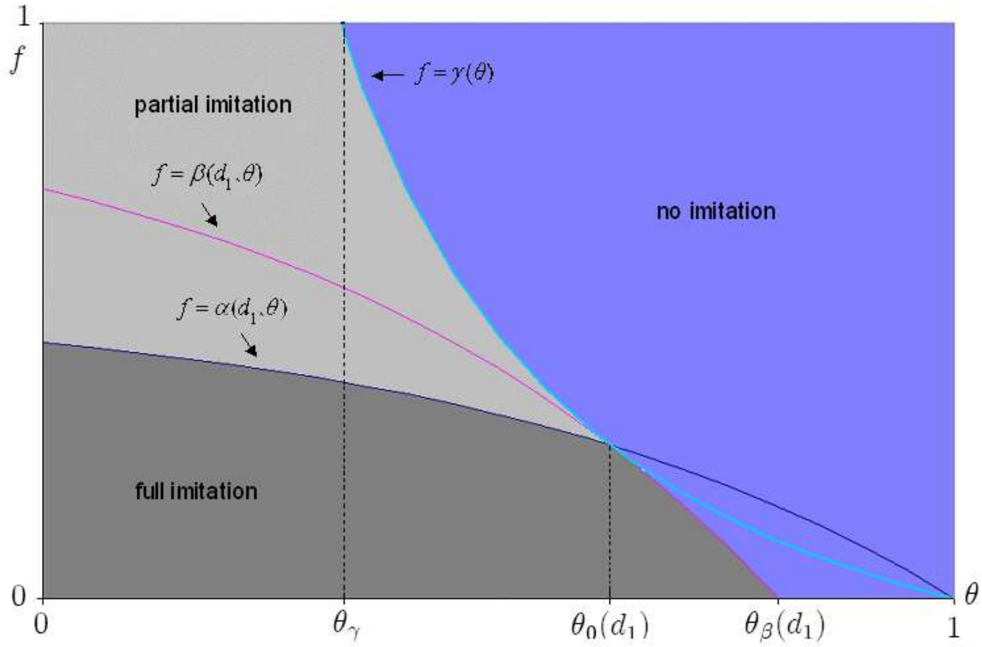

figure 2 $\left(2c - a < d_1 < \frac{9c-4a}{5}\right)$

This proposition is illustrated in figure 2 in which we assume that $2c - a < d_1 < \frac{9c-4a}{5}$. It follows from $\alpha(d_1, 0) = \frac{4}{9}\frac{a-d_1}{c-d_1}$ that $\alpha(d_1, 0) < 1 \Leftrightarrow d_1 < \frac{9c-4a}{5}$. Therefore it is worth distinguishing medium innovations $\left(2c - a < d_1 < \frac{9c-4a}{5}\right)$ from small innovations $(d_1 > 2c - a)$.

The imitator's choice of d_2 is affected by three variables: the cost parameter f, the patent strength θ and the innovation size $c - d_1$. Define θ_γ and $\theta_\beta(d_1)$ respectively as the solutions to the equations: $\gamma(\theta) = 1$ and $\beta(d_1, \theta) = 0$. For a given innovation d_1 such that $d_1 > 2c - a$, the effect of the cost imitation parameter f on the imitation level d_2 depends on the value of the patent strength θ in $[0, 1]$ in a specific way that we now describe.

When $\theta < \theta_\gamma$, the patent is very weak and imitation occurs whatever the cost imitation parameter f for two reasons. First, the risk of infringing a very weak patent is not sufficiently dissuasive: even if an infringement lawsuit occurs, damages will be paid with a very low probability θ. Second, imitation is not expensive enough to deter imitation of a small or medium innovation ($d_1 > 2c - a$). Therefore, imitation is either partial or full according to the imitation cost f. It is only partial if f is above the threshold $\alpha(d_1, \theta)$ and it is full if f is below this threshold.

When $\theta_\gamma \leq \theta \leq \theta_0(d_1)$, the patent is stronger and imitation becomes more expensive since the payment of damages occurs with a higher probability θ. Therefore, imitation may be either absent, partial or full, according to the imitation cost parameter value f. There is no imitation at all when f is higher than $\gamma(\theta)$. Imitation is only partial when f is below $\gamma(\theta)$ and above the previous threshold $\alpha(d_1, \theta)$ and is full when f is lower than $\alpha(d_1, \theta)$.

A third situation occurs when $\theta_0(d_1) < \theta < \theta_\beta(d_1)$. In this case, infringing is much more expensive because the patent will be upheld by the court with a higher probability θ. However, keeping the old technology $d_2 = c$ is also very detrimental to the follower. Therefore, imitation is either full or absent according to whether f is below or above the lower threshold $\beta(d_1, \theta)$.

Finally, a fourth situation occurs for the highest values of θ (i.e. $\theta > \theta_\beta(d_1)$). In this case, it is no longer profitable to imitate even when imitation is costless, because the patent protection is very strong. The imitation cost does not matter anymore and the presumptions that the patent will be upheld by the court and that imitation will be judged as being an infringement are so high that the patent protection plays fully its role against imitation. Since $\theta_\beta(d_1) < 1$, it is interesting to note that less than perfect protection is sufficient to deter imitation. Therefore, it is justified to refer to the value $\theta_\beta(d_1)$ as the *safe protection level.* A patent that protects against imitation does not need to be 100% perfect and the safe protection level depends on the importance of the innovation itself. As the innovation is less important (d_1 increases), the safe protection level $\theta_\beta(d_1)$ increases.

This important result suggests that smaller innovations require stronger protection, since they are likely to be imitated. This is a serious argument against the "one size fits all" protection principle.

The effect of d_1 over $d_2^P(f, \theta, d_1)$ for a given (θ, d_1) is interesting: as d_1 decreases, leading to an innovation involving a higher cost reduction, the partial imitation area increases because $\alpha(d_1, \theta)$ decreases, the full imitation area decreases because both $\alpha(d_1, \theta)$ and $\beta(d_1, \theta)$ decrease and the no-imitation area increases because $\beta(d_1, \theta)$ decreases.

1.4.2. Non-patented innovation

The follower's optimal imitation strategy under regime S can be simply derived from its optimal imitation strategy under regime P by taking $f = 1$ and $\theta = 0$. The next proposition summarizes our findings when the innovator chooses to use secrecy to protect its innovation.

Proposition 3 *Under the secrecy regime, the follower's optimal imitation strategy $d_1 \rightarrow d_2^S(d_1)$ is as follows:*

- If $d_1 \leq 2c - a$ then the follower does not imitate ($d_2^S(d_1) = c$) and is driven out of the market.

- If $2c - a < d_1 < \frac{9c-4a}{5}$ then the follower partially imitates ($d_2^S(d_1) = 9c - 4(a + d_1) < d_1$).

- If $\frac{9c-4a}{5} \leq d_1 \leq c$ then the follower fully imitates ($d_2^S(d_1) = d_1$).

Note that large innovations ($d_1 \leq 2c - a$) are never imitated under regime S while they are fully imitated under regime P when $f < \rho(d_1, \theta)$. The explanation of this rather counter-intuitive result simply derives from the previous remark that $\rho(d_1, 0) < 1$ for any $d_1 \leq$

$2c-a$.Then under the patent regime where some enabling knowledge is disclosed, the cost imitation parameter f may be so low (more specifically $f<\rho(d_1,\theta)$) that it may be profitable to incur the low imitation cost $fI(d_1)$ even when damages are paid with a high probability θ.

1.5. Protection stage

Which protection regime will the innovator choose given its anticipation of the follower's imitation level under each of the two regimes? To answer this question, we need to compare its equilibrium expected profit under the patent regime

$$\tilde{\Pi}_1^P(d_1,f,\theta)\equiv\Pi_1^P(d_1,d_2^P(d_1,f,\theta),\theta)$$

to its equilibrium profit under the secrecy regime

$$\tilde{\Pi}_1^S(d_1)\equiv\Pi_1^S(d_1,d_2^S(d_1))$$

Let us first determine the forces that drive the innovator's protection regime choice. Consider two (exogenous) imitation levels $d_2,d_2'\in[d_1,c]$. The difference $\Pi_1^P(d_1,d_2,\theta)-\Pi_1^S(d_1,d_2')$ can be decomposed in the following way:

$$\Pi_1^P(d_1,d_2,\theta)-\Pi_1^S(d_1,d_2')=\left(\Pi_1^P(d_1,d_2,\theta)-\Pi_1^S(d_1,d_2)\right)+\left(\Pi_1^S(d_1,d_2)-\Pi_1^S(d_1,d_2')\right)$$

The first term of this decomposition, namely the difference $\Pi_1^P(d_1,d_2,\theta)-\Pi_1^S(d_1,d_2)$, corresponds to what we call the *damage effect*. Given an imitation level d_2, the innovator can expect some damages if it patents its innovation, which is not the case if it chooses to keep it secret. Let us show that the damage effect is always (weakly) negative and (weakly) increasing in θ. This is equivalent to show that function $\theta \longrightarrow \Pi_1^P(d_1, d_2, \theta)$ is (weakly) increasing for any $d_2 \in [d_1, c]$. For any $d_2 < \min\left(c, \frac{a+d_1}{2}\right)$, one obtains:

$$\frac{\partial \Pi_1^P(d_1, d_2, \theta)}{\partial \theta} = \frac{6\theta(a-d_2)(d_2-d_1) + 6(a-2d_2+d_1) + 2(a-2d_2+d_1)^2}{3-\theta}$$

When $d_2 < \min\left(c, \frac{a+d_1}{2}\right)$, one can check that this derivative is strictly positive. In particular, this leads to :

$$\Pi_1^P(d_1, d_2, \theta) > \Pi_1^P(d_1, d_2, 0) = \Pi_1^S(d_1, d_2)$$

When $d_2 \geq \frac{a+d_1}{2}$, we have shown that $\Pi_1^P(d_1, d_2, \theta)$ does not depend on the parameter θ. In particular, $\Pi_1^P(d_1, d_2, \theta) = \Pi_1^P(d_1, d_2, 0) = \Pi_1^S(d_1, d_2)$. It follows that $\Pi_1^P(d_1, d_2, \theta) - \Pi_1^S(d_1, d_2) \geq 0$ for any $\theta \in [0, 1]$ and $d_2 \in [d_1, c]$.

Turn now to the second term. The difference $\Pi_1^S(d_1, d_2) - \Pi_1^S(d_1, d_2')$ corresponds to what we call the *competition effect.* The innovator's profits decline as it is more imitated: $\Pi_1^S(d_1, d_2)$ is an increasing function of d_2, which implies that the sign of $\Pi_1^S(d_1, d_2) - \Pi_1^S(d_1, d_2')$ is the same as the sign of $d_2 - d_2'$.

Thus, if the innovator anticipates that it will be less (or equally) imitated under the patent regime than under the secrecy regime then both the damage effect and the competition effect suggest the same protection regime, namely the patent regime. But if the innovator anticipates that it will be more imitated under the patent regime than under the secrecy regime, then the damage effect and the competition effect are opposite. The first one increases the incentives for the innovator to choose the patent regime while the second one increases the incentives to choose the secrecy regime. The following lemma, that summarizes and completes what precedes, is useful for the subsequent analysis:

Lemma 4 *If the innovator is less (or equally) imitated under regime P than under regime S then its optimal protection regime is the patent regime P. In particular, when the innovator anticipates that it will be fully imitated under the secrecy regime or that it will not be imitated at all under the patent regime, it chooses to patent its innovation.*

In order to determine the innovator's optimal protection regime, we distinguish the three cases that appeared in the imitation stage discussion.

Case 1: $d_1 < 2c - a$ (large innovations)

In this case, we know that the innovator is not imitated at all when it chooses to keep secrecy ($d_2^S(d_1) = c$). Its equilibrium expected profit under regime S is then given by:

$$\tilde{\Pi}_1^S(d_1) = \Pi_1^S(d_1, d_1) = \frac{(a - d_1)^2}{4}$$

We have also shown that under the patent regime, the innovator is fully imitated or not imitated at all, according to whether $f < \rho(d_1, \theta)$ or $f > \rho(d_1, \theta)$. Hence, its equilibrium expected profit under regime P is given by:

$$\tilde{\Pi}_1^P(d_1, \theta, f) = \begin{cases} \frac{(a-d_1)^2}{(3-\theta)^2} & if \quad f < \rho(d_1, \theta) \\ \frac{(a-d_1)^2}{4} & if \quad f > \rho(d_1, \theta) \end{cases}$$

implying that : $\tilde{\Pi}_1^P(d_1, \theta, f) < \tilde{\Pi}_1^S(d_1)$ if $f < \rho(d_1, \theta)$ and $\tilde{\Pi}_1^P(d_1, \theta, f) = \tilde{\Pi}_1^S(d_1)$ if $f > \rho(d_1, \theta)$.

This leads to the following proposition illustrated in figure 3.

Proposition 5 *If the innovation is large enough ($d_1 < 2c - a$), the innovator prefers to keep its innovation secret if $f < \rho(d_1, \theta)$ and is indifferent between patent protection and* *secrecy if* $f > \rho(d_1, \theta)$.

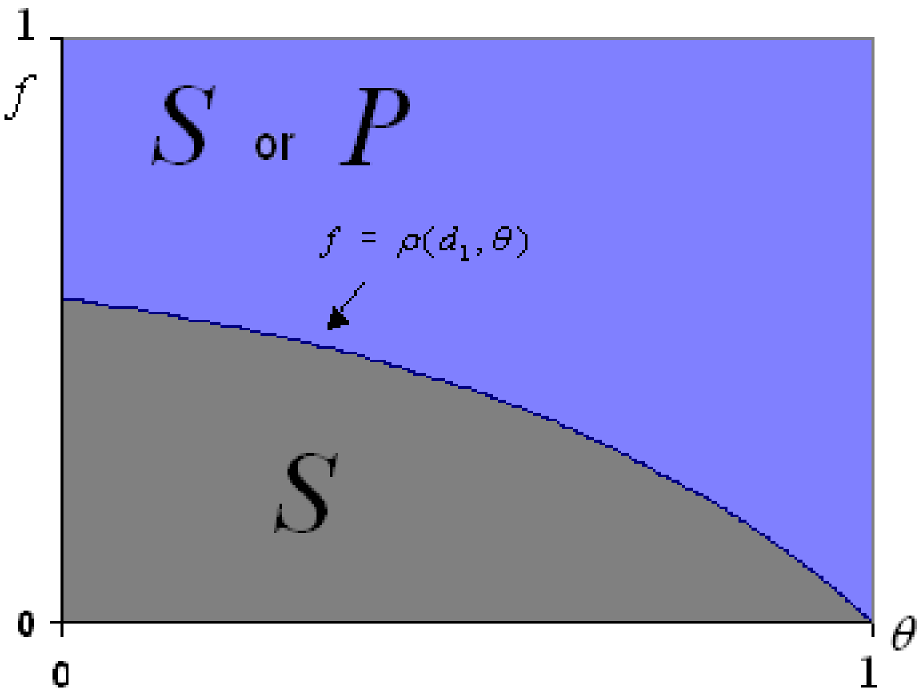

figure 3 $(d_1 < 2c - a)$

Hence, keeping secrecy is always an optimal strategy to the innovator when the innovation is large ($d_1 < 2c - a$). Such a choice may hinder the diffusion of large innovations, which may be detrimental to society since large innovations are likely to be those bringing breakthroughs and opening big opportunities for technological improvements (cumulative innovation). Proposition 5 suggests one way to make innovators patent their very inventive innovations: this may be induced either by reducing the level of compulsory disclosure which is equivalent, in our model, to increasing the value of the parameter f or by increasing the value of the expected damages.

Case 2: $2c-a<d_1<\frac{9c-4a}{5}$ (medium innovations)

In this case, the innovator is partially imitated under regime S ($d_2^S(d_1)=9c-4a-4d_1<d_1$) and its equilibrium expected profit under this regime is given by:

$$\tilde{\Pi}_1^S(d_1)=\Pi_1^S(d_1,9c-4a-4d_1)=(3c-a-2d_1)^2$$

Three subcases must be distinguished according to the value of $d_2^P(d_1,f,\theta)$ which affects $\tilde{\Pi}_1^P(d_1,\theta,f)=\Pi_1^P(d_1,d_2^P(d_1,f,\theta))$.

Subcase 2.1: $f<Min(\alpha(d_1,\theta),\beta(d_1,\theta))$

We know that for such values of parameters, the innovator is fully imitated under regime P ($d_2^P(d_1)=d_1$). Its equilibrium expected profits under regime P are :

$$\tilde{\Pi}_1^P(d_1,\theta,f)=\Pi_1^P(d_1,d_1)=\frac{(a-d_1)^2}{(3-\theta)^2}$$

Some straightforward calculations lead to:

$$\tilde{\Pi}_1^P(d_1,\theta,f)>\tilde{\Pi}_1^S(d_1)\Longleftrightarrow\theta>\tilde{\theta}(d_1)=\frac{9c-4a-5d_1}{3c-a-2d_1} \tag{1.8}$$

Hence the innovator chooses to keep its innovation secret if $\theta<\tilde{\theta}(d_1)$ and to patent it if $\theta>\tilde{\theta}(d_1)$. Note that $\tilde{\theta}(d_1)$ is a decreasing function of $d_1\in\left[2c-a,\frac{9c-4a}{5}\right]$ such that $\tilde{\theta}(\frac{9c-4a}{5})=0$ and $\tilde{\theta}(2c-a)=1$.

It is interesting to compare this new threshold $\tilde{\theta}(d_1)$ to the previously defined safe protection level $\theta_\beta(d_1)$. Since $A(\theta)$ is strictly decreasing, the comparison of $\theta_\beta(d_1)$ and $\tilde{\theta}(d_1)$ can be derived from the comparison of $A(\theta_\beta(d_1))$ and $A(\tilde{\theta}(d_1))$. From $\beta(d_1,\theta_\beta(d_1))=0$ we derive: $A(\theta_\beta(d_1))=\frac{(d_1-2c+a)^2}{9(a-d_1)^2}$ and, using the above expression of $\tilde{\theta}(d_1)$, one obtains: $A(\tilde{\theta}(d_1))=$

$\frac{3(d_1-2c+a)(3c-a-2d_1)}{(a-d_1)^2}$. Therefore we just need to compare $\frac{d_1-2c+a}{9}$ and $3\left(3c-a-2d_1\right)$. For any $d_1 \in \left]2c-a, \frac{9c-4a}{5}\right[$,we have $\frac{d_1-2c-a}{9} < \frac{a-c}{45}$ and $3\left(3c-a-2d_1\right) > 3(a-c)$ and so we get $A(\theta_\beta(d_1)) < A(\tilde{\theta}(d_1))$ which is equivalent to $\tilde{\theta}(d_1) < \theta_\beta(d_1)$.

This result shows that even if a medium innovation is expected to be fully imitated under the patent regime, a patent protection is still preferred by the innovator if the patent holder expects to recover the infringer's profit with a sufficiently high probability. Note that this probability $\tilde{\theta}(d_1)$ is lower than the safe protection level $\theta_\beta(d_1)$ previously defined. Therefore, patents will be filed even if their protection level is strictly lower than the safe protection level warranting perfect protection against imitation (see figure 4).

Subcase 2.2 : $\alpha\left(d_1, \theta\right) < f < \gamma(\theta)$

In this subcase, the innovator is partially imitated under regime P ($d_2^P(d_1) = d_2^{int}(d_1, f, \theta)$) and under regime S ($d_2^S(d_1) = 9c - 4(a + d_1) < d_1$). The following lemma (proven in appendix A3) compares these imitation levels under regime P and regime S.

Lemma 6 *When the innovator is partially imitated under both protection regimes, two cases arise:*

- If $f < 9A(\theta)$ then the innovator is more imitated under regime P than under regime S.

- If $f > 9A(\theta)$ then the innovator is more imitated under regime S than under regime P.

Finally, by combining the two previous lemmas, we reach the conclusion that when $f > 9A(\theta)$, the innovator chooses to patent its innovation since the damage effect and the competition effect go in the same direction. However, if $f < 9A(\theta)$ the damage effect and the competition effect are opposite. The following lemma is crucial in order to compare $\tilde{\Pi}_1^P\left(d_1, \theta, f\right)$ to $\tilde{\Pi}_1^S\left(d_1\right)$ in this case.

Lemma 7 *Along any curve $f = KA(\theta)$ in the (θ, f) space, where K is a strictly positive parameter, the innovator's equilibrium profit $\tilde{\Pi}_1^P(d_1, \theta, f)$ increases in the patent strength θ as long as partial imitation occurs.*

Proof. We showed in appendix A3 that the follower's level of imitation $d_2^{int}(d_1, f, \theta)$ depends on the parameters θ and f only through $\frac{f}{A(\theta)}$. This implies that $d_2^{int}(d_1, f, \theta)$ remains constant as one moves on the curve $f = KA(\theta)$. Then lemma 7 appears as a simple corollary of the result according to which the function $\theta \longrightarrow \Pi_1^P(d_1, d_2, \theta)$ is increasing, for given marginal costs d_1 and d_2, This result appeared when we introduced the damage effect. ■

Using this lemma, we derive the following result (proven in appendix A4).

Lemma 8 *For medium innovations ($2c - a < d_1 < \frac{9c-4a}{5}$), when $\alpha(d_1, \theta) < f < 9A(\theta)$, there exists a threshold $\lambda(d_1, \theta)$ decreasing in the patent strength θ such that the innovator keeps its innovation secret if $f < \lambda(\theta, d_1)$ and $\theta < \tilde{\theta}(d_1)$ and patents it if $f > \lambda(\theta, d_1)$ or $\theta > \tilde{\theta}(d_1)$. The threshlod function $\lambda(d_1, \theta)$ satisfies the following two conditions : $\lambda(d_1, 0) = 1$ and $\lambda(d_1, \tilde{\theta}(d_1)) = \alpha(d_1, \tilde{\theta}(d_1))$.*

The first condition states that the innovator is indifferent between patenting and keeping secrecy when $\theta = 0$ and $f = 1$ and the second that it is indifferent between these two regimes when $\theta = \tilde{\theta}(d_1)$ and $f = \alpha(d_1, \tilde{\theta}(d_1)$ which is consistent with our previous findings (see figure 4).

Subcase 2.3: $\beta(d_1, \theta) < f < \alpha(d_1, \theta)$ or $f > Max(\alpha(d_1, \theta), \gamma(\theta))$

In this subcase, the innovator is not imitated at all under the patent regime ($d_2^P(d_1) = d_1$).

We derive from lemma 4 that the innovator's optimal protection regime is regime (P).

Finally, the following proposition summarizes the case $2c - a < d_1 < \frac{9c-4a}{5}$.

Proposition 9 *For medium process innovations* $\left(2c - a < d_1 < \frac{9c-4a}{5}\right)$, *there exist a threshold function* $\tilde{\theta}(d_1)$ *decreasing in the innovation size* $c - d_1$ *and a threshold function* $\lambda(d_1, \theta)$ *decreasing in the patent strength* θ *such that :*

- *If* $\theta < \tilde{\theta}(d_1)$ *and* $f < \lambda(d_1, \theta)$ *then the innovator chooses the secrecy regime*

- *If* $\theta > \tilde{\theta}(d_1)$ *or* $f > \lambda(d_1, \theta)$ *then the innovator chooses the patent regime.*

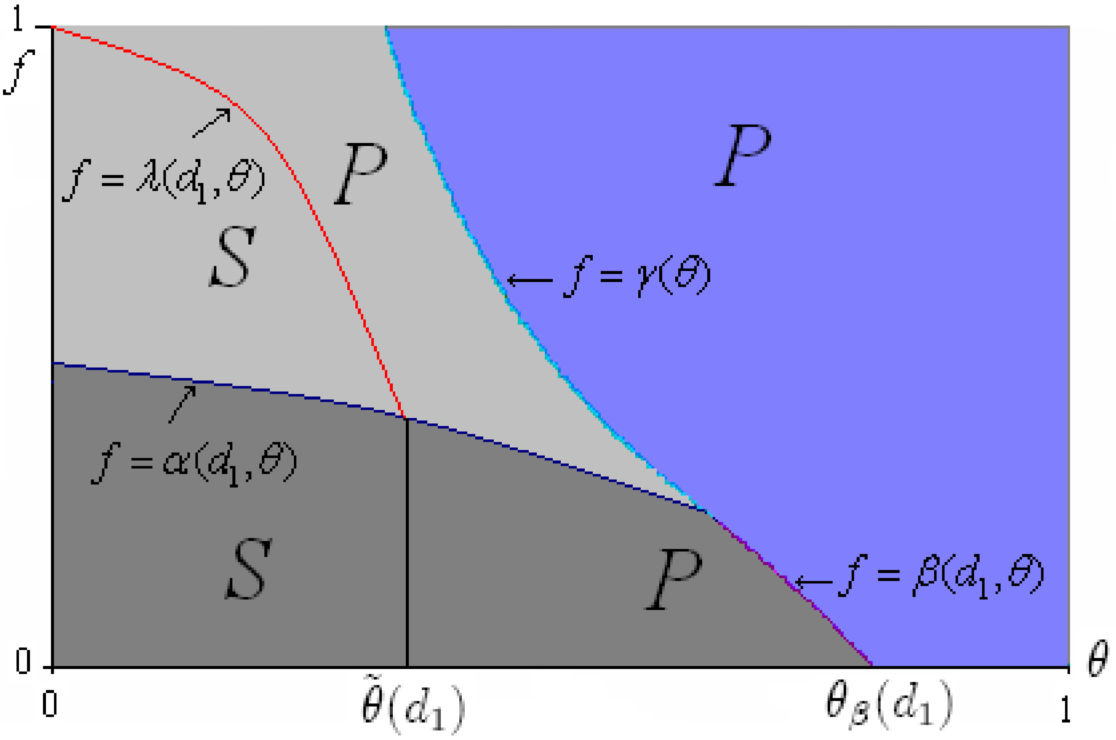

*figure*4 $\left(2c - a < d_1 < \frac{9c-4a}{5}\right)$

This proposition, illustrated in figure 4, can be interpreted as follows. When the patent is strong enough ($\theta > \tilde{\theta}(d_1)$) the innovator always chooses to patent its innovation whatever the imitation cost parameter f. This means that the imitation cost allowed by disclosure does not matter anymore when the patent strength is above $\tilde{\theta}(d_1)$. In particular, it patents its innovation even if disclosure makes the imitation costless. This result does not hold for weak patents ($\theta < \tilde{\theta}(d_1)$). In this case, the competition effect and the damage effect may go in opposite directions leading to different protection regimes according to whether the disclosure effect of patenting is high enough ($f < \lambda(d_1, \theta)$) or not. The effect of the innovation size on the secrecy region is clear: since $\tilde{\theta}(d_1)$ decreases as the innovation size decreases (d_1 increases), the corresponding area shrinks as the innovation is smaller.

Case 3: $d_1 > \frac{9c-4a}{5}$ (small innovations)

In this case, the innovator will be fully imitated if it chooses regime S ($d_2^S(d_1) = d_1$). Therefore, according to lemma 4, the innovator's optimal protection regime is the patent regime (P).

Proposition 10 *Small innovations ($d_1 > \frac{9c-4a}{5}$) are always patented.*

This result can be explained in the following way : since small innovations are fully imitated under secrecy, patenting is preferred for two reasons. First, it may deter imitation leading to higher market profits. Second, even when the patent strength and the disclosure effect are such that imitation cannot be deterred, it allows the innovator to expect some damages compensating its market profit loss due to imitation, which is not the case under secrecy regime.

We develop now an overall discussion of our results by comparing them to those of Anton and Yao (2004) and embedding them in a broader perspective.

The latter proposition states that inventors of small process innovations always prefer protection induced by patent to the protection conferred by a trade secret. This result is a reminder of the "little patents" expression coined by Anton and Yao (2004). However, the argument behind this common result is different in our model. Under the secrecy regime, small process innovations are fully imitated for two reasons that reinforce one another.

First, imitating a small innovation is not very costly and second, there is no threat of an infringement lawsuit when the innovation is kept secret. Under patent protection, such a threat exists and it may overrun the benefits that the infringer expects from imitating the leader. Note that in our model, small innovations may be imitated under the patent regime while this does not occur in Anton and Yao.

For large process innovations, our results are similar to the " big secrets" characterization obtained in Anton and Yao (2004). Large process innovations are never imitated when they are kept secret, while the enabling knowledge disclosed by a patent may reduce the imitation cost in a way that renders their imitation attractive. This classical tradeoff in the economics of patents explains why "big secrets" are preferred to "big patents". Note however that our model does not totally discard the possibility of patenting some large process innovations. This may occur when their imitation under the patent regime is too costly. In this case we have shown that the innovator is indifferent between secret and patent, because in both cases, the innovation is not imitated.

Finally, it is for medium process innovations that our results significantly differ from those of Anton and Yao. We have shown that keeping a medium process innovation secret does not avoid imitation. It is precisely for medium process innovations that partial innovation occurs under the secrecy regime. However, under the patent regime, imitation may be either absent, partial and total. We have also shown that the innovator may patent or keep secrecy while in Anton and Yao, medium process innovations are always patented and partially disclosed.

1.6. Licensing agreements as alternatives to litigation

In this section, we allow for licensing agreements between the innovator and the follower. Since our purpose is not to study all the possible agreements that may emerge, we restrict our attention to the simple case of process innovations leading to a small cost reduction, i.e. $d_1 > \frac{9c-4a}{5}$. We know from the previous section that those innovations are always patented. We analyze licensing agreements between the innovator and an imitator that avoid litigation to be completed until the court's decision.

We study two-part tariff licenses (r, F) where r is a royalty rate and F a fixed fee. Let us first examine the equilibrium outcomes when the innovator and the follower agree on a license (r, F). Gross profits can be written as

$$\Pi_1^L(x_1, x_2, d_1, r, F) = (a - x_1 - x_2 - d_1)\, x_1 + r x_2 + F$$

$$\Pi_1^L(x_1, x_2, d_1, r, F) = (a - x_1 - x_2 - (d_1 + r))\, x_2 - F$$

An equilibrium of the competition stage under the license regime (L) is given by

$$x_1^L = \frac{a - d_1 + r}{3}$$

$$x_2^L = \frac{a - d_1 - 2r}{3}$$

and leads to the following equilibrium price:

$$p^L = \frac{a + 2d_1 + r}{3}$$

We assume hereafter that $0 \leq r \leq \frac{a-d_1}{2}$. This assumption does not entail any loss of generality since any royalty rate such that $r > \frac{a-d_1}{2}$ leads to the same boundary solution as $r = \frac{a-d_1}{2}$:

$$x_1^L = \frac{a - d_1}{2}$$

$$x_2^L = 0$$

Equilibrium gross profits are given by:

$$\tilde{\Pi}_1^L(d_1, r, F) = \left(\frac{a - d_1 + r}{3}\right)^2 + r\frac{a - d_1 - 2r}{3} + F \tag{1.9}$$

$$\tilde{\Pi}_2^L(d_1, r, F) = \left(\frac{a - d_1 - 2r}{3}\right)^2 - F \tag{1.10}$$

1.6.1. A benchmark: the non-collusive agreement

First, we address the following question: does there exist a license (r, F) such that the profits under the license regime replicate the profits under the patent regime (without licensing), i.e:

$$\tilde{\Pi}_1^L(d_1, r, F) = \tilde{\Pi}_1^P(d_1, \theta)$$

and

$$\tilde{\Pi}_2^L(d_1, r, F) = \tilde{\Pi}_2^P(d_1, \theta)$$

Since we do not set any restriction on the fixed fee F (in particular, we allow F to be negative), it is clear that such a question amounts to the existence of a royalty rate r such that the industry profits $\tilde{\Pi}^L(d_1, r) = \tilde{\Pi}_1^L(d_1, r, F) + \tilde{\Pi}_2^L(d_1, r, F)$ under the license regime replicate the industry profits under the patent regime without licensing, i.e:

$$\tilde{\Pi}^L(d_1, r) = \tilde{\Pi}_1^P(d_1, \theta) + \tilde{\Pi}_2^P(d_1, \theta) \tag{1.11}$$

which can be rewritten as:

$$\frac{1}{9}[-r^2 + (a - d_1)r + 2(a - d_1)^2] = \frac{2 - \theta}{(3 - \theta)^2}(a - d_1)^2 \tag{1.12}$$

Solving this equation, we find a unique solution :

$$\tilde{r}(\theta) = \frac{\theta}{3 - \theta}(a - d_1) \tag{1.13}$$

The equilibrium price is then given by:

$$\tilde{p}^L(\theta) = \frac{a + d_1(2 - \theta)}{3 - \theta}$$

The royalty rate $\tilde{r}(\theta)$ and equilibrium price $\tilde{p}^L(\theta)$ are increasing in θ. We can now derive the fixed fee $\tilde{F}(\theta)$ from the equation $\tilde{\Pi}_2^L(d_1, \tilde{r}(\theta), \tilde{F}(\theta)) = \tilde{\Pi}_2^P(d_1, \theta)$:

$$\tilde{F}(\theta) = \frac{\theta(\theta - 1)}{(3 - \theta)^2}(a - d_1)^2$$

Thus, we obtain the following properties of the function $\tilde{F}$:

- $\tilde{F}(0) = \tilde{F}(1) = 0$.

- $\tilde{F}(\theta) < 0$ for any $\theta \in]0, 1[$.

- $\tilde{F}(\theta)$ is strictly decreasing over $\left[0, \frac{3}{5}\right]$ and strictly increasing over $\left[\frac{3}{5}, 1\right]$.

Note that the license $(\tilde{r}(\theta), \tilde{F}(\theta))$ not only replicates the expected profits of the innovator and the follower under regime P (without licensing) but also induces the same equilibrium outputs and then the same equilibrium price. It is indeed easy to check that :

$$\tilde{x}_1^L(\theta) = \frac{a - d_1 + \tilde{r}(\theta)}{3} = \frac{(1-\theta)(a-d_1)}{3-\theta} = x_1^P(\theta)$$

and

$$\tilde{x}_2^L(\theta) = \frac{a - d_1 - 2\tilde{r}(\theta)}{3} = \frac{a-d_1}{3-\theta} = x_2^P(\theta)$$

which obviously lead to

$$\tilde{p}^L(\theta) = p^P(\theta)$$

where $p^P(\theta)$ is the equilibrium price under the patent regime without licensing.

Thus, the fact that competition stage outcomes are identical under both regimes (patent under the shadow of infringement and license defined by $(\tilde{r}(\theta), \tilde{F}(\theta))$), allows us to consider the license regime defined by $(\tilde{r}(\theta), \tilde{F}(\theta))$ as a *benchmark* to which we could compare the outcomes of the licensing agreements that are likely to emerge.

Consider Shapiro's criterion that a settlement, e.g. a licensing agreement, should not be allowed unless it does not make consumers worse off relative to the situation where litigation occurs. The non collusive agreement we fully characterized above meets this requirement since it does not result in a higher price. However, there exists a fundamuntal difference between Shapiro (2003) and our setting: competition occurs and prices are set in Shapiro's model *after* the issues of validity and infringement are resolved while in our setting, it happens *before* (competition occurs under the shadow of infringement). Due to this difference, it is possible to design a settlement that improves firm's profits without decreasing consumers' *expected* surplus in Shapiro (2003) while this is not possible in our setting.[9]

1.6.2. *The most collusive agreement*

One of the agreements which are more likely to emerge is a two-part tariff license thet maximizes the firms' joint profits. The part of the industry profits allocated to each firm is then determined by the fixed fee F.

Consider such a licensing agreement, denoted by $\left(\hat{r}(\theta), \hat{F}(\theta)\right)$. Since $\hat{r}(\theta)$ is defined by $\hat{r}(\theta) = Arg \max_{0 \leq r \leq \frac{a-d_1}{2}} \Pi^L(d_1, r)$, the royalty $\hat{r}(\theta)$ does not depend on the patent strength θ. Moreover, it is easy to see that $\Pi^L(d_1, r)$ is an increasing function of r over interval $\left[0, \frac{a-d_1}{2}\right]$. This yields:

$$\hat{r}(\theta) = \hat{r} = \frac{a-d_1}{2} = \tilde{r}(1), \ \forall \theta \in [0,1]$$

leading to the price:

$$\hat{p}(\theta) = \hat{p} = \frac{a+d_1}{2} = \tilde{p}^L(1) \geq \tilde{p}^L(\theta), \ \forall \theta \in [0,1]$$

This result means that the patent strength is no longer reflected by the price paid by consumers. In particular, low quality patents which generate lower prices when litigated, generate the same price as would high quality patents. This obviously harm the consumers. Note that $x_2^L = \frac{a-d_1-2r}{3} = 0$ when $r = \hat{r}$. Hence, when the follower accepts the license $\left(\hat{r}, \hat{F}(\theta)\right)$ it implicitly accepts to stay out of the market and the industry profits are then captured by the patentee. Nevertheless, the innovator transfers a part of these monopoly profits to the licensee through the negative fixed fee $\hat{F}(\theta)$. In other words, the license

[9] In our model, the only agreement that is acceptable by both parties and does not harm consumers is the non collusive agreement determined above.

$\left(\hat{r}, \hat{F}(\theta)\right)$ is equivalent to an agreement whereby the innovator pays its competitor to stay out of the market. Hence this illustrates in our framework the issue, raised by Shapiro (2003), that licensing agreements used to settle patent litigation could actually be used as a way to reach anticompetitive outcomes.

The patentee's and the licensee's profits when they agree on the license $\left(\hat{r}, \hat{F}(\theta)\right)$ are given by:

$$\Pi_1^L(d_1, \theta) = \frac{(a - d_1)^2}{4} + \hat{F}(\theta)$$

$$\Pi_2^L(d_1, \theta) = -\hat{F}(\theta)$$

This type of agreement is accepted by both the innovator and the follower as long as the fixed fee $\hat{F}(\theta)$ is in the interval $\left[\Pi_1^P(d_1, \theta) - \frac{(a-d_1)^2}{4}, -\Pi_2^P(d_1, \theta)\right]$, that is:[10]

$$\hat{F}(\theta) \in \left[-A(\theta)\,(5 - \theta)\,(a - d_1)^2\,, -A(\theta)\,(a - d_1)^2\right] \tag{1.14}$$

The possibility that such two-part tariffs involving negative fees emerge in licensing agreements is a big concern for competition authorities. We know that in the pharmaceutical industry, agreements of this kind have been actually used by some patent holders in their negociations with generic challengers under the Hatch-Waxman Act. They obviously harm consumers and this is why patent settlements, which take the form of licensing agreements, must be under the scrutiny of competition authorities (Shapiro, 2003).

[10] The license $(\hat{r}, F_0(\theta))$ where $F_0(\theta) = -A(\theta)\,(5 - \theta)\,(a - d_1)^2$ is the optimal license, from the innovator's perspective, among all the licenses that maximize industry profits. It is likely to emerge if the innovator has a "take it or leave it" bargaining power. Indeed, it is clear that with such a bargaining power the innovator will pay its competitor the minimum amount that makes it accept to stay out the market.

1.7. Conclusion

Departing from the usual convention that patents are perfect forms of protection opens a lot of research avenues. One of the most important issues is to know under what conditions patent protection is preferred to secrecy. Our model provides a theoretical answer to this question for a process innovation. For each class of cost reduction (small, medium and large) we have obtained specific results. First, we have determined the imitation level in each regime. Second, in the space of the two key parameters (patent strength and relative imitation cost) we have derived the partition that delineates areas where one protection regime dominates the other. How can one use these results for a policy purpose? This is an interesting and complex issue for which we suggest preliminary insights. Consider the relative imitation cost. In a world where patent design is independent of the invention, particularly concerning the same compulsory disclosure for all patents, it seems very hard to determine *a priori* what would be the value of the imitation cost parameter. One can simply reach a rather vague idea of the secrecy effectiveness of the invention, that leads to an idiosyncratic characterization covering a large spectrum of possibilities, running from the "naked idea" case to the "perfectly hidden idea" case. This type of assessment would depend on some priors on whether the invention could be more or less easily discovered by reverse-engineering. But in a world where a patent is not designed around the "one size fits all" principle, some flexibility could be introduced by allowing each innovator to choose a patent inside a menu of characteristics. For instance an innovator may have to choose between a patent with strong property rights and high disclosure requirements and a patent with weak property rights and low disclosure requirements. If an incentive mechanism built around this principle could be achieved, it would be an appropriate answer to the rather disappointing result that "little patents and big secrets" are the preferred forms of protection. Small innovations could be easily imitated because their rights are weak. Large innovations could be patented because their rights are strong. In both cases, diffusion of innovation would be enhanced. The construction of such an incentive mechanism is left for future research.

Our model analyzes also the licensing agreements between a patent holder and a competitor. Such agreements avoid the litigation to go until completion. One of the possible consequences of a patent settlement as an alternative to a trial raises some concerns. The royalty rate paid by the licensee does not depend on the patent strength as a natural benchmark would command. Licensing very bad quality patents may occur with as high royalty rate as if the patent were full-proof. Moreover, the patentee pays a fixed fee to the licensee to compensate its loss in the market. While the two parties maximize their joint profits, it is clear that such a settlement harms consumers and creates a big concern for society. Shapiro (2003) and Farrell and Shapiro (2005) reach the same conclusion by using quite different models.

Finally, while some economists (Ayres and Klemperer, 1999) find that probabilistic rights open welfare improving opportunities (entry occurs under the shadow of punishment) it is also important to stress some of their negative consequences. Adopting trade secrecy for large inventions may reduce the diffusion of innovation. Moreover, patent settlements of the sort examined in this paper are detrimental to society. This is one reason why patent quality is probably one of the most challenging issues to which we are now confronted.

1.8. References

Anton, J.J. and D. A. Yao, 2004, "Little patents and big secrets: managing intellectual property", *Rand Journal of Economics* 35, 1-22.

Anton, J.J. and D. A. Yao, 2005, "Finding "Lost Profits": an equilibrium analysis of patent infringement damages", W.P. Duke University.

Anton, J.J., H. Greene and D. A. Yao, 2005, "Policy implications of weak patent rights", mimeo.

Bebchuk, L.A., "Litigation and settlement under imperfect information", *Rand Journal of Economics*, 15, 404-415.

Chang, H. F. 1995. "Patent Scope, Antitrust Policy, and Cumulative Innovation," *Rand Journal of Economics* 26, 34-57.

Cohen, W.M, R.R. Nelson and J.P. Walsh, 2000, "Protecting their intellectual assets: appropriability conditions and why US manufacturing firms patent or not", NBER W.P. 7552.

Crampes, C, 1986, "Les inconvénients d'un dépôt de brevet pour une entreprise innovatrice", *L'Actualité Economique, Revue d'Analyse Economique*, 62, 4, 521-534.

Crampes, C. and C. Langinier, Litigation and settlement in patent infringement cases,*Rand Journal of Economics*, 33(2), 258-274.

Encaoua, D. and A. Hollander, 2004, "Competition policy and innovation", *Oxford Review of Economic Policy* 18, 63-79.

Encaoua, D., D. Guellec and C. Martinez, 2005, "Patent systems for encouraging innovation: Lessons from economic analysis", W.P., University Paris I.

Farrell, J. and C. Shapiro, 2007, "How strong are weak patents?", W. P. University of

California at Berkeley, forthcoming American Economic Review.

Gallini, N.T., 1992, "Patent policy and costly innovation", *Rand Journal of Economics* 23, 52-63.

Green, J. and S. Scotchmer, 1995, "On the division of profit in sequential innovation", *Rand Journal of Economics* 26, 20-23.

Horstman, I., G.M. MacDonald and A. Slivinski, 1985, "Patents as information transfer mechanisms: To patent or (Maybe) not to patent", *Journal of Political Economy* 95-5, 837-858.

Lemley, M.A. and C. Shapiro, 2005, "Probabilistic patents", *Journal of Economic Perspectives* 19-2, 75-98.

Levin, R.C., A.K. Klevorick, R.R. Nelson and S.G. Winter, 1987, "Appropriating the returns from industrial R&D", *Brooking Papers on Economic Activity*, 783-820.

Mansfield, E., 1986, "Patents and innovation: an empirical survey", *Management Science* 32, 173-181.

Merges, R., 1999, "As many as six impossible patents before breakfast: property rights for business concepts and patent system reform", *Berkeley High Technology Law Journal* 14, 577-615.

Meurer, M., 1989, "The settlement of patent litigation", *Rand Journal of Economics*, 20, 77-91.

Pakes, A. and Z. Griliches, "Patents and R&D at the firm level: a first look", *Economics Letters* 5-4, 377-381.

Scherer, F.M., 1965, "Market structure, opportunity and the output of patented inventions", *American Economic Review* 55, 1097-1125.

Scherer, F.M., 1967, "Market structure and the employment of scientists and engineers", *American Economic Review* 57, 524-531.

Scherer, F.M., 1983, "The propensity to patent", *International Journal of Industrial Organization* 1, 107-128.

Shapiro, C., 2003, "Antitrust analysis of patent settlements between rivals", *Rand Journal of Economics* 34-2, 391-411.

Scotchmer, S., 2004, *Innovation and Incentives*, The MIT Press, Cambridge, Ma.

Scotchmer, S. and J. Green, 1990, "Novelty and disclosure in patent law", *Rand Journal of Economics* 21, 131-146.

1.9. Appendix

A1. Proof of Proposition 1

Since $\Pi_2^P(d_1, d_2, f, \theta) = 0$ for any $d_2 \in \left[\frac{a+d_1}{2}, c\right]$, it follows from the fact that imitation is costly that the follower's best imitation level over this interval is $d_2 = c$:

$$\underset{d_2 \in \left[\frac{a+d_1}{2}, c\right]}{Arg \max}\, G_2^P(d_1, d_2, f, \theta) = c \tag{1.15}$$

which implies that the follower's optimal imitation level is necessarily equal to either c or $\underset{d_2 \in \left[d_1, \frac{a+d_1}{2}\right]}{Arg \max}\, G_2^P(d_1, d_2, f, \theta)$.

In order to determine the maximum value of $G_2^P(d_1, d_2, f, \theta)$ over the interval $\left[d_1, \frac{a+d_1}{2}\right]$, we must distinguish two cases:

Case 1 : $f < 8A(\theta)$

The function $d_2 \to G_2^P(d_1, d_2, f, \theta)$ is convex over $\left[d_1, \frac{a+d_1}{2}\right]$ since $G_2^P(d_1, d_2, f, \theta) =$

$H(d_1, d_2, f, \theta)$ over this interval. Moreover, it is straightforward to show that $d_2^{int}(d_1, f, \theta) < \frac{a+d_1}{2}$. Then, there are two possibilities according to whether $d_2^{int}(d_1, f, \theta) > d_1$ or $d_2^{int}(d_1, f, \theta) \leq d_1$.

- If $d_2^{int}(d_1, f, \theta) > d_1$ then $d_2 \to G_2^P(d_1, d_2, f, \theta)$ is decreasing over the interval $\left[d_1, d_2^{int}(d_1, f, \theta)\right]$ and is increasing over the interval $\left[d_2^{int}(d_1, f, \theta), \frac{a+d_1}{2}\right]$, which entails that $d_2 \to G_2^P(d_1, d_2, f, \theta)$ reaches its maximum value over the interval $\left[d_1, \frac{a+d_1}{2}\right]$ at either $d_2 = d_1$ or $d_2 = \frac{a+d_1}{2}$.

- If $d_2^{int}(d_1, f, \theta) < d_1$ then $d_2 \to G_2^P(d_1, d_2, f, \theta)$ is increasing over the interval $\left[d_1, \frac{a+d_1}{2}\right]$ which implies that it reaches its maximum value at $d_2 = c$.

The crucial point is that in both cases, $d_2^P(d_1, f, \theta) \in \left\{d_1, \frac{a+d_1}{2}, c\right\}$. Since we know that the follower prefers not to imitate rather than imitate at a level $d_2 = \frac{a+d_1}{2}$, it is sufficient to compare $G_2^P(d_1, d_1, f, \theta)$ to $G_2^P(d_1, c, f, \theta) = 0$ in order to determine $d_2^P(d_1, f, \theta)$. Hence, two subcases arise:

- If $f < 2A(\theta)\left(\frac{a-d_1}{c-d_1}\right)^2$ then $G_2^P(d_1, d_1, f, \theta) > 0$ which results in $d_2^P(d_1, f, \theta) = d_1$ (full imitation)

- If $2A(\theta)\left(\frac{a-d_1}{c-d_1}\right)^2 < f < 8A(\theta)$ then $G_2^P(d_1, d_1, f, \theta) < 0$ which results in $d_2^P(d_1, f, \theta) = c$ (no imitation).

Case 2 : $f > 8A(\theta)$

In this case, $d_2^{int}(d_1, f, \theta) > c > \frac{a+d_1}{2}$ and $d_2 \to G_2^P(d_1, d_2, f, \theta)$ is concave over the interval $\left[d_1, \frac{a+d_1}{2}\right]$. Then, the function $d_2 \to G_2^P(d_1, d_2, f, \theta)$ is increasing over the interval $\left[d_1, \frac{a+d_1}{2}\right]$, which results in $\underset{d_2 \in \left[d_1, \frac{a+d_1}{2}\right]}{Arg \max}\ G_2^P(d_1, d_2, f, \theta) = \frac{a+d_1}{2}$. Since $G_2^P(d_1, d_2, f, \theta) = 0$ for any $d_2 \geq \frac{a+d_1}{2}$, we conclude that $d_2^P(d_1, f, \theta) = c$ (no imitation).

A2. Proof of Proposition 2

The imitator must compare the maximal net profit it can get when it imitates, i.e. $\sup_{d_2 \in [d_1, c[} G_2^P(d_1, d_2, f, \theta) =$

$\sup_{d_2 \in [d_1,c[} H(d_1,d_2,f,\theta)$, to the net profit it derives from keeping its old technology, ie. $G_2^P(d_1,c,f,\theta) = \frac{(d_1-2c-a)^2}{9}$.Two cases must be distinguished:

Case 1 : $f < 8A(\theta)$

In this case, $d_2^{int}(d_1,f,\theta) > c$ and $d_2 \rightarrow G_2^P(d_1,d_2,f,\theta)$ is convex over the interval $[d_1,c[$, which entails that $d_2 \rightarrow G_2^P(d_1,d_2,f,\theta)$ is decreasing over the interval $[d_1,c[$, and results in $\sup_{d_2 \in [d_1,c[} G_2^P(d_1,d_2,f,\theta) = G_2^P(d_1,d_1,f,\theta)$ which has to be compared to $G_2^P(d_1,c,f,\theta)$. This leads us to distinguish two subcases.

Define $\beta(d_1,\theta) = 2A(\theta)\left(\frac{a-d_1}{c-d_1}\right)^2 - \frac{2}{9}\left(\frac{d_1-2c+a}{c-d_1}\right)^2$.

- If $f < \beta(d_1,\theta)$ then $G_2^P(d_1,d_1,f,\theta) > G_2^P(d_1,c,f,\theta)$ which results in $d_2^P(d_1,f,\theta) = d_1$(full imitation).

- If $f > \beta(d_1,\theta)$ then $G_2^P(d_1,d_1,f,\theta) < G_2^P(d_1,c,f,\theta)$ which results in $d_2^P(d_1,f,\theta) = c$ (no imitation).

Case 2 : $f > 8A(\theta)$

In this case, $d_2^{int}(d_1,f,\theta) < c$ and $d_2 \rightarrow G_2^P(d_1,d_2,f,\theta)$ is concave over the interval $[d_1,c[$.

Two subcases must be distinguished :

- If $d_2^{int}(d_1,f,\theta) < d_1$ then $d_2 \rightarrow G_2^P(d_1,d_2,f,\theta)$ is decreasing over the interval $[d_1,c[$, which implies that $\sup_{d_2 \in [d_1,c[} G_2^P(d_1,d_2,f,\theta) = G_2^P(d_1,d_1,f,\theta)$.

- If $d_2^{int}(d_1,f,\theta) > d_1$ then $d_2 \rightarrow G_2^P(d_1,d_2,f,\theta)$ reaches its maximum over $[d_1,c[$ at $d_2 = d_2^{int}(d_1,f,\theta)$: $\sup_{d_2 \in [d_1,c[} G_2^P(d_1,d_2,f,\theta) = G_2^P(d_1,d_2^{int}(d_1,f,\theta),f,\theta)$.

Consider the condition $d_2^{int}(d_1,f,\theta) < d_1$. It is straightforward to show that this inequality can be rewritten as :

$$f < \alpha(d_1,\theta) = \frac{4(a-d_1)}{c-d_1}A(\theta)$$

Then, the two previous subcases can be written as:

- If $8A(\theta) < f < \frac{4(a-d_1)}{c-d_1}A(\theta)$ then $\sup\limits_{d_2\in[d_1,c[} G_2^P(d_1,d_2,f,\theta) = G_2^P(d_1,d_1,f,\theta)$ which has to be compared to $G_2^P(d_1,c,f,\theta)$ (this has been previously done).

- If $f > \frac{4(a-d_1)}{c-d_1}A(\theta)$ then $\sup\limits_{d_2\in[d_1,c[} G_2^P(d_1,d_2,f,\theta) = G_2^P(d_1,d_2^{int}(d_1,f,\theta),f,\theta)$ which has to be compared to $G_2^P(d_1,c,f,\theta)$. Comparing these two terms is equivalent to comparing f to the threshold $\gamma(\theta) = \frac{8}{\frac{(3-\theta)^2}{1-\theta}-9}$. More specifically:

- If $\frac{4(a-d_1)}{c-d_1}A(\theta) < f < \gamma(\theta)$ then $G_2^P(d_1,d_2^{int},f,\theta) > G_2^P(d_1,c,f,\theta)$ which leads to $d_2^P(d_1,f,\theta) = d_2^{int}(d_1,f,\theta)$ (partial imitation).

- If $f > \gamma(\theta)$ then $G_2^P(d_1,d_2^{int},f,\theta) < G_2^P(d_1,c,f,\theta)$ which leads to $d_2^P(d_1,f,\theta) = c$ (no imitation).

Let us now show that the equations $\alpha(d_1,\theta) = \beta(d_1,\theta)$ and $\alpha(d_1,\theta) = \gamma(\theta)$ have the same solution $\theta_0(d_1)$ over the interval $[0,1[$,which means that the curves $f = \alpha(d_1,\theta)$, $f = \beta(d_1,\theta)$ and $f = \gamma(\theta)$ meet at the same point.

Some straightforward computations show that the equation $\alpha(d_1,\theta) = \beta(d_1,\theta)$ is equivalent to the equation:

$$A(\theta) = \frac{1}{9}\frac{d_1-2c+a}{a-d_1}$$

Therefore, the equations $\alpha(d_1,\theta) = \beta(d_1,\theta)$ and $\alpha(d_1,\theta) = \gamma(\theta)$ have the same solution in θ over the interval $[0,1[$ if (and only if) $\frac{1}{9}\frac{d_1-2c+a}{a-d_1}$ is a solution in X to the equation $\frac{4(a-d_1)}{c-d_1}X = \frac{8}{X-9}$. It is easy to check that this is satisfied.

A3. Proof of lemma 6

It is easy to see that $d_2^{int}(d_1,f,\theta)$ depends on the parameters (f,θ) only through the parameter $f\frac{(3-\theta)^2}{1-\theta}$.With a slight modification of notations, we can write $d_2^{int}(d_1,f,\theta) = d_2^{int}(d_1,\frac{f}{A(\theta)})$. It is also clear that the imitation level d_2^{int} is increasing in $\frac{f}{A(\theta)}$. Hence, lemma 6 simply derives from $d_2^S(d_1) = d_2^{int}(d_1, \frac{f}{A(\theta)} = 9)$.

A4. Proof of lemma 8.

Consider first the case $\theta < \tilde{\theta}\,(d_1)$. Let us show that equation $\Pi_1^P(d_1, d_2^{int}(d_1, f, \theta), \theta) = \Pi_1^S(d_1, 9c - 4a - 4d_1)$ which expresses that the innovator is indifferent between patenting and keeping secrecy (in this subcase) has a unique solution in f over the interval $[\alpha(d_1, \theta), 9A(\theta)]$. Since the function $f \longrightarrow d_2^{int}(d_1, f, \theta)$ is strictly increasing over the interval $[\alpha(d_1, \theta), 9A(\theta)]$, this is equivalent to state that equation $\Pi_1^P(d_1, d_2, \theta) = \Pi_1^S(d_1, 9c-4a-4d_1)$ has a unique solution in d_2 over the interval $\left[d_2^{int}(d_1, \alpha(d_1, \theta), \theta), d_2^{int}(d_1, 9A(\theta), \theta)\right]$. The latter interval can simply be written as $[d_1, 9c - 4a - 4d_1]$.Note that the function $F_\theta : d_2 \longrightarrow \Pi_1^P(d_1, d_2, \theta) - \Pi_1^S(d_1, 9c - 4a - 4d_1)$ is a convex parabolic function then it is either i/ increasing over $[d_1, 9c - 4a - 4d_1]$ or ii/ U-shaped over $[d_1, 9c - 4a - 4d_1]$. Since $\theta < \tilde{\theta}\,(d_1)$, we have $F_\theta(d_1) < 0$ (see subcase 2.1). We also know from lemma 4 that $F(9c - 4a - 4d_1) \geq 0$. It follows that in both cases i/ and ii/ equation $F_\theta(d_2) = 0$ has a unique solution over $[d_1, 9c - 4a - 4d_1]$. Hence, the equation $\Pi_1^P(d_1, d_2^{int}(d_1, f, \theta), \theta) = \Pi_1^S(d_1, 9c - 4a - 4d_1)$ has a unique solution in f that we denote by $\lambda(d_1, \theta)$. Note that $d_2^{int}(d_1, 1, 0) = d_2^S\,(d_1) = 9c - 4a - 4d_1$ which leads to $\lambda(d_1, 1) = 0$. Note also that :

$$\Pi_1^P(d_1, d_2^{int}(d_1, f, \theta), \theta) > \Pi_1^S(d_1, 9c - 4a - 4d_1) \text{ if and only if } f > \lambda\,(d_1, \theta) \tag{1.16}$$

Furthermore, we know that $d_2^{int}(d_1, \alpha(d_1, \tilde{\theta}\,(d_1)), \tilde{\theta}\,(d_1)) = d_1$ and $\Pi_1^P(d_1, d_1, \theta) = \Pi_1^S(d_1, 9c-4a-4d_1)$ (see subcase 2.1) so $\Pi_1^P(d_1, d_2^{int}(d_1, \alpha(d_1, \tilde{\theta}\,(d_1)), \tilde{\theta}\,(d_1)), \tilde{\theta}\,(d_1)) = \Pi_1^S(d_1, 9c-4a-4d_1)$ which leads to $\lambda(d_1, \tilde{\theta}(d_1)) = \alpha(d_1, \tilde{\theta}(d_1))$.

Consider now the case $\theta > \tilde{\theta}\,(d_1)$. Let $\theta_0 > \tilde{\theta}\,(d_1)$ and $f_0 \in\,]\alpha(d_1, \theta), 9A(\theta)[$. The point

(θ_0, f_0) belongs to the curve $f = \frac{f_0}{A(\theta_0)} A(\theta)$. It is easy to see (graphically or analytically) that the curve $f = \frac{f_0}{A(\theta_0)} A(\theta)$ necessarily meets either the curve defined by $\theta = \tilde{\theta}(d_1)$ and $f \leq \alpha(d_1, \theta)$ or the curve defined by $f = \lambda(d_1, \theta)$ and $\theta \leq \tilde{\theta}(d_1)$ at a point (θ_1, f_1) such that $\theta_1 < \theta_0$. Since in any point of the latter two curves the innovator's profit under regime P is equal to its profit under regime S, and $\theta_1 < \theta_0$, we derive from lemma 7 that the innovator's profit under regime P is greater than its profit under regime S when $(\theta, f) = (\theta_0, f_0)$. Hence, the innovator chooses to patent its innovation.

Let us now show that $\lambda(d_1, \theta)$ is strictly decreasing in θ. Consider θ_1 and θ_2 such that $\theta_1 < \theta_2 \leq \tilde{\theta}(d_1)$. The points (θ_1, d_1) and $\left(\theta_2, \frac{\lambda(d_1,\theta_1)}{A(\theta_1)} A(\theta_2)\right)$ belong to the curve defined by $f = \frac{\lambda(d_1,\theta_1)}{A(\theta_1)} A(\theta)$. We derive from lemma 7 that $\Pi_1^P(d_1, d_2^{int}(d_1, \frac{\lambda(d_1,\theta_1)}{A(\theta_1)} A(\theta_2), \theta_2), \theta_2) < \Pi_1^P(d_1, d_2^{int}(d_1, \lambda(d_1, \theta_1), \theta_1), \theta_1) = \Pi_1^S(d_1, 9c - 4a - 4d_1)$ which leads, according to (1.17), to $\frac{\lambda(d_1,\theta_1)}{A(\theta_1)} A(\theta_2) > \lambda(d_1, \theta_2)$. Furthermore we know that $A(\theta)$ is positive and decreasing, so $\frac{A(\theta_2)}{A(\theta_1)} < 1$. This allows us to state that $\lambda(d_1, \theta_1) > \lambda(d_1, \theta_2)$.

Chapter 2

Licensing Uncertain Patents: Per-Unit Royalty vs Fixed Fee[1]

2.1. Introduction

Licensing intellectual property is a key element in the innovation process and its diffusion. A license is a contract whereby the owner of intellectual property authorizes another party to use it, in exchange for payment.[2] The properties and virtues of licensing (Kamien, 1992, Scotchmer, 2004) have mainly been analyzed in a framework in which intellectual property rights guarantee perfect protection and give their owners a right to exclude as strong as physical property rights do. This framework does not correspond to what we observe in practice. In the real world patents do not give the right to exclude but rather a more limited right to "*try to exclude*" by asserting the patent in court (Ayres and Klemperer, 1999, Shapiro, 2003, Lemley and Shapiro, 2005). The exclusive right of a patentholder

[1]This chapter is based on a joint work with David Encaoua.

[2]According to some surveys (Taylor and Silberstone,1973, Rostoker,1984, and Anand and Khanna, 2000), the per-unit royalty rate and the fixed fee mechanism are the most frequent licensing schemes.

can be enforced only if the court upholds the patent validity. For this reason, patents are considered as *probabilistic rights* rather than *ironclad rights.* This paper is devoted to the analysis of licensing patents that are *uncertain*, i.e. patents that have a positive probability to be invalidated by a court if they are challenged.[3]

Many reasons explain the inherent uncertainty attached to a patent. First, the standard patentability requirements, namely the subject matter, utility, novelty and non-obviousness (or inventive step in Europe) are difficult to assess by patent office examiners. Legal uncertainty over the patentability standards is especially pervasive in the new patenting subject matters for which the prior art is rather scarce, like software or business methods. Moreover, the claims granted by the patent office are supposed to delineate the patent scope, but their *ex post* validation depends on the judicial doctrine adopted by the court, and it may be difficult for a patentholder and a potential infringer to know exactly what the patent protects. Second, the resources devoted to the patentability standards review by the patent office are in general insufficient to allow an adequate review of each patent application.[4] Many innovations are granted patent protection even though they do not meet patentability standards. This results in many "weak patents", i.e. patents that have a high probability to be invalidated by a court if they are challenged. Finally, it has been argued that incentives inside the patent offices make it easier and more desirable for examiners to grant patents rather than reject them (Farrell and Merges, 2004, IDEI report, 2006).

The patent quality problem raises many concerns particularly in the US.[5] We may ask,

[3]This uncertainty does not necessarily imply asymmetric information or different beliefs about patent validity among involved parties. Uncertainty may occur even if the parties share the same beliefs on the patent validity. For a different view, see Bebchuk (1984), Reinganum and Wilde (1986), Meurer (1989), Hylton (2002).

[4]The average time spent by an examiner on each patent is about 15-20 hours in the USPTO (Jaffe & Lerner, 2004) and around 30 hours in the EPO. The gap between the massive growth of patent applications and the insufficient resources at the patent office creates a "*vicious circle*" (Caillaud and Duchêne, 2005). Incentives to file "bad applications" increase the patent office overload, and a larger overload leads to further deterioration of the examination process.

[5]Europe is also concerned by the patent quality problem even though the post-grant opposition at the

first: are bad quality patents harmful or not? Lemley (2001) claims that it is reasonably efficient to maintain a low standard of patent examination, in accordance with the "*rational ignorance principle*". Specifically, he argues that the cost of a thorough examination for each application would be prohibitive while inducing only a small benefit. Firstly, the majority of patents turn out to have insignificant market value implying that the social cost of granting them is small even if they are invalid. Secondly, if a weak but profitable patent is granted, some market players will probably bring the case before a court to settle the validity issue, if the patent is licensed at too high a price.

These arguments have attracted much criticism. First, there are many reasons to think that individual incentives to challenge a weak patent are rather low. A patentee generally cares more about winning than a potential infringer does, since by winning against a single challenger, a patentee establishes the validity of the patent against many other potential infringers. By contrast, when infringers are competitors, a successful challenge obtained by one of them benefits all (Farrell and Merges, 2004, Lemley and Shapiro, 2005).[6] Consequently, according to the free-riding argument, the individual incentives to challenge a patent validity are weak. Moreover, according to the so-called pass-through argument, licensees are induced to accept a high per-unit royalty rate when they can decide to pass-on the royalty to their customers.[7] Finally, an unsuccessful attacker may be in jeopardy or even evicted from the market once deprived from the new technology, or required to pay a higher price than the licensees who have accepted the licensing contract.

EPO alleviates it (see Graham *et al.* 2003). The European situation in terms of patent quality is analyzed in Guellec and von Pottelsberghe de la Potterie (2007) and the IDEI report (2006).

[6]Following the *Blonder-Tongue* decision (1971), it became clear that "the attacker is not able to exclude others from appropriating the benefit of its successful patent attack", Blonder-Tongue Labs., Inc. v. Univ. of Illinois Found, 402, U.S. 313, 350 (1971).

[7]When multiple infringers compete in a product market, royalties are often passed-through, at least in part, to consumers. The pass-through will be stronger the more competitive the product market, the more symmetric the royalties, the more elastic the industry supply curve, and the less elastic the industry demand curve" (Farrell and Merges, 2004).

All these arguments suggest that individual incentives to challenge a patent may be rather low. The probabilistic nature of patent protection and the low individual incentives to challenge a patent may thus strengthen the market power of the licensor. The owner of a probabilistic right and a potential user will come to a licensing agreement as a private settlement to avoid the uncertainty of a court resolution. An agreement benefits the holder of a weak patent while litigation and possible invalidation by a court would deprive the licensor from any licensing revenue. However the licensing contract will be accepted by the licensee only if its expected profit is at least as large as when the patent validity is challenged. Therefore licensing an uncertain patent under the shadow of patent litigation raises an interesting trade-off. We show in this paper that different factors explain the issue of this trade-off: i) the nature of the licensing scheme (per-unit royalty vs. up-front fee); ii) the patent strength measured by the probability that it will be upheld; iii) the importance of the innovation; iv) the type of commitment when dealing with an unsuccessful challenger; v) the possibility to engage in collective negotiations of the licensing contract; vi) some market structure variables such as the size of the industry and the intensity of market competition.[8]

The literature on licensing and the properties of the different licensing mechanisms has extensively examined the case of perfect patent protection. Based largely on previous works by Arrow (1962), Katz and Shapiro (1985, 1986), Kamien and Tauman (1984, 1986), Kamien *et al.* (1992), the survey by Kamien (1992) summarizes the major results, especially by comparing the patentholder's profits under different licensing schemes. The patentee's profits are highest when licensing is made through an auction, in which the patentee announces the number of licenses on offer and the latter accrue to the highest

[8]Since a non-licensee suffers a negative externality when a competitor becomes a licensee, more intense competition in the product market increases the licensor's market power.

bidders. The per-unit royalty scheme and the up-front fee mechanism have been set against each other. While the earlier literature claimed that a per unit royalty always generates lower profits than a fixed fee, regardless of the industry size and the magnitude of the innovation (Kamien and Tauman,1984 and 1986), a more recent work has shown that when the number of firms in the industry is sufficiently high, the innovator's payoff is higher with royalty licensing than with a fixed fee or an auction (Sen, 2005). Moreover, some licensing methods induce full diffusion, while others lead to only partial diffusion of the innovation: the number of licensees depends on the licensing method and the magnitude of the cost reduction. In a more recent contribution, Sen and Tauman (2007) generalize these findings by allowing the optimal combination of an auction and a per-unit royalty in situations where the innovator may be either an outsider or an insider in the industry.[9]

Let us now consider that a patent is a probabilistic right. Rough intuition suggests that licensing an uncertain patent in the shadow of patent litigation leads to a license price which is proportional to the patent strength. This intuition is not always correct for the following reason: when imperfect competition occurs in the industry, the free riding argument mentioned above lowers the individual incentives to challenge the patent's validity and this benefits the patentholder. Farrell and Shapiro (2007) establishes two important properties for a minor cost reducing innovation: (i) For weak patents, the royalty rate is as high as if the patent were certain: it is equal to the magnitude of the cost reduction allowed by the innovation; (ii) Whatever the patent strength, the royalty rate obtained in

[9]Another burgeoning literature explores the consequences of informational asymmetries on licensing. Aoki and Hu (1996) examines how the choice between strategic licensing and litigating is affected by the levels of the litigation costs and their allocation between the plaintiff and the defendant. Brocas (2006) identifies two informational asymmetries: the moral hazard due to the inobservability of the innovator's R&D effort, and the adverse selection due to the private value of holding a license. Macho-Stadler *et al.* (1996) introduces know-how transfer and shows that the patentholder prefers contracts based on per-unit royalties rather than fixed fee payments. Other contributions, emphasizing either risk aversion (Bousquet *et al.*, 1998), strategic delegation (Saracho, 2002), strategic complementarity (Muto, 1993, Poddar and Sinha, 2004), or the size of the oligopoly market (Sen, 2006) reach the same conclusion stating the superiority of the royalty licensing scheme.

the shadow of patent litigation exceeds the expected value of the royalty resulting from the patent challenge. These strong properties have been obtained by considering a two-part licensing contract mechanism combining a per-unit royalty and a fixed fee, allowing for instance a high royalty rate to be compensated by a negative transfer (i.e. an up-front fee paid by the licensor to the licensee).[10] Two restrictive assumptions have been used in Farrell and Shapiro to obtain these results: first, they restrict their analysis to small process innovations, i.e. innovations leading to a small cost reduction; second, they assume that the best patentholder's licensing strategy is to sell a license to all firms in the industry, rather than to restrict the license supply to some firms, leaving it to others to possibly initiate a litigation process.

In this paper we assess the robustness of these results by separately investigating two of the most common licensing mechanisms, namely the per-unit royalty rate and the up-front fee. We analyze the properties of these mechanisms which let the licensor choose the number of licensees whatever the innovation size. For both types of licensing schemes, we develop a three-stage game in which the patentholder, acting as a Stackelberg leader, determines either a royalty rate or a fixed fee at the first stage. At the second stage, each firm independently decides whether to accept the licensing contract. If it does not, it challenges the patent validity. If the patent is found valid, the unsuccessful challenger is bound to use the old technology. If the patent is found invalid, all the firms in the oligopolistic industry have free access to the technology. In the last stage, licensees and non-licensees compete in the product market. Different variants of this basic model are examined in this paper, by introducing the possibility of a collective challenge or by allowing renegotiation between the patentholder and an unsuccessful challenger.

[10] Farrell and Shapiro also investigate a two-part tariff in which the fixed fee is constrained to be non-negative. However, in this case, their main result holds only under additional restrictions.

Our paper departs from Farrell and Shapiro (2007) in several ways. First, unlike Farrell and Shapiro who focus on a single licensing scheme combining a per-unit royalty and a fixed fee, we separately investigate these two schemes; second, while they only consider the case where the cost reduction is small, we investigate the consequences of any cost reduction; third, we relax the crucial assumption of their paper stating that the patentholder licenses every firm in the industry, by endogeneizing the number of licensees. We show below that this endogeneization has important consequences, particularly when comparing the properties of the per-unit royalty rate and the up-front fee licensing schemes. We also challenge the assumption that an unsuccessful challenger is offered a license at a price that captures its entire surplus.

We contribute to the literature on licensing uncertain patents on five points. First, we show that while it is generally possible for the patentholder to reap some "extra profit" by selling an uncertain patent under the per-unit royalty regime, the opportunity to do so under the up-front fee regime disappears. This is due to the fact that the patentee's profit under a fixed fee regime is always equal to the expected profit in case of litigation.[11] Second, we show in the case of a linear demand under Cournot competition that the patentee's profits may be higher with a per-unit royalty than with a fixed fee. This result - which confirms Sen (2005) - rests on a completely different argument based on patent uncertainty. Third, for the per-unit royalty regime, we obtain sufficient conditions under which the royalty rate resulting from a collective challenge is lower than the expected royalty from an individual challenge. Fourth, we show that there exist situations in which

[11] Under a fixed-fee licensing regime, the patentholder can avoid patent litigation only by offering a fee at most equal to a proportion of the fee that he would choose if the patent were perfect, this proportion being exactly equal to the patent strength. Under a per-unit royalty regime, the patentholder prefers to avoid patent litigation if and only if the royalty is higher than some threshold. Sufficient conditions under which this threshold is larger than the expected value of the royalty rate in case of litigation are established. Other sufficient conditions ensure that the threshold may be lower than this benchmark. These properties are a novel contribution on licensing uncertain patents.

the per-unit royalty for a weak patent is below the expected royalty in case of litigation. The latter result is obtained under general assumptions on the profit functions and is confirmed when post-trial renegotiation is introduced. Finally, we show that the results obtained with perfect patents also hold when patents are uncertain but strong: in this case, litigation never occurs.

The paper is organized as follows. Section 2.2 examines the per-unit royalty scheme. It starts with the derivation of the maximal value of the per-unit royalty that deters any litigation. This value is compared to two benchmarks: i/ the expected value of the royalty in case of litigation; ii/ the royalty that would prevail under collective challenges of the patent validity. The patentholder's optimal royalty rate and its licensing revenues are then determined. The conditions under which litigation is avoided at the subgame perfect equilibrium are established. Section 2.3 analyzes the fixed fee licensing scheme. It derives the demand for licenses and the licensing revenues as a function of the up-front fee. These revenues are then compared to the expected revenues in case of litigation. In Section 2.4, the two licensing mechanisms are compared from the licensor's perspective. Section 2.5 concludes by summarizing the results, putting them in an economic policy perspective, and suggesting new research directions.

2.2. Royalty licensing schemes

We consider an industry consisting of n identical risk-neutral firms producing at a marginal cost c (fixed production costs are assumed to be zero). A firm P outside the industry holds a patent covering a technology that would allow each firm to reduce its marginal cost from c to $c - \epsilon$.

In this section, we examine licensing schemes involving a pure royalty rate. More specifically, we seek to determine the subgame perfect equilibria of the following three-stage game:

First stage: The patentholder P proposes a licensing contract whereby a licensee can use the patented technology against the payment of a per-unit royalty rate r.

Second stage: The n firms in the industry simultaneously and independently decide whether to purchase a license at the royalty rate r. If a firm does not accept the license offer, it can challenge the patent validity before a court.[12] The outcome of such a trial is uncertain: with probability θ the patent is upheld by the court and with probability $1-\theta$ it is invalidated. The parameter θ measures the patent strength. If the patent is upheld, then a firm that does not purchase the license uses the old technology,[13] thus producing at marginal cost c whereas those who accepted the license offer use the new technology and pay the royalty rate r to the patentholder, having thus an effective marginal cost equal to $c-\epsilon+r$. If the patent is invalidated, all the firms, including those who accepted the offer can use for free the new technology and their common marginal cost is $c-\epsilon$. [14]

Third stage: The n firms produce under the cost structure inherited from stage 2. The kind of competition that occurs is not specified. It is only assumed that there exists a

[12] In the US, a firm can seek a declaratory judgement against the validity of a patent if it has a "reasonable apprehension" of being sued for infringement by the patentholder. A firm that is planning to use a patented technology, or is currently using it, without a license can reasonably fear to be sued for infringement.

[13] This assumption may seem quite strong but recall that IP laws do not compel patentholders to license others, particularly those who challenge the validity of a patent or sue the patentholder for infringement of their own patents. To illustrate, when Intergraph (a company producing graphic work stations) sued Intel (micro-processors) for infringement of its Central Processing Unit patent, Intel countered by removing Intergraph from its list of customers and threatening to discontinue the sale of Intel microprocessors to Intergraph (See Encaoua and Hollander, 2002). We relax later this assumption by introducing renegotiation between the unsuccessful challenger and the patentholder.

[14] Note that in our model the plaintiff is the potential licensee and the defendant is the patentholder while in Farrell and Shapiro (2007) the roles are inversed. Both situations occur in the real world. Nonetheless, note that, in a setting without litigation costs, as in our model and Farrell and Shapiro's, who the plaintiff/defendant is does not matter. What matters in both models is that a trial in which patent validity will be examined by the court, will occur whenever at least one firm does not accept the licensing contract. In Farrell and Shapiro, a patentholder always finds it optimal to sue a firm that uses its technology without a license and the alleged infringer challenges the patent validity as a defense strategy. In our model a firm that refuses the licensing contract always finds it optimal to challenge the patent validity.

unique Nash equilibrium in the competition game between the members of the oligopoly for any cost structure of the firms.

We sum-up the outcome of the third stage by denoting $\pi(x, y)$ the equilibrium profit function of a firm producing with marginal cost x while its $(n-1)$ competitors produce with marginal cost y.[15] The case where $\pi(x, y) = 0$ is not excluded.

We assume the following general properties that are satisfied by a large class of profit functions (See Amir and Wooders, 2000 and Boone, 2001).

A1. A firm's equilibrium profit $\pi(x, y)$ is continuous in both it arguments over $[0, +\infty[\times [0, +\infty[$ and twice differentiable in both its arguments over the subset of $[0, +\infty[\times [0, +\infty[$ in which $\pi(x, y) > 0$. Moreover $\pi(c, c) > 0$.

A2. A firm's equilibrium profit is decreasing in its own cost : If $\pi(x, y) > 0$ then $\pi_1(x, y) < 0$, and if $\pi(x, y) = 0$ then $\pi(x', y) = 0$ for any $x' > x$.

A3. A firm's equilibrium profit is increasing in its competitors' costs : If $\pi(x, y) > 0$ then $\pi_2(x, y) > 0$, and if $\pi(x, y) = 0$ then $\pi(x, y') = 0$ for any $y' < y$.

A4. In a symmetric oligopoly, an identical drop in all firms' costs raises each firm's equilibrium profit: If $\pi(x, x) > 0$ then $\pi_1(x, x) + \pi_2(x, x) < 0$, and if $\pi(x, x) = 0$ then $\pi(y, y) = 0$ for any $y > x$.

Given A2 and A3, A4 means that own cost effects dominate rival's cost effects. This assumption, while being fullfilled in a wide range of competitive settings including Cournot oligopoly and differentiated Bertrand oligopoly with linear demand, may not be satisfied

[15] We restrict the notation to the situations where at least $(n-1)$ firms produce with the same marginal cost because this will be the case in all the equilibria under royalty licensing as we will see in the subsequent analysis.

under Cournot competition when the demand is "very convex" (see Kimmel 1992, Février and Linnemer 2004).

We solve for the subgame perfect Nash equilibria of the game using backward induction.

2.2.1. Accepting or not the patentholder's offer: second stage

Let us determine the set of royalty rates r such that all firms accepting the licensing contract is a Nash equilibrium of the second stage. This occurs if and only if no firm has an incentive to deviate by refusing to buy a license at this rate and challenging the patent validity.[16] The expected profit from such a unilateral deviation is $\theta\pi(c, c-\epsilon+r)+(1-\theta)\pi(c-\epsilon, c-\epsilon)$ because:

- If the challenger does not succeed in invalidating the patent, which happens with probability θ, it produces at the cost c while its competitors produce at the effective cost $c-\epsilon+r$,
- If the patent is invalidated, which happens with probabiliy $1-\theta$, all firms produce at the cost $c-\epsilon$.

Thus, all firms accepting a per-unit royalty r is a Nash equilibrium if and only if:

$$\pi(c-\epsilon+r, c-\epsilon+r) \geq \theta\pi(c, c-\epsilon+r)+(1-\theta)\pi(c-\epsilon, c-\epsilon) \tag{2.1}$$

Note that the royalty rate r affects both sides of this inequality. Due to assumption A4, the left-hand side, that is, a firm's profit when all firms accept the license is decreasing in r. Due to assumption A3, the right-hand side, that is, the expected profit of a challenger when all other firms accept the license offer, is (weakly) increasing in r. Thus, a lower royalty rate makes the license option more attractive to a potential licensee for two reasons: it increases

[16]Since there are no litigation costs, a firm that refuses the licensing contract always finds it optimal to challenge the patent validity. For an analysis of the effects of litigation costs on licensing under the shadow of litigation, see Aoki & Hu (1999).

the payoff from the license option and it decreases the payoff from the outside option namely the challenge option. Note that the latter indirect effect arises only if $\pi(c, c-\epsilon+r) > 0$. However it may happen that the extent of the cost asymmetry between the licensees and an unsucessful challenger result in zero profit for the latter, that is, $\pi(c, c-\epsilon+r) = 0$. We distinguish between two cases according to whether such royalty rate values exist.

Case 1: $\pi(c, c-\epsilon) = 0$.

This case may occur for a sufficiently large innovation (high value of ϵ) or a sufficiently intense competition (e.g. large number n of firms). Using assumptions A1 and A3 and the intermediate value theorem, one easily shows that there exists a threshold $\hat{r} \in [0, \epsilon]$ such that $\pi(c, c-\epsilon+r) = 0$ if $r \leq \hat{r}$ and $\pi(c, c-\epsilon+r) > 0$ if $r > \hat{r}$. In other words, an unsuccessful challenger will not be viable if the royalty rate is below some threshold $\hat{r}$, and will be viable if the royalty rate is above the threshold $\hat{r}$.

Consider first a contract involving a royalty rate $r \leq \hat{r}$. In this case, condition (2.1) can be rewritten as:

$$\pi(c-\epsilon+r, c-\epsilon+r) \geq (1-\theta)\pi(c-\epsilon, c-\epsilon) \tag{2.2}$$

Let $\hat{\theta} \in [0,1]$ be the unique solution in θ to the equation $\pi(c-\epsilon+\hat{r}, c-\epsilon+\hat{r}) = (1-\theta)\pi(c-\epsilon, c-\epsilon)$. The following lemma introduces a threshold that is useful for characterizing the licensing contracts accepted by all firms:

Lemma 1 *Assume that $\pi(c, c-\epsilon) = 0$. The equation $\pi(c-\epsilon+r, c-\epsilon+r) = (1-\theta)\pi(c-\epsilon, c-\epsilon)$ has a unique solution in r over $[0, \hat{r}]$ for any $\theta \in \left[0, \hat{\theta}\right]$. This solution, denoted $r_2(\theta)$, satifies the following properties: i/ $r_2(\theta)$ is differentiable and increasing in θ over*

$\left[0, \hat{\theta}\right]$ [17], *ii/* $r_2(0) = 0$ *and* $r_2(\hat{\theta}) = \hat{r}$.

Proof. See Appendix. ■

Consider now a contract involving a royalty rate $r > \hat{r}$. It will be accepted by all firms if and only if inequality (2.1) is satisfied.

Lemma 2 *Assume that* $\pi(c, c-\epsilon) = 0$. *The equation* $\pi(c-\epsilon+r, c-\epsilon+r) = \theta\pi(c, c-\epsilon+r) + (1-\theta)\pi(c-\epsilon, c-\epsilon)$ *has a unique solution in* r *over* $[\hat{r}, \epsilon]$ *for any* $\theta \in \left[\hat{\theta}, 1\right]$. *This solution, denoted* $r_1(\theta)$, *satisfies the following properties: i/* $r_1(\theta)$ *is differentiable and increasing in* θ *over* $\left[\hat{\theta}, 1\right]$, *ii/* $r_1(\hat{\theta}) = \hat{r}$ *and* $r_1(1) = \epsilon$.

Proof. See Appendix ■

We can now characterize the set of royalty rates that are accepted by all firms when $\pi(c, c-\epsilon) = 0$.

Proposition 3 *If* $\pi(c, c-\epsilon) = 0$ *then all firms accepting the royalty rate* r *is a Nash equilibrium if and only if* $r \leq r(\theta)$ *where:*

$$r(\theta) = \begin{cases} r_2(\theta) & \text{if} \quad \theta \in \left[0, \hat{\theta}\right] \\ r_1(\theta) & \text{if} \quad \theta \in \left]\hat{\theta}, 1\right] \end{cases}$$

Proof. See Appendix. ■

When the innovation is large or/and the intensity of competition is high, proposition 3 shows that the firms' incentives to accept a given royalty rate crucially depend on whether the patent is relatively weak (i.e. $\theta \leq \hat{\theta}$) or relatively strong (i.e. $\theta > \hat{\theta}$). When the patent is strong, the positive effect of a higher royalty rate on the outside option profit

[17] Throughout this paper, a function f will be said to be differentiable over a closed interval $[a, b]$ if it is differentiable at any point of the open interval $]a, b[$, right-differentiable at a and left-differentiable at b.

(i.e. a challenger's profit) plays a role in constraining the royalty rates acceptable by all firms. Indeed, $\pi\left(c, c-\epsilon+r_1\left(\theta\right)\right)>0$ because $r_1\left(\theta\right)>\hat{r}$ for all $\theta>\hat{\theta}$. However, when the patent is weak, this indirect effect does not play a role since $\pi\left(c, c-\epsilon+r_2\left(\theta\right)\right)=0$, due to $r_2\left(\theta\right)\leq\hat{r}$ for all $\theta<\hat{\theta}$. In this sense, a firm has an additional incentive not to accept a licensing contract when the patent is strong enough.[18]

Remark 1 : From lemma (1) and (2), it is clear that the maximal royalty rate $r\left(\theta\right)$ acceptable by all firms is increasing and continuous over $[0,1]$. Moreover, it is differentiable over $\left[0,\hat{\theta}\right]$ and $\left[\hat{\theta},1\right]$ but its left-sided derivative is different from its right-sided derivative at point $\theta=\hat{\theta}$. One can show that the former is greater than the latter which is in line with our previous observation that an extra force (stemming from the indirect effect we pointed out) constrains the royalty rates acceptable by all firms when $\theta>\hat{\theta}$.

The following property of $r_2(\theta)$ will be useful for the comparison of the maximal royalty rate $r\left(\theta\right)$ accepted by all firms to some benchmarks we define later.

Lemma 4 *$r_2(\theta)$ is convex over $\left[0,\hat{\theta}\right]$ if (and only if) $\pi\left(x,x\right)$ is convex in x over $\left[c-\epsilon, c-\epsilon+\hat{r}\right]$.*

Proof. See Appendix ■

Note that the convexity of a firm's profit $\pi\left(x,x\right)$ in a symmetric industry holds for a wide range of competitive environments, including Cournot oligopoly with linear or iso-elastic demand and differentiated Bertrand oligopoly with linear demand.[19]

[18] One can get to the same interpretation using a more formal argument: defining the threshold $r_2\left(\theta\right)$ not only for $\theta\in\left[0,\hat{\theta}\right]$ but for all $\theta\in[0,1[$ as the unique solution to the equality derived from to inequality (2), we can show that $r_1\left(\theta\right)<r_2\left(\theta\right)$ for all $\theta\in\left]\hat{\theta},1\right[$.

[19] It can be shown that, if $\pi\left(x,x\right)$ is concave in x over $[c-\epsilon,c]$ then $r_2(\theta)$ is concave over $\left[0,\hat{\theta}\right]$. However, it is difficult, if not impossible, to find a simple demand function leading to the concavity of the equilibrium profit function $\pi\left(x,x\right)$ under neither Cournot nor differentiated Bertrand competition.

Case 2: $\pi(c, c-\epsilon) > 0$

In this case, whatever the royalty rate $r \geq 0$ proposed by the patentholder, the profit of an unsuccessful challenger remains positive even when all other firms purchase a license: $\pi(c, c-\epsilon+r) \geq \pi(c, c-\epsilon) > 0$. Therefore, in this case, we use the same notation $r_1(\theta)$ for the unique solution in r to the equation $\pi(c-\epsilon+r, c-\epsilon+r) = \theta\pi(c, c-\epsilon+r)+(1-\theta)\pi(c-\epsilon, c-\epsilon)$ for all $\theta \in [0,1]$.[20] The existence, uniqueness and properties of $r_1(\theta)$ can be established as under case 1. These are stated in the following lemma:

Lemma 5 *Assume that $\pi(c, c-\epsilon) > 0$. The equation $\pi(c-\epsilon+r, c-\epsilon+r) = \theta\pi(c, c-\epsilon+r)+(1-\theta)\pi(c-\epsilon, c-\epsilon)$ has a unique solution in r over $[0, \epsilon]$ for any $\theta \in [0,1]$. This solution, denoted $r_1(\theta)$, satisfies the following properties: i/ $r_1(\theta)$ is differentiable and increasing in θ over $[0,1]$, ii/ $r_1(0) = 0$ and $r_1(1) = \epsilon$.*

Proof. See Appendix ■

The next proposition characterizes the set of royalty rates acceptable by all firms.

Proposition 6 *If $\pi(c, c-\epsilon) > 0$ then, for any $\theta \in [0,1]$, all firms accepting the royalty rate r is a Nash equilibrium if and only if $r \leq r(\theta) = r_1(\theta)$.*

Proof. See Appendix ■

Considering again the indirect effet that captures the positive externality of a higher royalty rate on a challenger's expected profit, we can state that this effect is always at work in constraining the royalty rates acceptable by all firms when the innovation is small or/and the competition intensity is low.

[20]The threshold $r_1(\theta)$ that could be denoted $r_1(\theta, \epsilon)$ to explicitly display its dependence upon ϵ, has been previously defined for the values of ϵ such that $\pi(c, c-\epsilon) = 0$, and for patent strength values $\theta \in \left[\hat{\theta}, 1\right]$ (see lemma 2). Here, this threshold is defined for the values of ϵ that satisfy $\pi(c, c-\epsilon) > 0$ and for all patent strength values $\theta \in [0,1]$.

We now illustrate those results in the case of Cournot oligopoly with linear demand.

Example 1: Cournot oligopoly with linear demand

Under Cournot competition with linear demand $Q = a - p$, it is straightforward to show that: $\pi(c, c-\epsilon) = 0 \Longleftrightarrow \epsilon \geq \frac{a-c}{n-1}$.

Therefore, if $\epsilon < \frac{a-c}{n-1}$, a licensing contract with a royalty rate r is accepted by all firms if and only if $r \leq r_1(\theta)$ where $r_1(\theta)$ is the unique positive solution in r to equation (2.1) which, in the case of Cournot competition with linear demand, is equivalent to the following equation: $(a-c+\epsilon-r)^2 = \theta[a-c-(n-1)(\epsilon-r)]^2 + (1-\theta)(a-c+\epsilon)^2$.

Assume now that $\epsilon \geq \frac{a-c}{n-1}$. Let us determine the value $\hat{r}$ such that the inequality $\pi(c, c-\epsilon+r) > 0$ is equivalent to $r > \hat{r}$. A simple calculation leads to $\hat{r} = \epsilon - \frac{a-c}{n-1}$. Therefore, a licensing contract based on a royalty rate $r \leq \hat{r}$ is accepted by all firms if and only if $r \leq r_2(\theta)$ where $r_2(\theta)$ is the unique solution in $r \in [0, \epsilon]$ to the following equation: $[a-c+\epsilon-r]^2 = (1-\theta)[a-c+\epsilon]^2$.

The positive solution to this equation is given by $r_2(\theta) = [1-\sqrt{1-\theta}](a-c+\epsilon)$. This expression can be used to determine the patent strength threshold $\hat{\theta} = 1 - [\frac{n(a-c)}{(n-1)(a-c+\epsilon)}]^2$ such that $r_2(\hat{\theta}) = \hat{r}$.

Thus, if $\epsilon \leq \frac{a-c}{n-1}$, the maximal royalty rate the patentholder can make all firms accept is $r(\theta) = r_1(\theta)$ for all $\theta \in [0,1]$, while if $\epsilon \geq \frac{a-c}{n-1}$, the maximal royalty rate the patentholder can make all firms accept is given by:

$$r(\theta) = \begin{cases} r_2(\theta) = [1-\sqrt{1-\theta}](a-c+\epsilon) & \text{if} \quad 0 \leq \theta \leq \hat{\theta} \\ r_1(\theta) & \text{if} \quad \hat{\theta} \leq \theta \leq 1 \end{cases}$$

2.2.1.1. Royalty rate benchmarks

Now that we have characterized the maximal royalty $r(\theta)$ acceptable by all firms, it is interesting to compare it to two benchmarks: i/ the expected value of the maximal royalty rate in case of litigation, which we denote by $r^e(\theta)$; ii/ the royalty rate deterring a collective challenge, which we denote by $r^c(\theta)$.

First benchmark: the expected value of the maximal royalty rate in case of litigation

This benchmark can be easily computed: with probability θ the patent is upheld by the court, hence becoming an ironclad right that can be licensed at a maximal per-unit royalty $r(1) = \epsilon$, and with probability $1-\theta$ the patent is invalidated and the firms can use it for free, leaving the patentholder with a royalty $r(0) = 0$. Therefore, the expected value of the maximal royalty rate in case of litigation is equal to $r^e(\theta) = \theta r(1) + (1-\theta)r(0) = \theta\epsilon$. The expected value of the maximal royalty in case of litigation is thus proportional to the patent strength θ. This benchmark is interpreted in Farrell and Shapiro (2007) as the *ex ante* value of the per-unit royalty rate that an applicant of a process innovation reducing the cost by ϵ can expect when the patent has a probability θ to be granted by the patent office.

Second benchmark: the royalty rate deterring a collective challenge

Suppose that at stage 2 the firms cooperatively agree on whether to buy the license or refuse it and challenge all together the patent validity.[21] In this case, the firms will

[21] Firms are allowed to challenge collectively the validity of a patent, at least in the US. An example is the PanIP Group Defense Fund which is a coalition of fifteen e-retailers that has been created to invalidate a patent covering some key aspects of electronic commerce, hold by Pangea Intellectual Properties (US patent number 5.576.951).

cooperatively accept a licensing contract involving a royalty rate r if and only if:

$$\pi(c-\epsilon+r, c-\epsilon+r) \geq \theta\pi(c,c) + (1-\theta)\pi(c-\epsilon, c-\epsilon)$$

The function w defined by $w(r) = \pi(c-\epsilon+r, c-\epsilon+r) - \theta\pi(c,c) - (1-\theta)\pi(c-\epsilon, c-\epsilon)$ is continuous, strictly decreasing (by A4) and satisfies the conditions $w(0) \geq 0$ and $w(\epsilon) \leq 0$. Hence there exists a unique solution $r^c(\theta) \in [0,\epsilon]$ to the equation $w(r) = 0$, and the inequality $w(r) \geq 0$ is equivalent to $r \leq r^c(\theta)$. This means that all firms cooperatively accept to buy a license at a royalty rate r if and only if $r \leq r^c(\theta)$. Some properties of the benchmark $r^c(\theta)$ are presented in the next proposition.

Proposition 7 *The function $r^c(\theta)$ satisfies the following properties:*

i/ $r^c(\theta)$ is increasing over $[0,1]$ and $r^c(0) = 0$, $r^c(1) = \epsilon$,

ii/ $r^c(\theta)$ is convex over $[0,1]$ if (and only if) the function $x \to \pi(x,x)$ is convex over $[c-\epsilon, c]$ and in this case $r^c(\theta) \leq r^e(\theta) = \theta\epsilon$.[22]

Proof. See Appendix. ■

Thus the royalty rate deterring a collective challenge $r^c(\theta)$ is lower than the expected royalty rate in case of litigation $r^e(\theta)$ if the (reasonable) assumption that $\pi(x,x)$ is convex holds.

2.2.1.2. Comparison of $r(\theta)$ to $r^e(\theta) = \theta\epsilon$

Analyzing the shape of the function $\theta \to r(\theta)$ allows us to compare the per-unit royalty rate $r(\theta)$ that deters individual challenge to the benchmark $r^e(\theta) = \theta\epsilon$ which represents

[22] It can be shown that $r^c(\theta)$ is concave over $[0,1]$ if and only if the function $x \to \pi(x,x)$ is concave over $[c-\epsilon, c]$. In this case $r^c(\theta) \geq r^e(\theta) = \theta\epsilon$. However, as previously mentioned, the concavity of $\pi(x,x)$ is unlikely in usual settings.

the expected maximal royalty rate in case of individual litigation.

Recall first that when the innovation ϵ is such that $\pi(c, c-\epsilon) = 0$, we have $r(\theta) = r_2(\theta)$ over the interval $\left[0, \hat{\theta}\right]$.

We would like to compare $r_2(\theta)$ to the benchmark $r^e(\theta) = \theta\epsilon$ for any $\theta \in \left[0, \hat{\theta}\right]$. If we tackle the case of sufficiently weak patents we can derive a comparison of $r_2(\theta)$ to $r^e(\theta)$ for θ small enough from the comparison of $r_2'(0)$ to ϵ. Indeed, if $r_2'(0) \geq \epsilon$ (resp. $r_2'(0) < \epsilon$) then for θ sufficiently small, but different from 0, we will have $r_2(\theta) > \theta\epsilon$ (resp. $r_2(\theta) < \theta\epsilon$).[23]

Since $r_2(0) = 0$, we have:

$$r_2'(0) = \frac{-\pi(c-\epsilon, c-\epsilon)}{(\pi_1 + \pi_2)(c-\epsilon, c-\epsilon)}$$

Therefore,

$$r_2'(0) \geq \epsilon \Longleftrightarrow \frac{-\pi(c-\epsilon, c-\epsilon)}{\epsilon(\pi_1 + \pi_2)(c-\epsilon, c-\epsilon)} \geq 1$$

Denoting $\lambda(\epsilon) = \pi(c-\epsilon, c-\epsilon)$, we obtain:

$$r_2'(0) \geq \epsilon \Longleftrightarrow \frac{\lambda(\epsilon)}{\epsilon\lambda'(\epsilon)} \geq 1 \Longleftrightarrow \eta(\epsilon) \leq 1 \tag{2.3}$$

where $\eta(\epsilon) = \frac{\epsilon\lambda'(\epsilon)}{\lambda(\epsilon)} = \frac{\epsilon n\lambda'(\epsilon)}{n\lambda(\epsilon)}$ is the elasticity of the industry profits with respect to a cost reduction ϵ.

Hence we can state that for sufficiently weak patents, the comparison of $r_2(\theta)$ and $r^e(\theta) = \theta\epsilon$ can be derived from the value of the elasticty $\eta(\epsilon)$ relative to 1.

Moreover, if the additional assumption that $\pi(x, x)$ is convex in x holds then we know from lemma (4) that $r_2(\theta)$ is convex. This makes it possible to derive the global comparison of $r_2(\theta)$ to $r^e(\theta) = \theta\epsilon$ over the whole interval $\left[0, \hat{\theta}\right]$, as the next proposition shows.

[23] Since $r_2(\theta)$ is strictly convex, even the equality $r_2'(0) = \epsilon$ yields $r_2(\theta) > \theta\epsilon$ for θ small enough (but different from 0).

Proposition 8 *Consider an innovation such that $\pi(c, c-\epsilon) = 0$. Assume that $\pi(x,x)$ is convex in x over $[c-\epsilon, c]$. The comparison of $r_2(\theta)$ with $r^e(\theta) = \theta\epsilon$ depends on $\eta(\epsilon)$ in the following way:*

1- if $\eta(\epsilon) \leq 1$ then $r_2(\theta) > \theta\epsilon$ for any $\theta \in \left]0, \hat{\theta}\right]$.

2- if $\eta(\epsilon) > 1$ then two subcases arise:

2-a- if $\hat{r} < \hat{\theta}\epsilon$ then $r_2(\theta) < \theta\epsilon$ for any $\theta \in \left]0, \hat{\theta}\right]$.

2-b- If $\hat{r} \geq \hat{\theta}\epsilon$ then there exists $\breve{\theta} \in \left]0, \hat{\theta}\right]$ such that $r_2(\theta) < \theta\epsilon$ for any $\theta \in \left]0, \breve{\theta}\right[$ and $r_2(\theta) \geq \theta\epsilon$ for any $\left[\breve{\theta}, \hat{\theta}\right]$.

Proof. See Appendix ■

The elasticity of the industry profits with respect to a cost reduction plays a crucial role in the comparison of the maximal royalty rate acceptable $r_2(\theta)$ by all firms and $r^e(\theta) = \theta\epsilon$. The intuition behind the result stated in Proposition (8) is that a low elasticity entails a low (negative) effect of an increase in the royalty rate on the firms' profit when they all purchase a license. Under such conditions, the patentholder may be able to impose a high royalty rate. In particular, the level of the royalty rate may be greater than the benchmark level $r^e(\theta)$ as case 1 shows. However, if the elasticity of the industry profits is high, the patentholder may not be able to charge a high royalty without triggering a challenge : such royalty would result in a relatively weak profit for the licensees hence making the challenge option more attractive.

The following examples illustrate the two cases presented in proposition (8).

Example 1 (continued): Cournot oligopoly with linear demand

Under this specification, $\eta(\epsilon) = \frac{2\epsilon}{a-c+\epsilon}$ which leads to $\eta(\epsilon) \leq 1 \Leftrightarrow \epsilon \leq a-c$. Moreover, the condition $\pi(c, c-\epsilon) = 0$ is equivalent to $\epsilon \geq \frac{a-c}{n-1}$. Therefore:

- If $\epsilon \in \left[\frac{a-c}{n-1}, a-c\right]$ then $\eta(\epsilon) \leq 1$ which yields $r_2(\theta) > \theta\epsilon$ for any $\theta \in \left]0, \hat{\theta}\right]$.

- If $\epsilon \geq a-c$ then $\eta(\epsilon) > 1$. This allows to state that we are under case 2 of proposition 8. Moreover if $n = 2$ it holds that $\hat{r} < \hat{\theta}\epsilon$ for any $\epsilon > a-c$, which implies that $r_2(\theta) < \theta\epsilon$ for any $\theta \in \left]0, \hat{\theta}\right]$ (that is, subcase 2-a applies) . However, if $n > 2$ then both subcases 2-a and 2-b are possible according to the value of $\epsilon \geq a-c$. We can show that for the values of ϵ sufficiently close to $a-c$, subcase 2-b applies while for greater values of ϵ, subcase 2-a applies.

Example 2: Differentiated Bertrand duopoly with linear demand

Consider a market with two firms producing differentiated goods. Assume that the inverse demand function for product $i = 1, 2$ is given by $p_i = a - (q_i + \gamma q_j)$ where $j \neq i$ and $\gamma \in [0, 1[$. Tedious calculations show that, under this specification, the condition $\pi(c, c-\epsilon) = 0$ is satisfied if and only if $\epsilon \geq \bar{\epsilon}(\gamma) = \frac{(2+\gamma)(1-\gamma)^2}{\gamma}(a-c)$. The threshold $\bar{\epsilon}(\gamma)$ is decreasing in γ which is intuitive as γ may be interpreted as an inverse measure of differentiation. Furthermore, the equilibrium profit when both firms produce at marginal cost $c-\epsilon$ is $\pi(c-\epsilon, c-\epsilon) = \frac{1-\gamma}{(2-\gamma)^2(1+\gamma)}(a-c+\epsilon)^2$ which yields $\eta(\epsilon) = \frac{2\epsilon}{a-c+\epsilon}$ and thus $\eta(\epsilon) > 1 \iff \epsilon > a-c$. Since $\bar{\epsilon}(\gamma) \underset{\gamma\downarrow 0}{\longrightarrow} +\infty$ and $\bar{\epsilon}(\gamma) \underset{\gamma\uparrow 1}{\longrightarrow} 0$ and $\bar{\epsilon}(\gamma)$ is continuous and (strictly) decreasing, there exists $\bar{\gamma} \in]0, 1[$ such that $\bar{\epsilon}(\bar{\gamma}) = (a-c)$ and $\bar{\epsilon}(\gamma) < a-c$ if and only if $\gamma < \bar{\gamma}$. Thus, we get the following results:

- If the differentiation between the two products is weak, i.e. $\gamma \geq \bar{\gamma}$, then $\eta(\epsilon) \leq 1$ if $\bar{\epsilon}(\gamma) < \epsilon \leq a-c$ and $\eta(\epsilon) > 1$ if $\epsilon > a-c$. Thus, the former case falls under case 1 of proposition (8): $r_2(\theta) > \theta\epsilon$ for any $\theta \in \left]0, \hat{\theta}\right]$, while the latter falls under case 2: $r_2(\theta) < \theta\epsilon$ at least for θ sufficiently small.

- If the differentiation between the two products is high, i.e. $\gamma < \bar{\gamma}$, then $\eta(\epsilon) > 1$ for any $\epsilon \geq a-c$. Here, it is always true that $r_2(\theta) < \theta\epsilon$ at least for θ sufficiently small.

We can also derive the position of $r_1(\theta)$ relative to $r^e(\theta) = \theta\epsilon$ for θ sufficiently close to 1 (i.e. sufficiently strong patents) from the comparison of $r_1'(1)$ to ϵ. Note that $r_1'(1) = \frac{\pi(c,c)-\pi(c-\epsilon,c-\epsilon)}{\pi_1(c,c)}$. Therefore if the slope $\frac{\pi(c,c)-\pi(c-\epsilon,c-\epsilon)}{\epsilon}$ is strictly greater (resp. smaller) than the negative partial derivative $\pi_1(c,c)$ then $r_1(\theta) < \theta\epsilon$ (resp. $r_1(\theta) > \theta\epsilon$) for sufficiently strong patents.

Figure 1 displays three possible shapes of $r(\theta)$ and illustrates how it may compare to the expected royalty $r^e(\theta) = \theta\epsilon$ in case of litigation.

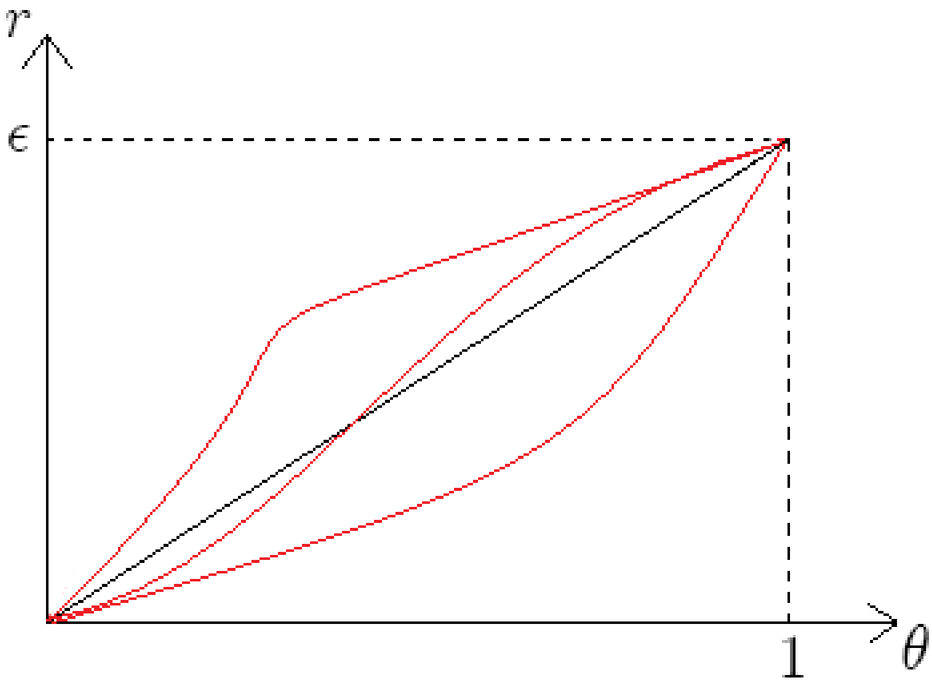

Fig 1: Some possible shapes of $r(\theta)$

2.2.1.3. Comparison of $r(\theta)$ to $r^c(\theta)$

Proposition 9 *The maximal royalty rate (non-cooperatively) accepted by all firms is higher than the maximal royalty rate deterring a collective challenge:* $r(\theta) \geq r^c(\theta)$ *for all* $\theta \in [0, 1]$.

Proof. *See Appendix* ■

An individual challenge by a firm has the following positive externality on the other firms : if the patent is invalidated, all firms, not only the challenger, benefit from the cost reduction for free. This public good nature of an individual challenge gives rise to a free-riding problem that is ruled out when firms act cooperatively. This is why $r(\theta) \geq r^c(\theta)$. Furthermore, the difference $r(\theta) - r^c(\theta)$ can be seen as a measure of the (positive) effect of the free-riding between firms on the patentholder's ability to extract higher royalties.

2.2.1.4. The second stage equilibria

Before turning to the patentholder's optimal licensing contract, we state the equilibria of the second stage in the following proposition

Proposition 10 *For a patentholder's offer involving a royalty rate r, the equilibria of the second stage depend on the royalty rate r as follows:*

i/ if $r \leq r(\theta)$ *then the unique equilibrium is given by all firms accepting the license offer,*

ii/ if $r(\theta) < r \leq \epsilon$ *then the equilibria are the situations where* $(n-1)$ *firms buy a license and one does not,*

iii/ if $r > \epsilon$ *the unique equilibrium is given by all firms refusing the license offer.*

Proof. See Appendix. ■

This proposition states that two potentially profitable possibilities are offered to a holder of an uncertain patent with strength θ when selling licenses through a per-unit royalty rate: either the royalty r is chosen below the maximal value $r(\theta)$ that deters any challenge, and in this case n licenses are sold, or the chosen royalty rate r is above this value ($r(\theta) < r \leq \epsilon$), and in this case one and only one firm challenges the patent validity ($n-1$ licenses are sold). Furthermore, if the royalty rate is above the cost reduction allowed by the innovation, then intuitively, no firm will purchase a license.

2.2.2. The patentholder's optimal license offer: first stage

We turn now to the patentholder's optimal decision at the first stage of the game. Denote $q(c - \epsilon + r, k)$ the individual output of a licensee when the per-unit rate r is accepted by k firms, and the $n - k$ remaining firms produce at marginal cost c.

The patentholder's expected licensing revenues $P(r)$ are given by

$$P(r) = \begin{cases} nrq(c - \epsilon + r, n) & if \quad r \leq r(\theta) \\ \theta(n-1)rq(c - \epsilon + r, n-1) & if \quad r(\theta) < r \leq \epsilon \\ 0 & if \quad r > \epsilon \end{cases}$$

Recall that whenever $r \in]r(\theta), \epsilon]$, one firm refuses the license offer and challenges the patent validity while the other $(n-1)$ firms buy a license (see proposition 10). Therefore, in this case, the patentholder's licensing revenues depend on the issue of litigation (the patent is upheld with probability θ).

We suppose that the following assumptions hold:

A5. A licensee's output is (weakly) decreasing in the number of licensees: $q(c-\epsilon+r, n-1) \geq q(c-\epsilon+r, n)$ for all $r \in [0, \epsilon]$.

A6. The aggregate output is (weakly) increasing in the number of licensees: $Q(c-\epsilon+r, n) \geq Q(c-\epsilon+r, n-1)$ for all $r \in [0, \epsilon]$.

A7. The function $r \rightarrow rq\,(c-\epsilon+r, k)$ is concave over $[0, \epsilon]$ for $k \in \{n-1, n\}$.

Assumption A7 ensures that the patentholder's licensing revenues are concave both in the range of the royalty values where litigation is deterred, that is, $[0, r(\theta)]$, and in the region of the royalty values where it is not, that is $]r\,(\theta)\,, \epsilon]$. This assumption is quite reasonable since a higher royalty rate is likely to have a negative effect on the (equilibrium) demand adressed to each licensee which would make the licensing revenues subject to two opposite effects, possibly resulting in a concave shape for the licensing revenues.

Note that assumptions A5, A6 and A7 are satisfied under Cournot competition with a linear demand.

Denote $\tilde{r}_k\,(\epsilon) = \underset{0 \leq r \leq \epsilon}{\arg\max} rq\,(c-\epsilon+r, k)$ for $k \in \{n-1, n\}$.

Determining the maximum of $P(r)$ over $[0, r(\theta)]$ and $[r(\theta), \epsilon]$ amounts to comparing ϵ and $\tilde{r}_k\,(\epsilon)$ for $k = n-1, n$. This leads to different outcomes according to the location of ε with respect to $\tilde{r}_{n-1}\,(\epsilon)$ and $\tilde{r}_n\,(\epsilon)$.

The following lemma is useful for the subsequent analysis:

Lemma 11 *If* $\epsilon \leq \tilde{r}_{n-1}\,(\epsilon)$ *then* $\epsilon \leq \tilde{r}_n\,(\epsilon)$.

Proof. See Appendix ■

A straightforward consequence of the lemma is that if $\epsilon > \tilde{r}_n\,(\epsilon)$ then $\epsilon > \tilde{r}_{n-1}\,(\epsilon)$ as well. Therefore, only three cases have to be investigated: i/ $\epsilon \leq \tilde{r}_{n-1}\,(\epsilon)$; ii/ $\tilde{r}_{n-1}\,(\epsilon) < \epsilon \leq \tilde{r}_n\,(\epsilon)$; iii/ $\epsilon > \tilde{r}_n\,(\epsilon)$.

The following propositions determine the patentholder's optimal choice $r^*(\theta)$ in each of these cases and identify the conditions under which the subgame perfect equilibrium involves no litigation (i.e no challenge of the patent validity).

Proposition 12 *If $\epsilon \leq \tilde{r}_{n-1}(\epsilon)$, the function $s(\theta)$ defined as the unique solution in r to the equation $nrq(c-\epsilon+r,n) = \theta(n-1)\epsilon q(c,n-1)$ is convex over $[0,1]$, satisfies $s(0)=0$, $s(1)<\epsilon$, and the per-unit royalty that maximizes the licensing revenues is given by:*

$$r^*(\theta) = \begin{cases} r(\theta) & \text{if} \quad r(\theta) \geq s(\theta) \\ \epsilon & \text{if} \quad r(\theta) < s(\theta) \end{cases}$$

In this case, litigation is deterred at equilibrium if and only if $r(\theta) \geq s(\theta)$

Proof. See Appendix ■

This proposition characterizes the optimal royalty rate for the patentholder when the magnitude ϵ of the cost reduction is such that $\epsilon \leq \tilde{r}_{n-1}(\epsilon)$. First, the function $s(\theta)$ defines the royalty rate level for which the patentholder is indifferent between selling n licenses at the price $r(\theta)$ and selling $(n-1)$ licenses at the higher price ϵ (in which case litigation occurs and the expected licensing revenues are $\theta(n-1)\epsilon q(c,n-1)$[24]). Note that when $\epsilon \leq \tilde{r}_{n-1}(\epsilon)$, if the license is sold to only $(n-1)$ firms, the optimal royalty rate is ϵ because the licensing revenue is an increasing concave function of r over $[0,\epsilon]$. Second, the comparison between the maximal rate $r(\theta)$ acceptable by all firms and the royalty rate $s(\theta)$ leads to the following decision: if $r(\theta) \geq s(\theta)$ it is optimal to set $r^*(\theta) = r(\theta)$ and this choice deters litigation; if $r(\theta) < s(\theta)$ it is optimal to set a higher price $r^*(\theta) = \epsilon$ and to let one

[24]Note that $q(c,n-1) = q(c,n)$. The LHS quantity refers to a situation where $n-1$ firms in the industry produce at a marginal cost $c-\epsilon$ and pay a per-unit royalty ϵ while one firm produces at marginal cost c. The RHS quantity refers to a situation where all firms produce at a marginal cost $c-\epsilon$ and pay a per-unit royalty ϵ. in both situations all firms in the industry produce at the same effective marginal cost, that is c which results in the equality stated above.

firm challenge the patent validity. Note that if $r(\theta)$ is convex and the curves $r(\theta)$ and $s(\theta)$ meet at only one point over $]0,1[$, then the curve $r(\theta)$ necessarily intersects the curve $s(\theta)$ from below since $r(0) = s(0) = 0$ and $s(1) < r(1) = \epsilon$. This implies that for low values of θ, we have $r(\theta) < s(\theta)$ and the optimal per-unit royalty rate is then independent of θ and is the same as if the patent were certain. A similar result appears in Farrell and Shapiro (2007) but the justification is different here. While Farrell and Shapiro consider only the case where the cost reduction magnitude ϵ is small enough and assume that all firms buy a license at equilibrium, we obtain this result by allowing the number of licensees to depend on the per-unit royalty. It is precisely when the royalty at which all firms accept to buy a license is too low (i.e. $r(\theta) < s(\theta)$) that the holder of a weak patent prefers to sell it at the higher price ϵ, triggering thus a patent litigation.

We turn now to the second case where $\tilde{r}_{n-1}(\epsilon) < \epsilon \leq \tilde{r}_n(\epsilon)$.

Proposition 13 *If $\tilde{r}_{n-1}(\epsilon) < \epsilon \leq \tilde{r}_n(\epsilon)$, then defining $v(\theta)$ as the unique solution in r to the equation $nrq(c-\epsilon+r,n) = (n-1)\theta\tilde{r}_{n-1}(\epsilon)\,q(c-\epsilon+\tilde{r}_{n-1}(\epsilon), n-1)$, and $\tilde{\theta}_{n-1}$ as the solution to the equation $r(\theta) = \tilde{r}_{n-1}(\epsilon)$, the function $v(\theta)$ is convex over $[0,1]$, $v(0) = 0$, $v(1) < \epsilon$, and we have:*

$$r^*(\theta) = \begin{cases} \tilde{r}_{n-1}(\epsilon) & \textit{if} \quad \theta < \tilde{\theta}_{n-1} \textit{ and } r(\theta) < v(\theta) \\ r(\theta) & \textit{otherwise} \end{cases}$$

In this case, litigation is deterred at equilibrium if and only if at least one of the two following conditions hold: $\theta \geq \tilde{\theta}_{n-1}$ or $r(\theta) \geq v(\theta)$

Proof. See Appendix. ■

To interpret this proposition, one must first note that if the patentholder finds it optimal to trigger a litigation by selling at a royalty $r > r(\theta)$, the optimal royalty rate is given by $\tilde{r}_{n-1}(\epsilon)$ since $\tilde{r}_{n-1}(\epsilon) < \epsilon$. The expected licensing revenues are therefore equal to $(n-1)\theta\tilde{r}_{n-1}(\epsilon)\, q(c-\epsilon+\tilde{r}_{n-1}(\epsilon), n-1)$. The function $v(\theta)$ defines the royalty rate level for which the patentholder is indifferent between selling n licenses at $r(\theta)$ and selling $(n-1)$ licenses at the price $\tilde{r}_{n-1}(\epsilon)$. Second, it is optimal to sell only $(n-1)$ licenses at the per-unit royalty $\tilde{r}_{n-1}(\epsilon)$ as long as $v(\theta) > r(\theta)$ and $\theta < \tilde{\theta}_{n-1}$ where $\tilde{\theta}_{n-1}$ is the solution to the equation $r(\theta) = \tilde{r}_{n-1}(\epsilon)$. This means that the holder of a weak patent $(\theta < \tilde{\theta}_{n-1})$ prefers to trigger a patent litigation by selling licenses at a per-unit royalty rate $\tilde{r}_{n-1}(\epsilon)$ when the royalty that all the firms accept is too low $(r(\theta) < v(\theta))$. Again, this extends the result obtained by Farrell and Shapiro (2007) in the sense that the optimal per-unit royalty rate $r^*(\theta)$ for a weak patent $(\theta < \tilde{\theta}_{n-1})$ is independent of the patent strength θ (provided that $r(\theta) < v(\theta)$). However, there is a major difference with Farrell and Shapiro (2007) and our finding in the case $\epsilon \leq \tilde{r}_{n-1}(\epsilon)$: the royalty $r^*(\theta)$ for such weak patents is not equal to the optimal royalty rate if the patent were certain. Indeed, it is equal to $\tilde{r}_{n-1}(\epsilon) < \epsilon = r^*(1)$. We turn now to the case $\epsilon > \tilde{r}_n(\epsilon)$.

Proposition 14 *If $\epsilon > \tilde{r}_n(\epsilon)$ then, defining $\tilde{\theta}_n$ as the unique solution to the equation $r(\theta) = \tilde{r}_n(\epsilon)$, we have*

$$r^*(\theta) = \begin{cases} \tilde{r}_{n-1}(\epsilon) & \text{if} \quad \theta \leq \min\left(\tilde{\theta}_{n-1}, \tilde{\theta}_n\right) \text{ and } r(\theta) < v(\theta) \\ \tilde{r}_n(\epsilon) & \text{if} \quad \theta \geq \tilde{\theta}_{n-1} \\ r(\theta) & \text{otherwise} \end{cases}$$

In this case, litigation is deterred at equilibrium if and only if at least one of the two following conditions hold: $\theta > \min(\tilde{\theta}_{n-1}, \tilde{\theta}_n)$ or $r(\theta) \geq v(\theta)$.

Proof. See Appendix. ■

The interpretation is the same as in the two previous propositions except that for $\epsilon > \tilde{r}_n(\epsilon)$ (implying that $\epsilon > \widetilde{r}_{n-1}(\epsilon)$), when it is optimal for the patentholder to trigger a litigation by selling at a royalty $r > r(\theta)$, the optimal royalty rate is given either by $\tilde{r}_{n-1}(\epsilon)$ or $\tilde{r}_n(\epsilon)$. Again the optimal per-unit rate of a weak patent may not depend on the patent's strength but it is not equal to the royalty rate as if the patent were certain. Indeed, whenever $\theta \leq \min\left(\tilde{\theta}_{n-1}, \tilde{\theta}_n\right)$ and $r(\theta) < v(\theta)$, it holds that $r^*(\theta) = \tilde{r}_{n-1}(\epsilon)$ whereas $r^*(1) = \tilde{r}_n(\epsilon)$.

The following proposition gives a sufficient condition for litigation deterrence.

Corollary 15 *If $r(\theta) > \theta\epsilon$ then the patentholder finds it optimal to deter litigation and $P(r^*(\theta)) > \theta P(r^*(1))$.*

Proof. See Appendix. ■

This corollary gives a justification for the use of the expected value of the royalty in case of litigation $r^e(\theta) = \theta\epsilon$ as a benchmark. If the maximal royalty rate $r(\theta)$ acceptable by all firms is above this value, then the patentholder will always prefer to deter litigation. More importantly, deterring a potential challenge in this case will not prevent the patentholder from making a relatively high profit. Specifically, it gets a profit higher than $\theta P(r^*(1))$ which represents the patentholder's expected profit if the patent validity issue were resolved *before* the license deal takes place. It can also be interpreted as the patentholder's expected profit if it were granted a full-proof patent by the patent office with probability θ (see Farrell and Shapiro 2007). Hence, the per-unit royalty scheme may allow the patentholder to reap some extra profit relative to the natural benchmark $\theta P(r^*(1))$.

Example 1 (continued): Cournot oligopoly with linear demand

Under this specification, we show that $\tilde{r}_n(\epsilon) = \frac{a-c+\epsilon}{2}$ and $\tilde{r}_{n-1}(\epsilon) = \frac{a-c+n\epsilon}{2n}$. Hence, in this case, it holds that $\tilde{r}_n(\epsilon) > \tilde{r}_{n-1}(\epsilon)$ for any $\epsilon > 0$. Furthermore, $\epsilon \geq \tilde{r}_{n-1}(\epsilon)$ holds if and only if $\epsilon \geq \frac{a-c}{n}$, and $\epsilon \geq \tilde{r}_n(\epsilon)$ holds if and only if $\epsilon \geq a-c$. The latter condition happens to be the condition defining a drastic cost-reducing innovation under Cournot competition (with linear demand). Hence, the condition $\epsilon \geq \tilde{r}_n(\epsilon)$ has a simple interpretation in this special case.

Consider now an innovation $\epsilon \in \left[\frac{a-c}{n-1}, a-c\right]$. Such innovation may fall either under the case $\epsilon \geq \tilde{r}_n(\epsilon)$ or under $\tilde{r}_{n-1}(\epsilon) \leq \epsilon < \tilde{r}_n(\epsilon)$. However, we know from a previous part of this example that for $\epsilon \in \left[\frac{a-c}{n-1}, a-c\right]$, it holds that $r(\theta) = r_2(\theta) > \theta\epsilon$ for any $\theta \in \left[0, \hat{\theta}\right]$. Therefore, corollary (15) and its proof allow us to state directly that litigation will not occur at equilibrium and the optimal royalty will be $r^*(\theta) = \min(r_2(\theta), \tilde{r}_n(\epsilon))$ for any $\theta \in \left[0, \hat{\theta}\right]$. Since $\epsilon \leq a-c$ then $\epsilon \leq \tilde{r}_n(\epsilon)$ which results in $r(\theta) \leq \tilde{r}_n(\epsilon)$ (because $r(\theta) \leq \epsilon$) and yields: $r^*(\theta) = r_2(\theta)$ for any $\theta \in \left[0, \hat{\theta}\right]$. The remaining part of corollary (15) ensures that such a royalty will result in a profit higher than the benchmark profit $\theta P(r^*(1))$.

2.2.3. Introduction of renegotiation

So far we have assumed that in case of litigation, an unsuccessful challenger produces with marginal cost c because the patentholder refuses to sell him a license. Whether such a commitment to refuse a license to an unsucessful challenger is credible or not must be discussed. From the challenger's perspective this commitment is equivalent to an offer of a new licensing contract involving a royalty rate $\bar{r} = \epsilon$. However, from the patentholder's perspective, this equivalence does not hold. Moreover a situation where an unsuccessful challenger is offered a new licensing contract involving a royalty rate $\bar{r} < \epsilon$ may be preferred by the patentholder to a situation where it is offered a contract based on $\bar{r} = \epsilon$. Such an issue is important since a potential challenger will take the decision whether to accept the license or contest the patent validity, anticipating what would happen if the patent is validated.

Formally if we allow for renegotiation when $(n-1)$ firms accept a licensing contract based on a royalty rate r and the remaining firm challenges the patent unsuccessfully, then the patentholder will offer to the challenger a contract involving a royalty rate $\bar{r} \in [0, \epsilon]$ that maximizes its licensing revenues

$$P(r, \bar{r}) = (n-1)\, rq^L\, (c - \epsilon + r, c - \epsilon + \bar{r}) + \bar{r}q^{NL}\, (c - \epsilon + r, c - \epsilon + \bar{r})$$

where $q^L\,(c - \epsilon + r, c - \epsilon + \bar{r})$ denotes the equilibrium quantity produced by each of the $(n-1)$ firms that accepted initially the license offer r and $q^{NL}\,(c - \epsilon + r, c - \epsilon + \bar{r})$ is the equilibrium quantity produced by the unsuccessful challenger who produces at marginal cost $c - \epsilon + \bar{r}$. If $\bar{r}(r)$ is the royalty rate that maximizes $P(r, \bar{r})$ with respect to $\bar{r}$, a licensing contract involving a royalty rate r will be accepted by all the firms if and only if:

$$\pi(c - \epsilon + r, c - \epsilon + r) \geq \theta\pi(c - \epsilon + \bar{r}\,(r)\,, c - \epsilon + r) + (1-\theta)\pi(c - \epsilon, c - \epsilon) \qquad (2.4)$$

Since $\bar{r}\,(r) \leq \epsilon$ we have $\pi(c - \epsilon + \bar{r}\,(r)\,, c - \epsilon + r) \geq \pi(c, c - \epsilon + r)$ which entails that constraint (2.4) is (weakly) more stringent than (2.1). More specifically, a royalty rate r could be accepted if the patentholder commits to refuse a license to a challenger or license him at $\bar{r} = \epsilon$, but not accepted if he cannot commit. This implies that the maximal royalty rate the patentholder can make the n firms pay is (weakly) smaller when renegotiation of a licensing contract (after patent validation) is introduced.

We investigate in more detail this issue in the following example.

Example 1 (continued): Cournot oligopoly with linear demand

Denote firm n the challenging firm and $\bar{r}$ the per-unit royalty rate at which a license is offered if the challenge fails. Cournot competition between $(n-1)$ firms (indexed by $i = 1, 2, ..., n-1$) whose marginal cost is $c - \epsilon + r$ and firm n whose marginal cost is $c - \epsilon + \bar{r}$ leads to the following equilibrium outputs:

$$q_i(r, \bar{r}) = \begin{cases} \frac{a-c+\epsilon-2r+\bar{r}}{n+1} & \text{if} \quad i = 1, .., n-1 \\ \frac{a-c+\epsilon-n\bar{r}+(n-1)r}{n+1} & \text{if} \quad i = n \end{cases}$$

For a given r, the value of the royalty rate $\bar{r}$ that maximizes the patentholder's licensing revenue is the solution to the following program:

$$\max_{\bar{r} \in [0,\epsilon]} P(r, \bar{r}) = (n-1)r\frac{a-c+\epsilon-2r+\bar{r}}{n+1} + \bar{r}\frac{a-c+\epsilon-n\bar{r}+(n-1)r}{n+1}$$

Suppose that the innovation is non-drastic, i.e. $\epsilon < a - c$. The unique unconstrained maximum of the concave function $\bar{r} \longrightarrow P(r, \bar{r})$ is given by the FOC $\frac{\partial P(r,\bar{r})}{\partial \bar{r}} = \frac{a-c+\epsilon+2(n-1)r-2n\bar{r}}{n+1} = 0$. The maximum of the function $P(r, \bar{r})$ over the interval $\bar{r} \in [0, \epsilon]$ is reached at

$$\bar{r}(r) = \min\left(\epsilon, \frac{(n-1)}{n}r + \frac{a-c+\epsilon}{2n}\right) = \min\left(\epsilon, r + \frac{1}{n}\left(\frac{a-c+\epsilon}{2} - r\right)\right)$$

Since $\epsilon < a-c$, we have $\frac{a-c+\epsilon}{2} > \epsilon$. Therefore, $r \in [0, \epsilon] \Longrightarrow \frac{a-c+\epsilon}{2} - r \geq 0$ and consequently $\bar{r}(r) \geq r$. Hence a firm which refuses a licensing contract and unsuccessfully challenges the patent validity will get a new licensing offer with a higher royalty rate than the royalty paid by licensees that have accepted the initial licensing contract.[25]

[25] It is obvious that the patentholder's position is stronger after the patent has been upheld by the court

Moreover, the condition $\bar{r}(r) < \epsilon$ is fulfilled if and only if $r < (\frac{2n-1}{2(n-1)})\epsilon - \frac{a-c}{2(n-1)} \equiv \varphi$, which is positive whenever $\epsilon > \frac{a-c}{2n-1}$. For such a royalty rate r, we have $\pi(c-\epsilon+\bar{r}(r), c-\epsilon+r) = [\frac{a-c+\epsilon-n\bar{r}(r)+(n-1)r}{n+1}]^2$, and the condition expressing that all firms accept the licensing contract r is:

$$\pi(c-\epsilon+r, c-\epsilon+r) \geq \theta\pi(c-\epsilon+\bar{r}(r), c-\epsilon+r) + (1-\theta)\pi(c-\epsilon, c-\epsilon)$$

Replacing $\bar{r}(r)$ by its value, one obtains:

$$\frac{(a-c+\epsilon-r)^2}{(n+1)^2} \geq \theta\left[\frac{a-c+\epsilon}{2(n+1)}\right]^2 + (1-\theta)\left[\frac{a-c+\epsilon}{(n+1)}\right]^2 = \frac{4-3\theta}{4}\left[\frac{a-c+\epsilon}{(n+1)}\right]^2$$

This inequality is satisfied if and only if:

$$r \leq (a-c+\epsilon)\left(1-\frac{\sqrt{4-3\theta}}{2}\right)$$

Hence a royalty rate $r < \varphi$ is accepted by all firms if and only if the previous inequality holds. Denoting $\bar{\theta}$ the unique solution in θ to the equation $(a-c+\epsilon)\left(1-\frac{\sqrt{4-3\theta}}{2}\right) = \varphi$, we can then state that for $\theta \leq \bar{\theta}$, the maximal royalty rate accepted by all firms when post-trial license offer is possible is given by:

$$r^p(\theta) = (a-c+\epsilon)\left(1-\frac{\sqrt{4-3\theta}}{2}\right)$$

Straightforward computations lead to $\frac{dr^p}{d\theta}(0) = \frac{3}{8}(a-c+\epsilon)$. It is easy to show that $\frac{dr^p}{d\theta}(0) < \epsilon$ for any $\epsilon \in \left]\frac{3}{5}(a-c), a-c\right[$. Consequently for such intermediate innovations, $r^p(\theta) < \theta\epsilon$ for sufficiently small values of θ. Note that for such innovations, the condition

than before.

$\epsilon > \frac{a-c}{2n-1}$ is satisfied since $\frac{3}{5}(a-c) > \frac{a-c}{2n-1}$ for any $n \geq 2$.

These results lead to the following proposition:

Proposition 16 *Assume renegotiation is possible. In a Cournot model with homogeneous product and a linear demand $Q = a - p$, the maximal per-unit royalty rate that induces a perfect subgame equilibrium in which all firms choose to buy a license of a patented technology that reduces the marginal cost by $\epsilon \in \left]\frac{3}{5}(a-c), a-c\right[$ is given by $r^p(\theta) = (a-c+\epsilon)\left(1 - \frac{\sqrt{4-3\theta}}{2}\right)$ for a patent strength θ smaller than a threshold $\overline{\theta} \in]0,1[$. The royalty $r^p(\theta)$ is sustained by a renegotiated royalty $\bar{r}(r^p(\theta)) < \epsilon$, and is smaller than the benchmark $r^e(\theta) = \theta\epsilon$ if the patent is sufficiently weak.*

Under the conditions stated in this proposition, the maximal royalty rate $r^p(\theta)$ (accepted by all firms) if renegotiation is possible is below the benchmark $r^e(\theta) = \theta\epsilon$ whereas the maximal royalty rate $r(\theta)$ if renegotiation is not possible is above the benchmark $r^e(\theta) = \theta\epsilon$ under the same conditions (see the application of Proposition 8 to Cournot competition). More generally, this proposition illustrates the role of post-trial renegotiation in licensing an uncertain patent. An individual challenge becomes more attractive when it is possible to renegotiate *ex post* a new royalty after the issue of the trial. Consequently, the patentee looses some of its market power in determining *ex ante* the per-unit royalty rate that deters litigation. For this reason, refusing a license to an unsuccessful challenger should not be allowed.

2.3. Fixed fee licensing schemes

In this section, we examine licensing contracts involving a fixed fee only. We consider the same three-stage game as in the per-unit royalty licensing scheme, simply replacing the royalty rate by a fixed fee in the licensing contract offered by the patentholder.

Denote $\pi^{L}(k)$ (respectively $\pi^{NL}(k)$) the equilibrium profit of a firm producing at a constant marginal cost $c-\epsilon$ (respectively c) in an industry of n firms, out of which k firms produce at marginal cost $c-\epsilon$ and the remaining $n-k$ firms produce at marginal cost c.

We set the following assumption which states that a licensee's profit (gross of the license fee) when all firms buy the license is higher than a non-licensee's profit whatever the number of licensees.

A8: $\pi^{NL}(k) < \pi^{L}(n)$ for all $k < n$.

Note that this assumption holds for instance under Cournot competition with linear demand. A stronger assumption would be to set the following two conditions : i/ a non-licensee's profit $\pi^{NL}(k)$ is decreasing in the number of licensees k, which would the counterpart of assumption A3 in this setting, ii/ a firm's profit if all firm buy a license is greater than a firm's profit if no firms purchases a license, i.e. $\pi^{NL}(0) < \pi^{L}(n)$, which would be the counterpart of assumption A4. It is straightforward that assumption A8 can be derived from i/ and ii/.

We start with a preliminary result describing what happens at equilibrium when not all firms accept the up-front fee.

Lemma 17 *Consider a Nash equilibrium of stage 2. If not all firms accept the licensing contract in this equilibrium then there is at least one firm (among those who do not accept the contract) that challenges the patent validity.*

Proof. See Appendix. ■

In order to derive the demand function for licenses, we set the following assumption:

A9: For all k between 0 and $n-1$,

$$\pi^L(k) - \pi^L(k+1) \geq \pi^{NL}(k-1) - \pi^{NL}(k)$$

This assumption, which holds for instance under Cournot competition with linear demand, states that a licensee's incremental profit is at least equal to a non-licensee's incremental profit when the number of licensees is reduced by one unit. Another way to put this assumption is to state that a firm's willingness to pay for a license, i.e. $\pi^L(k) - \pi^{NL}(k-1)$, is decreasing in the number of licensees k.

2.3.1. Demand function for licenses: second stage

The following proposition gives the demand for licenses at the Nash equilibrium of stage 2 as a function of the up-front fee F chosen by the patentholder P in stage 1:

Proposition 18 *Denote* $F_n(\theta) = \theta\left(\pi^L(n) - \pi^{NL}(n-1)\right)$ *and* $F_k = \pi^L(k) - \pi^{NL}(k-1)$ *for all* $k \leq n-1$.

- If $F < F_n(\theta)$ *then the unique Nash equilibrium of stage 2 is the situation where all firms accept the licensing contract.*

- If $F_n(\theta) < F < F_{n-1}$ *then the Nash equilibria of stage 2 are the situations where* $n-1$ *firms accept the licensing contract and one firm does not.*

- For any k *between* 0 *and* $n-2$*, if* $F_{k+1} < F < F_k$ *then the Nash equilibria of stage 2 are the situations where* k *firms accept the licensing contract and the remaining* $n-k$ *firms do not.*

- If $F > F_1$ *then the unique Nash equilibrium of stage 2 is the situation where all firms reject the licensing contract.*

Proof. See Appendix. ■

To avoid the multiple equilibria problem that arises when F is equal to one of the threshold values F_k we assume that a firm which is indifferent between accepting the license offer made the patentholder and refusing it chooses to accept it. Hence, we can define the number $k(F, \theta)$ of firms that accept at equilibrium the license offer F made by the patentholder:

$$k(F,\theta) = \begin{cases} n & if \quad F \leq F_n(\theta) \\ n-1 & if \quad F_n(\theta) < F \leq F_{n-1} \\ ... & . \quad ... \\ k & if \quad F_{k+1} < F \leq F_k \\ ... & . \quad ... \\ 0 & if \quad F > F_1 \end{cases}$$

Note that $k(F, \theta)$ depends on θ only through the threshold $F_n(\theta)$. More specifically, if we denote $F_n(1) = F_n$ we have $F_n(\theta) = \theta F_n$ and $F > F_n(\theta)$ implies $k(F, \theta) = k(F)$.

2.3.2. Choice of the fixed fee: first stage

The patentholder will choose F so as to maximize its licensing revenues anticipating the number of firms that will accept the license offer. If the up-front fee F is such that all firms accept the offer then the patentholder's licensing revenues are equal to nF. If the up-front fee is such that there is at least one firm that does not accept the offer then litigation occurs and the patentholder gets licensing revenues equal to $k(F)F$ only when the patent validity is upheld by the court. This happens with probability θ which entails that the expected licensing revenues of the patentholder when F induces a number of licensees k smaller than n are equal to $\theta k(F)F$. The expected licensing revenues of the patentholder as a function of the up-front fee F can be summarized as follows:

$$P(F,\theta) = \begin{cases} nF & if \quad F \leq F_n(\theta) \\ \theta(n-1)F & if \quad F_n(\theta) < F \leq F_{n-1} \\ ... & . \quad ... \\ \theta kF & if \quad F_{k+1} \leq F \leq F_k \\ ... & . \quad ... \\ 0 & if \quad F > F_1 \end{cases}$$

Since the demand function for licenses is stepwise, the maximization of $P(F,\theta)$ with respect to F will lead to one (or several) of the thresholds $F_n(\theta)$ and $F = F_k$, $k \leq n-1$. In other words, the maximization program $\max\limits_{F \geq 0} P(F,\theta)$ is equivalent to the maximization program:

$$\max_{F \in \{F_1,...,F_k,...,F_{n-1},F_n(\theta)\}} P(F,\theta)$$

Since $F_n(\theta) = \theta F_n$, the expected licensing revenues $P(F,\theta)$ for a value of F belonging to the set $\{F_1,...,F_k,...,F_{n-1},F_n(\theta)\}$ is given by:

$$P(F,\theta) = \begin{cases} n(\theta F_n) & if \quad F = \theta F_n \\ \theta(n-1)F_{n-1} & if \quad F = F_{n-1} \\ ... & . \quad ... \\ \theta kF_k & if \quad F = F_k \\ ... & . \quad ... \\ \theta F_1 & if \quad F = F_1 \end{cases}$$

This shows that for any $\theta \neq 0$, maximizing $P(F,\theta)$ over the set $\{F_1, ..., F_k, ..., F_{n-1}, F_n(\theta)\}$ is equivalent to maximizing $P(F,1)$ over the set $\{F_1, ..., F_k, ..., F_{n-1}, F_n\}$ in the following sense: if the maximum of $P(F,1)$ is reached at F_k then the maximum of $P(F,\theta)$ is reached at F_k if $k < n$ and at $F_n(\theta) = \theta F_n$ if $k = n$. Hence, we have the following result:

Proposition 19 *If the maximum of $P(F,1)$ is reached at $F^* = F_n$ then the patentholder offers a licensing contract with a fixed fee $F^*(\theta) = F_n(\theta) = \theta F_n$ that induces a number of licensees equal to the total number of firms. If the maximum of $P(F,1)$ is reached at $F^* = F_k$ with $k < n$ then the patentholder offers a licensing contract with an up-front fee $F^* = F_k$ that induces a number of licensees equal to k.*

This proposition entails the following two results:

Corollary 20 *The equilibrium number of licensees k^* does not depend on the patent strength θ.*

Proof. The previous proposition shows that the choice of the fixed fee by the patentholder does not depend on θ. Since the number of licensees is determined by the value of the up-front fee fixed by the patentholder it follows that the equilibrium number of licensees does not depend on the patent strength θ. ■

Corollary 21 *The equilibrium expected licensing revenues of the patentholder under an up-front fee regime, denoted $P_F^*(\theta) = P(F^*(\theta), \theta)$, are proportional to the patent strength, i.e.:*

$$P_F^*(\theta) = \theta P_F^*(1)$$

Proof. If the patentholder offers a licensing contract at an up-front fee $F = F_n(\theta) = \theta F_n$ then its equilibrium licensing revenues are $P_F^*(\theta) = n(\theta F_n) = \theta(nF_n) = \theta P_F^*(1)$. If the patentholder offers a licensing contract at an up-front fee $F = F_k$ where $k < n$, then its equilibrium licensing revenues are $P_F^*(\theta) = \theta (kF_k) = \theta P_F^*(1)$. ■

The results of this section lead to the conclusion that licensing an uncertain patent by means of a fixed fee is not affected by the uncertainty, in the sense that the number of licensees does not depend on the patent strength and the patentholder's licensing revenues are exactly proportional to the patent strength. These results are very different from those obtained with a per-unit royalty rate (previous section) or with a two-part tariff as in Farrell and Shapiro (2007). In particular under the up-front fee regime, the patentholder cannot reap any extra profit relative to the benchmark $\theta P_F^*(1)$.This leads to a first conclusion: licensing weak patents is very sensitive to the chosen licensing scheme.

We turn now to comparing the licensing revenues collected through the two schemes.

2.4. Royalty rate vs. fixed fee

In this section we show that, at least under some circumstances, the patentholder prefers to use a royalty rate rather than a fixed fee in the licensing contracts it proposes. Denote $P_r^*(\theta) = P(r^*(\theta))$ the optimal patentholder's profit when the per-unit royalty licensing scheme is used.

Proposition 22 *If the patentholder gets higher licensing revenues when using the royalty rate scheme than with the fixed fee scheme when the patent' svalidity is certain, i.e.* $\theta = 1$*, it will also prefer to use a royalty rate rather than a fixed fee when the patent validity is uncertain, i.e.* $\theta < 1$*, whenever* $r(\theta) > \theta\epsilon$*.*

Proof. This follows immediately from the fact that $P_r^*(\theta) > \theta P_r^*(1)$ whenever $r(\theta) > \theta\epsilon$, whereas it always holds that $P_F^*(\theta) = \theta P_F^*(1)$. Therefore, if $P_r^*(1) > P_F^*(1)$ then $P_r^*(\theta) >$

$\theta P_r^*(1) \geq \theta P_F^*(1) = P_F^*(\theta)$ which means that the patentholder's licensing revenues are higher when the royalty rate mechanism is used. ■

This proposition gives only a sufficient condition for royalty rate contracts to be preferred over up-front fee contracts when the innovation is covered by an uncertain patent. If royalties are preferred to fixed fees when $\theta = 1$, the former will be also preferred to the latter for $\theta < 1$ whenever $r(\theta) > \theta\epsilon$. However, this condition is far from necessary as the following example shows: The royalty rate mechanism may be preferred for some values of $\theta < 1$ even when the fixed fee mechanism is preferred when $\theta = 1$.

Example 1 (continued): Cournot competition with a linear demand

We know from Kamien and Tauman (1986) that in a full-proof patent setting, i.e. $\theta = 1$, the patentholder's licensing revenues are higher with an up-front fee than with a royalty rate.[26] We show hereafter that this ranking need not hold when the patent is uncertain: the patentholder may prefer to use the royalty rate mechanism rather than the fixed fee mechanism.

We consider innovations of intermediate magnitude, i.e. $\frac{a-c}{n-1} < \epsilon < a - c$ protected by relatively weak patents, i.e. $\theta \in \left]0, \hat{\theta}\right[$ with $\hat{\theta} = 1 - [\frac{n(a-c)}{(n-1)(a-c+\epsilon)}]^2$. We focus on this case because under those conditions the royalty rate the patentholder will set has a simple analytical form $r^*(\theta) = r_2(\theta) = [1 - \sqrt{1-\theta}](a - c + \epsilon)$. This allows us to compute the quantity produced at equilibrium by each firm: $q\left(c - \epsilon + r(\theta), n\right) = \frac{\sqrt{1-\theta}(a-c+\epsilon)}{n+1}$. The equilibrium licensing revenues derived from the royalty $r(\theta) = r_2(\theta)$ are thus given by:

$$P_r^*(\theta) = nr(\theta)q\left(c - \epsilon + r(\theta), n\right) = \frac{n\left(\sqrt{1-\theta} - 1 + \theta\right)(a - c + \epsilon)^2}{n+1}$$

[26] Sen (2005) shows that this result holds only when the number of firms in the downstream industry is not too high.

Kamien and Tauman (1986, proposition 2) gives the patentholder's profit expression when $\theta = 1$. Using this expression and corollary (21), we derive the value of the patentholder revenues for any $\epsilon \in \left]\frac{a-c}{n-1}, a-c\right[$:

$$P_F^*(\theta) = \begin{cases} \frac{2\theta n}{(n+1)^2}\epsilon^2\left[\frac{a-c}{2\epsilon}+\frac{n+2}{4}\right]^2 & \text{if} \quad \frac{a-c}{n-1} < \epsilon \leq \frac{2(a-c)}{n} \\ \frac{\theta n(n+2)}{(n+1)^2}\epsilon\,(a-c) & \text{if} \quad \frac{2(a-c)}{n} \leq \epsilon < a-c \end{cases}$$

Let us compare $P_r^*(\theta)$ and $P_F^*(\theta)$. First note that $P_F^*(\theta)$ is linear in θ while $P_r^*(\theta)$ is concave in θ. Second, these functions take the same value for $\theta = 0$.We can then state that a sufficient condition for $P_r^*(\theta)$ to be greater than $P_F^*(\theta)$ for all $\theta \in \left[0, \hat{\theta}\right]$ is that $P_r^*(\hat{\theta}) \geq P_F^*(\hat{\theta})$. The left-hand member of this inequality is given by $P_r^*(\hat{\theta}) = nr(\hat{\theta})q\left(c-\epsilon+r(\hat{\theta})\right) = n\hat{r}q\,(c-\epsilon+\hat{r}) = \frac{n^2}{(n-1)(n+1)}\left(\epsilon-\frac{a-c}{n-1}\right)(a-c)$ while the right-hand member depends on whether ϵ is such that $\frac{a-c}{n-1} \leq \epsilon \leq \frac{2(a-c)}{n}$ or $\frac{2(a-c)}{n} \leq \epsilon \leq a-c$.

Let us examine the subcase $\frac{2(a-c)}{n} \leq \epsilon \leq a-c$. When this condition is satisfied, we have $P_F^*(\hat{\theta}) = \frac{n(n+2)}{(n+1)^2}\epsilon\,(a-c)\left(1-[\frac{n(a-c)}{(n-1)(a-c+\epsilon)}]^2\right)$. Comparing $P_r^*(\hat{\theta})$ to $P_F^*(\hat{\theta})$ amounts then to comparing $\frac{n}{n-1}\left(\epsilon-\frac{a-c}{n-1}\right)$ to $\frac{n+2}{n+1}\epsilon\left(1-[\frac{n(a-c)}{(n-1)(a-c+\epsilon)}]^2\right)$. A sufficient (and necessary) condition to have $P_r^*(\hat{\theta}) \geq P_F^*(\hat{\theta})$ is $\left(\frac{n}{n-1}-\frac{n+2}{n+1}\right)\epsilon - \frac{n}{n-1}\frac{a-c}{n-1} \geq -\frac{n+2}{n+1}\epsilon[\frac{n(a-c)}{(n-1)(a-c+\epsilon)}]^2$. The left-hand side of this inequality is clearly increasing in ϵ while it is straighforward to show that the right-hand side is decreasing in ϵ. Therefore , to show that the previous inequality holds for any $\epsilon \in \left[\frac{2(a-c)}{n}, a-c\right]$, it is sufficient to show that it holds for $\epsilon = \frac{2(a-c)}{n}$. Taking the inequality for $\epsilon = \frac{2(a-c)}{n}$ and simplifying by $(a-c)$, we get $\frac{n}{n-1}\left(\frac{2}{n}-\frac{1}{n-1}\right) > \frac{2(n+2)}{n(n+1)}\left(1-\frac{n}{(n-1)\left(1+\frac{2}{n}\right)}\right)^2$ which can be shown after some algebraic manipulations to be equivalent to $\frac{n}{n-1} > \frac{2}{n+1}$, which is obviously true. Hence, the condition $P_r^*(\hat{\theta}) \geq P_F^*(\hat{\theta})$ holds for an innovation such that $\frac{2(a-c)}{n} \leq \epsilon \leq a-c$. We can then state the following result:

Proposition 23 *If the firms compete à la Cournot in a market where the demand is linear, then for an innovation* ϵ *of intermediate magnitude, i.e. such that* $\frac{2(a-c)}{n} \leq \epsilon \leq a - c$*, covered by a relatively weak patent, i.e. such that* $\theta \leq \hat{\theta}$*, the patentholder gets higher licensing revenues using a royalty rate rather than a fixed fee, whereas if the patent were perfect the reverse would be true.*

This proposition shows that the uncertainty over patent validity provides an alternative explanation as to why a patentholder may prefer the per-unit royalty scheme over the fixed fee scheme.

2.5. Conclusion

The consequences of licensing uncertain patents have been examined in this paper by addressing the following question: to what extent licensing a patent that has a positive probability to be invalidated if it is challenged favors the patentholder when confronted to potential users in an oligopolistic industry? Our results show that the answer to Farrell and Shapiro's question "How strong are weak patents?" is very sensitive to the choice of the licensing scheme. Two licensing schemes have been examined: the per-unit royalty rate and the up-front fee. The most salient result is that these two mechanisms lead to opposite consequences. While licensing uncertain patents by means of a royalty rate allows in general the patentholder to reap some extra profit relative to the expected profit after the court resolution of the patent validity, a fixed fee regime discards completely this possibility. Under a fixed fee the patentholder obtains exactly its expected revenue. These results mainly arise from letting the number of licensees depend on the price of the license chosen by the patentholder, either a per-unit royalty rate or an up-front fee. Another important result is that under the per-unit royalty licensing regime it may happen that the holder of a weak patent prefers to sell a license at the same royalty rate as if the patent were certain, taking thus the risk of triggering a litigation on patent validity. However, whether such an outcome is possible depends crucially on the innovation size and even when it does occur, the justification is completely different from Farrell and Shapiro (2007). It is precisely when the royalty rate acceptable by all the firms in the downward industry is too low that the holder of a weak patent may prefer to sell at the royalty rate that maximizes its licensing revenues as if the patent were certain. Moreover we have shown that even if fixed fees are preferred when the patent is very strong, royalties may be more profitable if the patent is uncertain, particularly if it is weak. Some classical properties of licensing certain patents may thus be reversed in the uncertain patent framework. We have also explored different policy levers affecting the patentholder's market power when using a per-unit royalty rate. We showed that its market power may be reduced in two ways: First, by preventing the patentholder's refusal to sell a license to an unsuccessful challenger. Second, by favoring collective challenges of patents' validity, particularly when competition intensity in the market is so high that individual incentives to challenge a patent are weak.

One important question concerns the patent quality problem. Since the patent system involves a two-tier process combining patent office examination and challenge by a court of the validity of the granted patent, there are two possible approaches to this problem.

The first approach is to find some ways to encourage third parties to bring to a court pieces of evidence in order to challenge the validity of presumably weak patents (post-grant opposition in Europe or post-grant reexamination in the United States). Giving more incentives to potential licensees to challenge a patent validity is necessary because the free riding aspect weakens individual incentives. In this perspective, two policy levers are suggested: the renegotiation of the licensing contract with an unsuccessful challenger and the cooperative approach among potential licensees to collectively accept or refuse a licensing contract. Incentives to renegotiate could be encouraged by not allowing a patentee to refuse a license to an unsuccessful chalenger. Encouraging a joint decision for accepting or refusing a licensing contract may also reduce the patentholder's market power.

The second approach to the patent quality problem is to improve the screening process inside the patent office itself through the strengthening of the patentability standards, turning back the Lemley's "rational ignorant patent office principle" (Lemley, 2001). This second approach could be interesting, particularly when the patent strength is no more common knowledge but a private information parameter. The patent office could thus propose to any applicant a menu involving the choice of either paying an extra fee to obtain a thorough examination process at the patent office signalling thus a high patent quality or paying a lower fee to simply obtain a "standard" examination process that may signal the weakness of the patent. Designing an efficient mechanism to implement such a procedure is left for future investigation.

2.6. References

Amir, R. and J. Wooders, 2000, One-way spillovers, endogenous innovator/imitator roles, and research joint ventures, *Games and Economic Behavior*, 31, 1-25.

Anand, B. and T. Khanna, 2000, The structure of licensing contracts, *The Journal of Industrial Economics*, 48, 1, 103-135.

Aoki, R. and J-L. Hu, 1999, Licensing versus litigation: the effect of the legal system on incentives to innovate, *The Journal of Economics and Management Strategy*, 8, 133-160.

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

Ayres, J. and P. Klemperer, 1999, Limiting patentees market power without reducing innovation incentives: the perverse benefits of uncertainty and non-injunctive remedies, *Michigan Law Review,* Vol. 97, 985-1033.

Bebchuk, L., 1984, Litigation and settlement under imperfect information, *The Rand Journal of Economics*, 15,3, 404-415.

Boone, J., 2001, Intensity of competition and the incentives to innovate, *International Journal of Industrial Organization,* 19, 705-726.

Bousquet, A., H. Cremer, M. Ivaldi and M. M. Wolkowicz, 1998, Technology transfer with moral hazard, *International Journal of Industrial Organization*, 16, 535-554.

Brocas, I., 2006, Designing auctions in R&D: optimal licensing of an innovation, *Topics in Economic Analysis & Policy*, 6, Issue 1, Article 11.

Caillaud, B. and A. Duchêne, 2005, Patent office in innovation policy: No-body's perfect, mimeo available at http://ceco.polytechnique.fr/Colloques/ETAPE/duchene.pdf.

Chiou, J-Y., 2008, The patent quality control process: Can we afford an (rationally) ignorant patent office?, mimeo.

Crampes, C. and C. Langinier, Litigation and settlement in patent infringement cases,*Rand Journal of Economics*, 33(2), 258-274.

Encaoua, D. and A. Hollander, 2002, Competition policy and innovation, *Oxford Review of Economic Policy,* 18, 1, 63-79.

Farrell, J. and R. Merges, 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, 2, 943-970.

Farrell, J. and C. Shapiro, 2007, How strong are weak patents?, Working Paper, University of California, Berkeley, forthcoming, *American Economic Review.*

Février, P and L. Linnemer, 2004, Idiosyncratic shocks in a asymetric Cournot oligopoly, *International Journal of Industrial Organization*, 22, 6, 835-848.

Graham, S., B. Hall, D. Harhoff and D. Mowery, 2003, Post-issue patent "quality control": a comparative study of US patent re-examinations and European patent oppositions, in W. Cohen and S. Merrill, eds.: *Patents in the Knowledge-Based Economy,* The National academic Press, Washington, D.C.

Guellec, D. and B. von Pottelsberghe de la Potterie, 2007, *The Economics of the European Patent System*, Oxford University Press, Oxford.

Hylton, K., 2002, An asymmetric information model of litigation, *International Review of Law and Economics,* 153, 166.

IDEI report, 2006, *Objectives and Incentives at the European Patent Office*, Institut d'Economie Industrielle, Toulouse.

Kamien, M., 1992, Patent licensing, In: Aumann, R.J., S. Hart (Eds.) *Handbook of Game Theory with Economic Applications*, Elsevier Science, North Holland, ch. 11, 331-354.

Kamien, M., S. Oren and Y. Tauman, 1992, Optimal licensing of cost-reducing innovation, *Journal of Mathematical Economics*, 21, 483-508.

Kamien, M., and Y. Tauman, 1984, The private value of a patent: a game-theoretic analysis, *Z. Nationalőkon*, *Journal of Economics,* 4 (Supplement), 93-118.

Kamien, M. and Y. Tauman, 1986, Fees versus royalties and the private value of a patent, *Quarterly Journal of Economics*, 101, 471-491.

Katz, M. and C. Shapiro, 1985, On the licensing of innovations, *Rand Journal of Economics*, 16, 504-520.

Kimmel, S., 1992, Effects of cost changes on oligopolists' profits, *Journal of Industrial Economics*, 40, 4, 441-449.

Lemley, M. and C. Shapiro, 2005, Probabilistic patents, *Journal of Economic Perspectives*, 19, 2, 75-98.

Macho-Stadler, I., X. Martinez-Giralt and D. Perez-Castrillo, 1996, The role of information in licensing contract design, *Research Policy*, 25, 25-41.

Meurer, M., 1989, The settlement of patent litigation, *The Rand Journal of Economics*, 20, 1, 77-91.

Muto, S., 1993, On licencing policies in Bertrand competition, *Games and Economic Behavior*, 5, 257-267.

Poddar, S. and U. Sinha, 2002, On patent licensing and spatial competition, *Economic Record*, 80, 208-218.

Reinganum, J. and L. Wilde, 1986, Settlement, litigation and the allocation of litigation costs, *The Rand Journal of Economics*, 17, 4, 557-566.

Rostoker, M. 1984, A survey of corporate licensing, *IDEA: Journal of Law and Technology*, 24, 59-92.

Saracho, A., 2002, Patent licensing under strategic delegation, *Journal of Economics and Management Strategy*, 11, 225-251.

Sen, D., 2005, Fee versus royalty reconsidered, *Games and Economic Behavior*, 53, 141-147.

Sen, D. and Y. Tauman, 2007, General licensing schemes for a cost-reducing innovation, *Games and Economic Behavior*, 59, 163-186.

Scotchmer, S., 2004, *Innovation and Incentives*, The MIT Press, Cambridge, Ma.

Taylor, C. and Z. Silberstone, 1973, *The economic Impact of the Patent System*, Cambridge

University Press, Cambridge, UK.

2.7. Appendix

Proof of Lemma 1

Denote $h(\theta, r) = \pi(c - \epsilon + r, c - \epsilon + r) - (1 - \theta)\pi(c - \epsilon, c - \epsilon)$ and consider, for a given $\theta \in \left[0, \hat{\theta}\right]$, the equation $h(\theta, r) = 0$. Note that $h(\theta, 0) = \theta\pi(c - \epsilon, c - \epsilon) \geq 0$ and $h(\theta, \hat{r}) = \left(\theta - \hat{\theta}\right)\pi(c - \epsilon, c - \epsilon) \leq 0$ for any $\theta \in \left[0, \hat{\theta}\right]$. Since $h(\theta, .)$ is continuous and strictly decreasing over $[0, \hat{r}]$ (due to A1 and A4), we can use the intermediate value theorem to state that the equation $h(\theta, r) = 0$ has a unique solution in r, which we denote $r_2(\theta)$, over $[0, \hat{r}]$. Moreover, assumption A1 implies that $h(., .)$ is continuously differentiable over $[0, \hat{r}] \times \left[0, \hat{\theta}\right]$, which allows to state (using the implicit function theorem for instance) that $r_2(\theta)$ is differentiable over $\left[0, \hat{\theta}\right]$ and

$$r_2'(\theta) = \frac{-\pi(c - \epsilon, c - \epsilon)}{(\pi_1 + \pi_2)(c - \epsilon + r_2(\theta), c - \epsilon + r_2(\theta))}$$

This implies that $r_2'(\theta) > 0$ since $\pi_1 + \pi_2 < 0$ over $[c - \epsilon, c - \epsilon + \hat{r}] \times [c - \epsilon, c - \epsilon + \hat{r}]$ (by A4). Therefore $r_2(\theta)$ increases in the patent strength θ over $\left[0, \hat{\theta}\right]$. Furthermore, it is obvious that $r_2(0) = 0$ and we derive from the definition of $\hat{\theta}$ that $r_2(\hat{\theta}) = \hat{r}$.

Proof of Lemma 2

Denote $g(\theta, r) = \pi(c - \epsilon + r, c - \epsilon + r) - \theta\pi(c, c - \epsilon + r) - (1 - \theta)\pi(c - \epsilon, c - \epsilon)$ and consider, for a given $\theta \in \left[\hat{\theta}, 1\right]$, the equation $g(\theta, r) = 0$. Note that $g(\theta, \hat{r}) = \left(\theta - \hat{\theta}\right)\pi(c - \epsilon, c - \epsilon) \geq 0$ for any $\theta \in \left[\hat{\theta}, 1\right]$ and $g(\theta, \epsilon) = (1 - \theta)\left[\pi(c, c) - \pi(c - \epsilon, c - \epsilon)\right] \leq 0$ for any $\theta \in \left[\hat{\theta}, 1\right]$ (by A4). Moreover, the function $g(\theta, .)$ is continuous and strictly increasing over $[\hat{r}, \epsilon]$. Then, using the intermediate value theorem, we state that the equation $g(\theta, r) = 0$ has a unique solution in r for any $\theta \in \left[\hat{\theta}, 1\right]$, which we denote by $r_1(\theta)$. Furthermore, assumption A1 ensures that $h(.,.)$ is continuously differentiable over $[\hat{r}, \epsilon] \times \left[\hat{\theta}, 1\right]$, which allows to state that $r_1(\theta)$ is differentiable over $\left[\hat{\theta}, 1\right]$ and:

$$r_1'(\theta) = \frac{\pi(c, c-\epsilon+r_1(\theta)) - \pi(c-\epsilon, c-\epsilon)}{(\pi_1+\pi_2)(c-\epsilon+r_1(\theta), c-\epsilon+r_1(\theta)) - \theta\pi_2(c, c-\epsilon+r_1(\theta))} \tag{2.5}$$

The denominator is negative due to A3 and A4. The numerator is negative as well because $\pi(c, c-\epsilon+r_1(\theta)) \leq \pi(c-\epsilon+r_1(\theta), c-\epsilon+r_1(\theta)) < \pi(c-\epsilon, c-\epsilon)$. The first inequality follows from $r_1(\theta) \leq \epsilon$ and the second one from A4. Thus, $r_1'(\theta) > 0$, that is $r_1(\theta)$ is increasing in the patent strength θ over $\left[\hat{\theta}, 1\right]$. Furthermore, it is obvious that $r_1(\epsilon) = 1$ and we derive from the definition of $\hat{\theta}$ that $r_1(\hat{\theta}) = \hat{r}$.

Proof of Proposition 3

We distinguish two cases:

Case 1: $\theta \in \left[0, \hat{\theta}\right]$.

Consider a royalty rate $r \leq \hat{r}$. This royalty rate is accepted by all firms if and only if condition (2.2) holds, that is, $h(\theta, r) \geq 0$ where h has been defined in the proof of lemma 1. Since $h(\theta, r)$ is decreasing in r, the royalty rate r will be accepted by all firms if and only if $r \leq r_2(\theta)$ where $r_2(\theta)$ is defined in lemma 1.

Consider now $r > \hat{r}$. A necessary and sufficient condition for this royalty to be accepted by all firms is that condition (2.1) holds, that is, $g(\theta, r) \geq 0$ where g has been defined in the proof of lemma 2. Since $g(\theta, r)$ and $h(\theta, r)$ are decreasing in r and $r_1(\theta) \leq \hat{r}$ for $\theta \in \left[0, \hat{\theta}\right]$ then for any $r > \hat{r}$, it holds that $g(\theta, r) \leq g(\theta, \hat{r}) \leq g(\theta, \hat{r}) = h(\theta, \hat{r}) \leq h(\theta, r_1(\theta)) = 0$ which shows that r will not be accepted.

Hence, for any $\theta \in \left[0, \hat{\theta}\right]$, a royalty rate r will be accepted by all firms if and only if

$r \leq \min(\hat{r}, r_2(\theta)) = r_2(\theta)$.

Case 2: $\theta \in \left[\hat{\theta}, 1\right]$

Consider now a royalty rate $r \leq \hat{r}$. This royalty rate is accepted by all firms if and only if $h(\theta, r) \geq 0$. Since $g(\theta, r)$ and $h(\theta, r)$ are decreasing in r, it holds that $h(\theta, r) \geq h(\theta, \hat{r}) = g(\theta, \hat{r}) \geq g(\theta, r_2(\theta)) = 0$, which shows that r will be accepted by all firms.

Consider now a royalty rate $r > \hat{r}$. This royalty rate is accepted by all firms if and only if $g(\theta, r) \geq 0$. Since the function $g(\theta, r)$ is decreasing in r, the royalty rate r will be accepted by all firms if and only if $r \leq r_1(\theta)$ where $r_1(\theta)$ is defined in lemma 1.

Hence, for any $\theta \in \left[\hat{\theta}, 1\right]$, a royalty rate r will be accepted by all firms if and only if $r \leq \max(\hat{r}, r_1(\theta)) = r_1(\theta)$.

Proof of Lemma 4

We know from the proof of lemma 1 that for any $\theta \in \left[0, \hat{\theta}\right]$:

$$r_2'(\theta) = \frac{-\pi(c-\epsilon, c-\epsilon)}{(\pi_1 + \pi_2)(c-\epsilon+r_2(\theta), c-\epsilon+r_2(\theta))}$$

Since the numerator is negative, the derivative $r_2'(\theta)$ is increasing (or equivalently $r_2(\theta)$ is convex) if and only if $(\pi_1 + \pi_2)(c-\epsilon+r_2(\theta), c-\epsilon+r_2(\theta))$ is increasing in θ. Since $r_2(\theta)$ is increasing over $\left[0, \hat{\theta}\right]$, the latter condition is equivalent to $x \to (\pi_1 + \pi_2)(x, x)$ being increasing over $[c-\epsilon, c-\epsilon+\hat{r}]$. Since $x \to (\pi_1 + \pi_2)(x, x)$ is the derivative of $x \to \pi(x, x)$, we can state that $r_2(\theta)$ is convex over $\left[0, \hat{\theta}\right]$ if and only if $\pi(x, x)$ is convex in x over $[c-\epsilon, c-\epsilon+\hat{r}]$.

Proof of Lemma 5

The existence and unicity can be proven as in lemma 2. A difference with lemma 2 is that we do not need to restrict to $\theta \in \left[\hat{\theta}, 1\right]$ to get the differentiability property. Indeed as $r \to \pi(c, c-\epsilon+r)$ remains positive for any $r \geq 0$, it is differentiable over $[0,\epsilon]$. This ensures the differentiability of $g(\theta, r)$ (defined in the proof of lemma 2) over $[0,\epsilon]$ and allows to state that $r_1(\theta)$ is differentiable over $[0,1]$ and $r_1'(\theta)$ has the expression given by (2.5), which ensures the increasingness of $r_1(\theta)$. The equalities $r_1(0)=0$ and $r_1(\epsilon)=1$ are straightforward.

Proof of Proposition 6

Let $\theta \in [0,1]$. A royalty $r \geq 0$ is accepted by all firms if and only if $g(\theta, r) \geq 0$. Since this is a decreasing function in r and $g(\theta, r_1(\theta)) = 0$, then a royalty rate $r \geq 0$ is accepted by all firms if and only if $r \leq r_1(\theta)$.

Proof of Proposition 7

Differentiating the equation $\pi(c-\epsilon+r^c(\theta), c-\epsilon+r^c(\theta)) = \theta\pi(c,c) + (1-\theta)\pi(c-\epsilon, c-\epsilon)$ with respect to θ, we get $\frac{dr^c(\theta)}{d\theta} = \frac{\pi(c,c)-\pi(c-\epsilon,c-\epsilon)}{(\pi_1+\pi_2)(c-\epsilon+r^c(\theta),c-\epsilon+r^c(\theta))}$. Both the numerator and the denominator are negative which implies that $r^c(\theta)$ is increasing.

Since $\pi(c,c) - \pi(c-\epsilon, c-\epsilon) < 0$ (due to A4), the derivative $\frac{dr^c(\theta)}{d\theta}$ is increasing in θ over $[0,1]$ (i.e. $r^c(\theta)$ is convex) if and only if $(\pi_1+\pi_2)(c-\epsilon+r^c(\theta), c-\epsilon+r^c(\theta))$ is increasing in θ over $[0,1]$. Since $r^c(\theta)$ is continuous and strictly increasing from $r^c(0)=0$ to $r^c(1)=\epsilon$, the latter condition is equivalent to $(\pi_1+\pi_2)(x,x)$ is increasing in x over $[c-\epsilon, c]$, which means that $x \to \pi(x,x)$ is convex over $[c-\epsilon, c]$. In this case, $r^c(\theta) \geq \theta r^c(1) + (1-\theta) r^c(0) = \theta\epsilon$.

Proof of Proposition 8

Let us tackle first the case $\eta(\epsilon) \leq 1$.We know from the equivalence relation (2.3) that this leads to $r_2'(0) \geq \epsilon$. Since $r_2'(\theta)$ is (strictly) increasing in θ, the latter inequality leads to $r_2'(\theta) > \epsilon$ for any $\theta \in \left]0, \hat{\theta}\right]$. Taking the integral of both sides of the inequality and using the equality $r_2(0) = 0$ leads to $r_2(\theta) > \theta\epsilon$ for any $\theta \in \left]0, \hat{\theta}\right]$.

Let us turn now to the case $\eta(\epsilon) > 1$. In this case, $r_2'(0) < \epsilon$, which entails that $r_2(\theta) < \theta\epsilon$ for θ small enough.

Assume first that $\hat{r} > \hat{\theta}\epsilon$. Since $r_2(\theta)$ is continuous over $\left[0, \hat{\theta}\right]$, and $r_2(\theta) < \theta\epsilon$ for θ small enough while $r_2(\hat{\theta}) = \hat{r} > \hat{\theta}\epsilon$ then we can state that the equation $r_2(\theta) = \theta\epsilon$ has at least one solution over $\left[0, \hat{\theta}\right]$. Denote $\breve{\theta}$ the smallest solution and let us show that there are no other solution to the equation. Since $r_2(\theta)$ is convex and intersects $r_2(\theta) = \theta\epsilon$ from below, the derivative at the intersection point is greater than the slope of the straight line $\theta\epsilon$, that is, $r_2'(\breve{\theta}) > \epsilon$. Since $r_2'(\theta)$ is (strictly) increasing in θ, $r_2(\theta)$ will remain above $\theta\epsilon$ for $\theta > \breve{\theta}$ (we can show it by integrating both sides of $r_2'(\breve{\theta}) > \epsilon$ and using the equality $r_2(\breve{\theta}) = \breve{\theta}\epsilon$).

Assume now that $\hat{r} \leq \hat{\theta}\epsilon$. This means that $r_2(\hat{\theta}) < \hat{\theta}\epsilon$. We argue that under this assumption $r_2(\theta)$ will remain above $\theta\epsilon$ for any $\theta \in \left[0, \hat{\theta}\right]$. The reason is that in case this does not hold, the curve of $r_2(\theta)$ would intersect the straight line $\theta\epsilon$ at least twice. This is impossible because, as we have shown, after an intersection with this straight line, the curve of $r_2(\theta)$ remains above the line.

Proof of Proposition 9

We have: $\pi(c, c) \geq \pi(c, c - \epsilon + r^c(\theta))$. Since $\pi(c - \epsilon + r^c(\theta), c - \epsilon + r^c(\theta)) = \theta\pi(c, c) + (1 - \theta)\pi(c - \epsilon, c - \epsilon)$ we obtain that $\pi(c - \epsilon + r^c(\theta), c - \epsilon + r^c(\theta)) \geq \theta\pi(c, c - \epsilon + r^c(\theta)) + (1 - \theta)\pi(c - \epsilon, c - \epsilon)$. The latter inequality implies that a royalty rate $r = r^c(\theta)$ will be non cooperatively accepted by all firms if proposed by the patentholder. Therefore $r^c(\theta) \leq r(\theta)$.

Proof of Proposition 10

We first show that if $r < \epsilon$ it is impossible to have an equilibrium in which the number k of firms accepting the offer is strictly less than $n-1$. If this were true then one of the $n-k \geq 2$ firms that have not accepted the licensing contract could get a higher expected profit by deviating unilaterally and accepting the contract. Indeed, if it deviates then litigation will still occur because there will remain at least one firm refusing the license offer. This would result in the deviating firm having a marginal cost $c-\epsilon+r$ instead of c in case the patent is upheld, while still having a marginal cost equal to $c-\epsilon$ if the patent is invalidated by the court. Hence, the number of firms accepting the license offer $r<\epsilon$ at equilibrium is at least equal to $n-1$. This remains true for $r=\epsilon$ under the assumption that a firm accepts the offer when indifferent between accepting or refusing it. Furthermore, if $r\leq r(\theta)$, condition (2.1) shows that an equilibrium cannot involve $k=n-1$ licensees. Thus, i/ is proven.

If $r>r(\theta)$, an outcome in which one firm refuses the license offer while the others accept it is a Nash equilibrium: condition (2.1) shows that the firm refusing the offer gets a higher profit than if it had accepted it, and it has been shown that the remaining firms do not benefit from refusing the license since the patent will be challenged anyway. This proves ii/.

Part iii/ of the proposition is straightforward.

Proof of Lemma 11

Let $k\in\{n-1,n\}$. Since the function $rq\left(c-\varepsilon+r,k-1\right)$ is concave in r and reaches its maximum at $\tilde{r}_k\left(\epsilon\right)$ then it is increasing over $\left[0,\tilde{r}_k\left(\epsilon\right)\right]$. Consequently, the following holds:

$$\epsilon\leq\tilde{r}_k\left(\epsilon\right)\Longleftrightarrow\frac{\partial}{\partial r}\left(rq\left(c-\varepsilon+r,k\right)\right)|_{r=\epsilon}\geq 0\Longleftrightarrow q\left(c,k\right)+\epsilon\frac{\partial}{\partial r}\left(q\left(c-\varepsilon+r,k\right)\right)|_{r=\epsilon}\geq 0$$

Let us compare $\frac{\partial}{\partial r}\left(q\left(c-\varepsilon+r,n\right)\right)|_{r=\epsilon}$ and $\frac{\partial}{\partial r}\left(q\left(c-\varepsilon+r,n-1\right)\right)|_{r=\epsilon}$. It is clear that $q(c,n)=q(c,n-1)$: both expressions refer to the individual output of a firm in a symmetric oligopoly consisting of n firms producing at marginal cost c. Thus, using assumption A5, we get: $\frac{q(c-\varepsilon+r,n-1)-q(c,n-1)}{r-\epsilon}\leq\frac{q(c-\varepsilon+r,n)-q(c,n)}{r-\epsilon}$ for all $r<\epsilon$. Taking the limit of both sides as $r \to \epsilon$, we obtain $\frac{\partial}{\partial r}\left(q\left(c-\varepsilon+r,n\right)\right)|_{r=\epsilon} \geq \frac{\partial}{\partial r}\left(q\left(c-\varepsilon+r,n-1\right)\right)|_{r=\epsilon}$. Hence, $q\left(c,n\right)+\epsilon\frac{\partial}{\partial r}\left(q\left(c-\varepsilon+r,n\right)\right)|_{r=\epsilon} \geq q\left(c,n-1\right)+\epsilon\frac{\partial}{\partial r}\left(q\left(c-\varepsilon+r,n-1\right)\right)|_{r=\epsilon}$. Therefore, the following chain of implications holds:

$$\begin{aligned}\epsilon &\leq \tilde{r}_{n-1}\left(\epsilon\right) \Longrightarrow q\left(c,n-1\right)+\epsilon\frac{\partial}{\partial r}\left(q\left(c-\varepsilon+r,n-1\right)\right)|_{r=\epsilon} \geq 0\\ &\Longrightarrow q\left(c,n\right)+\epsilon\frac{\partial}{\partial r}\left(q\left(c-\varepsilon+r,n\right)\right)|_{r=\epsilon} \geq 0 \Longrightarrow \epsilon \leq \tilde{r}_{n}\left(\epsilon\right)\end{aligned}$$

Proof of Proposition 12

Assume that $\epsilon \leq \tilde{r}_{n-1}\left(\epsilon\right)$. By lemma 11, the inequality $\epsilon \leq \tilde{r}_{n}\left(\epsilon\right)$ holds as well. In this case the maximum of $P^{*}(r)$ over $[0,r(\theta)]$ is reached at $r(\theta)$, and its maximum over $]r(\theta),\epsilon]$ is reached at ϵ. Therefore, we must compare $nr(\theta)q(c-\epsilon+r(\theta),n)$ to $(n-1)\theta\epsilon q(c,n-1)$. Consider a royalty rate $r \in [0,\epsilon]$. The inequality $nrq(r,n) \geq (n-1)\theta\epsilon q(\epsilon,n)$ is fulfilled if and only if $\frac{rq(c-\epsilon+r,n)}{\epsilon q(c,n)} \geq \frac{n-1}{n}\theta$. Since the function $r \longrightarrow \frac{rq(c-\epsilon+r,n)}{\epsilon q(c,n)}$ is strictly increasing and continuous in r and takes the value 0 for $r=0$ and 1 for $r=\epsilon$, there exists a unique solution to the equation $\frac{rq(c-\epsilon+r,n)}{\epsilon q(c,n)} = \frac{n-1}{n}\theta$, which is denoted by $s\left(\theta\right)$. The condition $\frac{rq(c-\epsilon+r,n)}{\epsilon q(c,n)} \geq \frac{n-1}{n}\theta$ can then be written as $r \geq s\left(\theta\right)$. Hence the inequality $nr(\theta)q(c-\epsilon+r(\theta),n) \geq n-1)\theta\epsilon q(c,n)$ amounts to $r\left(\theta\right) \geq s\left(\theta\right)$. The convexity of $s\left(\theta\right)$ can be derived from the concavity of $w: r \longrightarrow rq(c-\epsilon+r,n)$ and its increasingness over $[0,\epsilon]$: differentiating twice the equation $\frac{w(s(\theta))}{w(\epsilon)} = \theta\frac{n-1}{n}$, we get $w''\left(s\left(\theta\right)\right)\left(s'\left(\theta\right)\right)^{2}+w'\left(s\left(\theta\right)\right)s''\left(\theta\right)=0$ which leads to $s''\left(\theta\right) = -\frac{w''(s(\theta))(s'(\theta))^{2}}{w'(s(\theta))} > 0$. The property $s(0)=0$ is immediate and the property $s(1)<\epsilon$ derives from $\frac{n-1}{n}<1$.

Proof of Proposition 13

Assume that $\tilde{r}_{n-1}\left(\epsilon\right) < \epsilon \leq \tilde{r}_{n}\left(\epsilon\right)$. In this case, the maximum of $P(r)$ over $[0,r(\theta)]$ is reached at $r\left(\theta\right)$. Define $\tilde{\theta}_{n-1}$ as the unique solution in θ to the equation $r\left(\theta\right)=\tilde{r}_{n-1}\left(\epsilon\right)$ (it is straightforward to check that such a solution exists in $[0,1]$ and is unique). Two subcases must be distinguished:

- Subcase 1: $\theta \leq \tilde{\theta}_{n-1}$: The maximum of $P(r)$ over $]r(\theta),\varepsilon]$ is then reached at $\tilde{r}_{n-1}(\epsilon)$. Determining the royalty rate that maximizes $P(r)$ over $[0,\epsilon]$ amounts then to the comparison of $nr(\theta)q(c-\epsilon+r(\theta),n)$ and $(n-1)\theta\tilde{r}_{n-1}(\epsilon)\,q(c-\epsilon+\tilde{r}_{n-1}(\epsilon),n-1)$. The former is greater than the latter if and only if $r(\theta)$ is greater than $v(\theta)$ defined as the unique solution in r to the equation $nrq(c-\epsilon+r,n)=(n-1)\theta\tilde{r}_{n-1}(\epsilon)\,q(c-\epsilon+\tilde{r}_{n-1}(\epsilon),n-1)$. The existence, uniquess, increasingness and convexity with respect to θ of such a solution can be established in a similar way to that of $s(\theta)$. The function $v(\theta)$ satisfies as well the properties $v(0)=0$ and $v(1)<\epsilon$. The first inequality is straightforward to show and the second one derives from $n\varepsilon q(c,n) > n\tilde{r}_{n-1}(\epsilon)\,q(c-\epsilon+\tilde{r}_{n-1}(\epsilon),n)$ (which holds because $\tilde{r}_{n-1}(\epsilon) < \epsilon \leq \tilde{r}_n(\epsilon)$) and $n\tilde{r}_{n-1}(\epsilon)\,q(c-\epsilon+\tilde{r}_{n-1}(\epsilon),n) > (n-1)\,\tilde{r}_{n-1}(\epsilon)\,q(c-\epsilon+\tilde{r}_{n-1}(\epsilon),n-1)$. Indeed these two inequalities result in $n\varepsilon q(c,n) > (n-1)\,\tilde{r}_{n-1}(\epsilon)\,q(c-\epsilon+\tilde{r}_{n-1}(\epsilon),n-1) = nv(1)\,q(c-\epsilon+v(1),n)$ and consequently lead to $\epsilon > v(1)$.

- Subcase 2: $\theta > \tilde{\theta}_{n-1}$: The upper bound of $P(r)$ over $]r(\theta),\varepsilon]$ is then reached at $r(\theta)^{+}$. From the expression of $P(r)$, it is clear that $P(r(\theta)) > P(r(\theta)^{+})$. Hence, the maximum of $P(r)$ over $[0,\epsilon]$ is reached at $r(\theta)$.Consequently litigation is always deterred in this subcase.

Proof of Proposition 14

Assume that $\epsilon > \tilde{r}_n(\epsilon)$. By lemma (6) the inequality $\epsilon > \tilde{r}_{n-1}(\epsilon)$ holds as well. Analogously to $\tilde{\theta}_{n-1}$, define $\tilde{\theta}_n$ as the unique solution in θ to the equation $r(\theta)=\tilde{r}_n(\epsilon)$. Three subcases are distinguished:

- Subcase 1: $\theta \leq \min\left(\tilde{\theta}_{n-1},\tilde{\theta}_n\right)$: The maximum of $P(r)$ over $[0,r(\theta)]$ is then reached at $r(\theta)$ and its maximum over $]r(\theta),\varepsilon]$ is reached at $\tilde{r}_{n-1}(\epsilon)$. Hence the analysis conducted in subcase 1 in the proof of proposition 13 applies here.

- Subcase 2 : $\tilde{\theta}_{n-1} < \theta < \tilde{\theta}_n$: The maximum of $P(r)$ over $[0, r(\theta)]$ is then reached at $r(\theta)$ and its maximum over $]r(\theta), \varepsilon]$ is reached at $r(\theta)^+$. Therefore the maximum of $P(r)$ over $[0, \epsilon]$ is reached at $r(\theta)$ (see subcase 2 in the proof of proposition 13) which implies that litigation is deterred. Note that this subcase is not relevant if the inequality $\tilde{\theta}_{n-1} < \tilde{\theta}_n$ does not hold.

- Subcase 3: $\theta \geq \tilde{\theta}_n$: The maximum of $P(r)$ over $[0, r(\theta)]$ is then reached at $\tilde{r}_n(\epsilon)$. This is sufficient to state that the maximum of $P(r)$ over $[0, \epsilon]$ is reached at $\tilde{r}_n(\epsilon)$ .This follows from the fact that the function $r \longrightarrow nrq(c - \epsilon + r, n)$ reaches its unconstrained maximum at $\tilde{r}_n(\epsilon)$ and $nrq(c - \epsilon + r, n) > \theta(n-1)rq(c - \epsilon + r, n-1)$ for any $r \in [0, \epsilon]$. The latter inequality results from assumption A6: $nq(c - \epsilon + r, n) = Q(c - \epsilon + r, n) \geq Q(c - \epsilon + r, n-1) \geq (n-1)q(c - \epsilon + r, n-1)$.

Proof of Corollary 15

Assume that $r(\theta) > \theta\epsilon$. Since $s(\theta)$ is a convex function such that $s(0) = 0$ and $s(1) < \epsilon$ then $s(\theta) \leq \theta\epsilon$ for all $\theta \in [0, 1]$. Consequently a sufficient condition for the inequality $r(\theta) \geq s(\theta)$ to hold is that $r(\theta) \geq \theta\epsilon$. The same conclusion applies for the convex function $v(\theta)$. Given this, the first part of the corollary follows immediately from the three previous propositions.

Using the three previous propositions, it is straighforward to check that under the conditions $r(\theta) \geq s(\theta)$ and $r(\theta) \geq v(\theta)$ (which hold when $r(\theta) > \theta\epsilon$), the optimal royalty rate set by the patentholder simplifies as follows: $r^*(\theta) = \min(r(\theta), \tilde{r}_n(\epsilon))$. Using the inequality $r(\theta) > \theta\epsilon$, we get $r^*(\theta) > \min(\theta\epsilon, \tilde{r}_n(\epsilon)) > \min(\theta\epsilon, \theta\tilde{r}_n(\epsilon)) = \theta \min(\epsilon, \tilde{r}_n(\epsilon)) = \theta r^*(1)$. Hence for all $\theta \in [0, 1]$, $\theta r^*(1) < r^*(\theta) = \min(r(\theta), \tilde{r}_n(\epsilon))$. Since the function $P(r) = nrq(c - \epsilon + r, n)$ is concave in r over $[0, r(\theta)]$ then $P(\theta r^*(1)) > \theta P(r^*(1))$ and since it reaches its maximum at $\tilde{r}_n(\epsilon)$, it is increasing over $[0, r^*(\theta)]$ which entails that $P(r^*(\theta)) > P(\theta r^*(1))$. From the two previous inequalities, we obtain that $P(r^*(\theta)) > \theta P(r^*(1))$.

Proof of Lemma 17

Let us show that a situation where only $k < n$ firms accept the contract and none of the remaining $n-k$ firms challenges the patent validity cannot be a Nash equilibrium of stage 2. If one of these firms challenges the patent validity it gets an expected profit of $\theta\pi^{NL}(k) + (1-\theta)\pi^L(n)$, whereas it gets a profit equal to $\pi^{NL}(k)$ if no firm challenges the patent validity. From A8 it follows that $\theta\pi^{NL}(k) + (1-\theta)\pi^L(n) > \pi^{NL}(k)$ which means that a downstream firm that does not accept the licensing contract is always better off challenging the patent validity.

Proof of Proposition 18

The situation where the n firms accept the licensing contract F is a Nash equilibrium if and only if:

$$\pi^L(n) - F \geq \theta\pi^{NL}(n-1) + (1-\theta)\,\pi^L(n)$$

which can be rewritten as:

$$F \leq \theta\left(\pi^L(n) - \pi^{NL}(n-1)\right) \tag{2.6}$$

that is

$$F \leq F_n(\theta)$$

A situation where $n-1$ firms accept the licensing contract and one firm does not is a Nash equilibrium (of stage 2) if and only if:

$$\theta\pi^{NL}(n-1)+(1-\theta)\,\pi^{L}(n)\geq\pi^{L}(n)-F \tag{2.7}$$

and

$$\theta[\pi^{L}(n-1)-F]+(1-\theta)\,\pi^{L}(n)\geq\theta\pi^{NL}(n-2)+(1-\theta)\,\pi^{L}(n) \tag{2.8}$$

Condition (2.7) means that the one firm that does not accept the licensing contract and challenges the patent validity does not find it optimal to unilaterally deviate by accepting the licensing contract. Condition (2.8) means that none of the $n-1$ firms which accept the licensing contract find it optimal to unilaterally deviate by refusing the contract. When the number of firms accepting the contract is strictly less than n, litigation will occur (lemma 17) which entails that the firms accepting the contract pay the fixed fee F only if the patent validity is upheld, which happens with probability θ. With the complementary probability $1-\theta$, the patent is invalidated and all the firms get the same profit namely $\pi^{L}(n)$. It is straightforward to show that conditions (2.7) and (2.8) are equivalent to the following double inequality:

$$\theta[\pi^{L}(n)-\pi^{NL}(n-1)]\leq F\leq\pi^{L}(n-1)-\pi^{NL}(n-2)$$

that is

$$F_{n}(\theta)\leq F\leq F_{n-1}$$

Note that the inequality $\theta[\pi^{L}(n)-\pi^{NL}(n-1)]<\pi^{L}(n-1)-\pi^{NL}(n-2)$ follows immediately from A9 for $\theta=1$ and is *a fortiori* satisfied for $\theta<1$.

A situation where $k\leq n-2$ firms accept the licensing contract and the remaining do not is a Nash equilibrium of the stage 2 subgame if and only if:

$$\theta\left(\pi^L(k) - F\right) + (1-\theta)\,\pi^L(n) \geq \theta\pi^{NL}(k-1) + (1-\theta)\,\pi^L(n) \tag{2.9}$$

and

$$\theta\pi^{NL}(k) + (1-\theta)\,\pi^L(n) \geq \theta\left(\pi^L(k+1) - F\right) + (1-\theta)\,\pi^L(n) \tag{2.10}$$

Condition (2.9) means that none of the k firms accepting the licensing contract finds it optimal to unilaterally deviate by refusing the contract and condition (2.10) means that none of the $n-k$ firms refusing the licensing contract finds it optimal to unilaterally deviate by accepting the contract. It is easy to see that conditions (2.9) and (2.10) can be combined into the following double inequality that does not depend on θ:

$$\pi^L(k+1) - \pi^{NL}(k) \leq F \leq \pi^L(k) - \pi^{NL}(k-1)$$

that is:

$$F_{k+1} \leq F \leq F_k$$

Note that the inequality $\pi^L(k+1) - \pi^{NL}(k) \leq \pi^L(k) - \pi^{NL}(k-1)$ follows from A9. Thus, the role of assumption A9 is to guarantee that the set of values of F belonging to the interval $[F_{k+1}, F_k]$ is not empty.

A situation where no firm accepts the licensing contract is a Nash equilibrium if and only if:

$$\theta\pi^{NL}(0) + (1-\theta)\,\pi^L(n) \geq \theta\left(\pi^L(1) - F\right) + (1-\theta)\,\pi^L(n)$$

which can be rewritten as:

$$\pi^{NL}(0) \geq \pi^{L}(1) - F$$

or equivalently as:

$$F \geq \pi^{L}(1) - \pi^{NL}(0) = F_1$$

Chapter 3

Leniency Programs for Multimarket Firms: The Effect of Amnesty Plus on Cartel Formation[1]

3.1. Introduction

Experience garnered over many years has taught antitrust authorities in the United States (US) and the European Union (EU) that companies which have been colluding in one specific product or geographic market are more likely to have engaged in cartel activities in other adjacent markets.

Due to the high diversity of businesses in multinational firms, cartel activities bear all the marks of contagion between and especially within companies. The probably most well-known example for such a cross-linked collusive pattern is the conspiracy in the markets for various vitamins. The striking feature of this complex of infringements was the central

[1] This chapter is based on a joint work with Catherine Roux.

role played by Hoffmann-la Roche (HLR) and BASF, the two main vitamin producers, over the course of ten years in virtually every cartel affecting the whole extent of bulk vitamin production.[2] HLR, BASF and Rhône-Poulenc instigated the first main group of cartels which consisted of price fixing agreements in the markets for vitamins A and E. The initial success of these arrangements inspired their replication in other vitamin markets. Smaller producers such as Merck, Takeda and Daiichi joined the pioneers and simultaneously colluded in a number of vitamin products. Accordingly, the European Commission (EC) stated that "the simultaneous existence of the collusive arrangements in the various vitamins was not a spontaneous or haphazard development, but was conceived and directed by the same persons at the most senior levels of the companies concerned".[3] Rhône-Poulenc's disclosure of evidence on collusion in the markets for vitamins A and E led to the opening of an investigation. However, only BASF's comprehensive collaboration with the US Department of Justice (DoJ) under the Amnesty Plus Program accelerated inquiries and finally led to the successful prosecution of all participants. When Rhône-Poulenc plead guilty to its vitamin conspiracies under the US Amnesty Program and applied for leniency under the 1996 EC Leniency Notice, it did not provide any information on its participation in the vitamin D3 infringement and even pursued cartel activities in other product markets such as methionine and methylglucamine.[4]

In the US, convictions of global cartels in the 1990s suggest that at least a dozen firms were repeat offenders in related product industries (Connor, 2003). The DoJ has been investigating around 50 alleged international cartels in 2004, and half of them have been detected during inquiries on other markets (Hammond, 2004). With the objective

[2]Concerned were the markets for vitamins A, E, B1, B2, B5, B6, folic acid, C, D3, H, beta carotene and carotinoids.

[3]EC IP/01/1625 November 2001.

[4]EC IP/01/1625 November 2001, EC IP/02/976 July 2002, EC IP/02/1746 November 2002.

of fully exploiting the multimarket contact between colluding firms, the DoJ implemented the Amnesty Plus Program in 1999 as part of its Corporate Leniency Policy. According to Hammond, "The Division's Amnesty Plus program creates an attractive inducement for encouraging companies who are *already under investigation* to report the full extent of their antitrust crimes [...]" (Hammond, 2004, p.16).

Leniency programs reduce fines for cartel members that bring evidence to the antitrust authority. Amnesty refers to the complete exemption from fines. Amnesty Plus aims at attracting amnesty applications by encouraging subjects of ongoing investigations to consider whether they qualify for amnesty in other than the currently inspected markets where they engage in cartel activities. In particular, Amnesty Plus offers a firm, which currently plea-bargains an agreement for participation in one cartel, where it cannot obtain guaranteed amnesty, complete immunity in a second cartel affecting another market. Provided that the firm agrees to fully cooperate in the investigation of the conspiracy of which the DoJ was previously not aware, it is automatically granted amnesty for this second offense. Moreover, the company benefits from a substantial additional discount[5], i.e. the Plus, in the calculation of its fine in any plea agreement for the initial matter under investigation.[6]

Under the current EC Leniency Notice, Amnesty Plus does not exist. Although, in 2001, the Organization for Economic Co-operation and Development (OECD) recommended the inclusion of Amnesty Plus as part of the 2002 reforms of the EU Leniency Program, the EC did not seize the opportunity to follow the US example by introducing a similar policy.

[5]The size of the additional discount mainly depends on three factors: The strength of the evidence provided by the cooperating company, the potential significance of the revealed case measured in terms of volume of commerce involved, geographic scope and the number of co-conspirators, and the likelihood that the DoJ would have detected the cartel absent self-reporting (Hammond, 2006)

[6]As a counterpart of Amnesty Plus, the DoJ contemporaneously implemented the Penalty Plus Program. Penalty Plus increases the fines for companies that neglect to take advantage of Amnesty Plus but are nevertheless caught for a second time. The main reason why we do not include Penalty Plus in our analysis is that, whereas we want to focus on the difference between the US and the EU leniency policy, the clause of "aggravating circumstances" in the "2006 EC Guidelines on the method of setting fines" is very similar to the US Penalty Plus.

The present paper studies whether and how the Amnesty Plus policy affects firms' incentives to form a cartel. It seems intuitive that, following a conviction of one cartel, Amnesty Plus encourages firms to report another cartel by granting the first firm which applies for this program a substantial discount in the fine already imposed. However, we argue that, Amnesty Plus may have important consequences for cartel formation in particular because it increases the firms' incentives to report a cartel after a first detection through fine reductions.

We study two markets in which two identical firms play an infinitely repeated game of collusion. In each period, the firms can choose to form a cartel before interacting on the product market. Collusion generates incriminating evidence which the antitrust authority can discover with some probability. In addition, each firm can also bring this evidence to the authority. When a cartel is detected, either through an investigation or a firm's self-reporting, each cartel member, except the first reporting firm, has to pay a fine. Amnesty Plus becomes relevant when the firms have to decide whether to report a cartel in one market after they have already been convicted in the other market.

Our main result is that Amnesty Plus may affect cartel formation in two different ways: On the one hand, Amnesty Plus may have a pro-competitive effect by dissuading the firms to create one of their cartels when they would have formed both of them in the absence of Amnesty Plus. On the other hand, Amnesty Plus may also have an anticompetitive effect as it may encourage the firms to form both cartels when they would have formed only one of them under an antitrust policy without Amnesty Plus.

We also examine whether the firms can exploit their multimarket contact by linking punishment strategies across markets. Without Amnesty Plus, the firms can always treat the markets in isolation and thus, they use multimarket trigger strategies only if this facilitates collusion. Amnesty Plus however inherently links the markets. Moreover, it is the antitrust authority which decides on the implementation of Amnesty Plus, and the firms can only try to weaken its effectiveness by adapting their strategies. In particular, we find that if the markets do not differ substantially in terms of profitability, the use of multimarket strategies, while it does not directly affect the firms' ability to collude, lowers the pro-competitive effect of Amnesty Plus and increases its anticompetitive effect.

Surprisingly, although legal studies which mainly argue in favor of an Amnesty Plus policy in Europe are burgeoning, the existing literature contains virtually no formal economic analysis which attemps to clarify possible motives for the EC's non-adoption of Amnesty Plus, let alone to study the potential impact of such a policy on cartel formation. We take the first step towards filling this gap in the economic theory on leniency programs.

Recent academic research such as Harrington (2008), Chen and Rey (2007), Aubert et al.(2006), Spagnolo (2004) and Motta and Polo (2003) has elaborated on the differences in conception of leniency programs and their impact on the effectiveness of antitrust enforcement.[7] This line of research mainly highlights the basic trade-offs between destabilizing collusion and deterring cartel formation and explores whether and, if so, under which conditions, leniency programs do not deter but rather encourage the formation of a cartel. The results are embedded in a normative analysis of how the antitrust authority should design such programs to minimize their undesirable effect. Our analysis is close in purpose to this literature in that we examine how Amnesty Plus as a feature of leniency programs affects cartel stability. However, unlike in previous work, where the firms collude in one market only, we allow them to simultaneously participate in two collusive agreements.

Some studies suggest that leniency programs which not only reduce fines but offer a

[7] For an extensive overview of the economic literature on leniency programs see Spagnolo (2006).

positive reward to whistleblowing firms can deter collusion in a more effective way. In particular, Aubert et al.(2006) find the minimal reward necessary to induce a firm to report collusion and point out that this reward may be quite large. Spagnolo (2004) shows that an optimally designed leniency program rewards the first reporting company with an amount equal to the sum of the fines paid by its former partners. On this issue, economic theory however conflicts with legal practice. Although granting positive rewards may strengthen the deterrence power of leniency programs, remunerating antitrust offenders not only raises moral concerns but may also increase the risk of negative effects in that it may further lower the expected penalty level. Hence, antitrust authorities mostly refrain from rewarding informants.[8] However, it may be argued that Amnesty Plus is equivalent to granting more than 100% leniency because it not only waives the entire penalty in the second cartel but also gives a fine discount for the initial infringement (Wils, 2007). From this perspective, the justification for an Amnesty Plus policy does not seem obvious.

Another strand of literature studies the role of multimarket contact between firms in sustaining collusion when monitoring is perfect. In their seminal paper, Bernheim and Whinston (1990) build on the idea, first raised by Edwards (1955) and further developed in a finite oligopoly games context by Harrington (1987), that multimarket contact across firms may foster anticompetitive outcomes. As a benchmark, they establish an irrelevance result: with identical firms and markets and constant returns to scale technology, multimarket contact does not affect the opportunities for cooperation. However, they also identify various plausible circumstances in which strategically linking markets facilitates collusion by slackening the incentive constraints that limit the firms' ability to sustain collusive be-

[8]Note however that for other forms of multiagent crime, like government fraud, the US False Claim Act substantially rewards the cooperation of individual informants. Moreover, Korea has introduced a reward scheme for reporting parties in antitrust cases.

havior in settings of repeated interactions. Spagnolo (1999) refutes this irrelevance result and shows that, if the firms' static objective functions are strictly concave, multimarket contact always makes collusion viable in a set of markets even if, in its absence, it could not be sustained in any of these markets.

Relatively few papers examine the effect on collusive behavior of the interaction between multimarket contact and imperfect information. Thomas and Willig (2006) find that exploiting multimarket contact by strategically linking markets may be unprofitable. This surprising result occurs because strategic linkage may promote contagion which allows adverse shocks to spread from one market to another. Matsushima (2001) shows that efficient collusion can be achieved in the limit through the linkage of a sufficiently large number of identical markets.

We look at the interaction between Amnesty Plus and multimarket contact and its effect on the firms' ability to collude when information is perfect. Our analysis allows to distinguish the anticompetitive effect of strategic linkage from the one of Amnesty Plus. Whereas multimarket strategies are chosen by the firms and used only if they are pro-collusive, Amnesty Plus is implemented by the antitrust authority. Hence, it links markets both when it facilitates but also when it hinders collusion.

The remainder of the paper is organized as follows. Section 3.2 sets up the model. Sections 3.3 and 3.4 analyze cartel formation when firms use standard trigger strategies both under a European antitrust policy without Amnesty Plus and under a US antitrust policy with Amnesty Plus. In section 3.5, we graphically present our main findings. In sections 3.6 and 3.7, we allow firms to use multimarket punishment. Section 8 briefly concludes. All proofs can be found in Appendix A. In Appendix B, we discuss how the relaxation of two important assumptions affects our results.

3.2. The model

3.2.1. Set-up

We consider two markets $k = 1, 2$ in which two identical firms play an infinitely repeated game where, in each period, they can choose to form a cartel before interacting on the product market. Communication is necessary for collusion and generates hard evidence which makes it possible to establish the antitrust offense. Markets 1 and 2 may differ in profitability. In particular, market 1 is at least as profitable as market 2. Firms discount future payoffs by a common discount factor $\delta \in [0, 1]$. We compare the firms' cartel formation decisions under the EC Leniency Program and the US Amnesty Program whose sole difference here is that the latter allows for Amnesty Plus.[9] Amnesty Plus signifies that a firm which has been caught colluding in one market can get a discount in the fine already imposed by reporting the remaining cartel in the other market.

The collusive joint profit is $2\pi_k > 0$, and thus, each firm makes a cartel profit equal to π_k. Denote by $\lambda = \frac{\pi_2}{\pi_1} \in]0, 1]$ the profit ratio of the two cartels. If firms compete, they make zero profits. In case one firm unilaterally deviates from the cartel while the other continues to collude, the deviating firm earns the whole short-term cartel profit $2\pi_k$ alone, whereas the other firm gets nothing. Firms use (grim) trigger strategies. The punishment firms agreed upon starts the period following the deviation and lasts forever after.

At the time firms decide whether to enter an illegal agreement, they observe the strictness of the enforcement policy that is summarized by a conviction probability $q \in]0, 1]$ with which the Antitrust Authority (AA) opens an investigation in one market leading to the

[9] More specifically, we examine the effect of Amnesty Plus on the best collusive subgame-perfect equilibrium that can be sustained through trigger strategies.

conviction of the colluding firms with certainty.[10] Detection across markets is independent. Each convicted firm pays a market specific fine F_k which is reduced under Amnesty Plus to $F_k - R_k$ in return for the disclosure of the second cartel. $R_k \in]0, F_k]$ represents the fine reduction granted to the first informant. The higher R_k the more generous the Amnesty Plus policy. The firm which is eligible for Amnesty Plus is the first company which reports the second infringement and thus, it also receives amnesty in that market. If both firms simultaneously apply for Amnesty Plus, each is first with probability $\frac{1}{2}$.

To keep the analysis simple, we assume that the evidence of collusion lasts for only one period. Thus, even after a firm has deviated from a collusive agreement it is held liable for its cartel behavior and can be fined until the end of the period in which the deviation occurred.[11] Each cartel member has the possibility to bring the incriminating evidence to the AA. The first informant is eligible for total immunity from fines under a standard Amnesty Program. In our model, the only strategic implication of this standard Amnesty Program is that, since a defecting firm must still fear conviction, a unilateral deviation is always immediately followed by the reporting of the cartel.

In practice, fines are set according to judicial principles but are often related, directly or indirectly, to the nature and importance of the anticompetitive behavior, and thus, to the profits from collusion. We assume that the AA sets the fine as a function of the per period collusive profits[12], $F_k = F(\pi_k)$ where $F(\cdot)$ is an increasing function. Let then $\theta_k = \frac{F_k}{\pi_k}$

[10] To keep the model simple, we identify investigation and conviction with a single probability. However, we could introduce uncertainty with respect to the AA's ability to prove guilty a detected cartel by substituting qs for q where s is the probability with which the investigation succeeds. See Chen and Rey (2007) for an analysis of optimal leniency rates before any and once an investigation is opened which distinguishes the probability of launching the investigation from the probability with which it succeeds.

[11] The limitation period of the liability for antitrust offenses is generally a positive number of years. Article 25 of the EC Council Regulation 1/2003 states that the Commission can sue for Administrative Action until five years from the date of the infringement. Moreover, "[...]in the case of continuing or repeated infringements, time shall begin to run on the day on which the infringement ceases".

[12] Since the evidence that incriminates a cartel lasts only for one period, the assumption that the AA takes the collusive profit *per period* and not cumulated over the whole duration of the cartel as a basis for the determination of the fine seems plausible.

denote the fine-profit ratio for market $k = 1, 2$ and suppose that $\theta_2 \geq \theta_1$. This reflects the idea that the fine rises proportionally or less than proportionally with the cartel profit, i.e. that $\theta_k = \frac{F_k}{\pi_k}$ is weakly decreasing in π_k.[13] Fine records tend to support this assumption.[14]

Following a cartel conviction, we assume that the AA closely monitors the previously collusive industry and thus, firms compete and never return back to collusion in the same market.

3.2.2. Timing

The time structure of the game is as follows:

- $t = 0$:

  Stage 0: Both firms decide in each market whether to enter a collusive agreement. If at least one firm decides not to collude in market k, competition takes place in this market. If this happens in both markets, the game ends for that period. If both firms choose to collude in market k, their communication leaves some hard evidence.

  Stage 1: Each firm decides whether to deviate or not from the collusive agreement(s). Its rival does not observe this decision until the end of stage 2.

  Stage 2: Each firm decides whether to report the evidence to the AA. The AA detects the cartel with probability 1 if at least one firm self-reports. The first informant gets complete immunity from fines in this market, whereas the other firm has to pay the

[13]E.g., let $F(\cdot)$ be an increasing concave function or let F_k be an affine transformation of π_k: $F(\pi_k) = a + b\pi_k$ with $a \geq 0$ and $b \geq 0$. In Appendix B1, we discuss how the relaxation of the assumption $\theta_2 \geq \theta_1$ affects our analysis.

[14]E.g. Vitamin cartel: In the US, a fine equal to 127% of the collusive overcharge was imposed in the market of vitamin B2 whereas it ranged between 63% and 88% of the collusive overcharge in the more profitable vitamin C market. In the EU, the fines ranged between 63% and 88% for vitamin B2 and between 30% and 60% for vitamin C (Connor, 2005).

full fine. If each cartel formed in stage 0 is reported in this stage, the game ends for this period; otherwise:

Stage 3: Each cartel formed in stage 0 and not reported in stage 2 is detected with probability q. If the AA does not detect the cartel(s) formed in stage 0, the game ends for that period. If the AA however detects the cartel(s) formed in stage 0, the colluding firms pay the corresponding fines, and the game ends for that period. If the firms have formed both cartels in stage 0, and the AA has detected only one of them, then:

Stage 4: Each firm chooses whether to report the remaining cartel.

- $t \geq 1$: If both cartels have been formed but none of them has been convicted (detected or reported) in the previous period, the same time structure applies to this period. If either no cartel has been formed or the cartel(s) formed has (have) been convicted, the firms compete in both markets.[15] If either one cartel has been formed and not convicted or both cartels have been formed but only one cartel has been convicted, then:

  Stage 0: Each firm decides whether to enter a collusive agreement in the market where the cartel has gone undetected in the previous periods. If at least one firm chooses not to collude, competition takes place in this market, and the game ends for this period.

  Stage 1: Each firm decides whether to deviate from the collusive agreement. Its rival does not observe this decision until the end of stage 2.

[15]We assume here that the firms can only form a cartel if this cartel has already been formed in the previous period. In Appendix B2, we relax this assumption and discuss the strategy where the firms form cartel 1 until it is detected and then form cartel 2.

Stage 2: Each firm decides whether to report the evidence to the AA.

Stage 3: The AA detects the cartel in market k with probability q. If the cartel is not detected the game ends for this period. If it is detected, the colluding firms pay the corresponding fines.

3.2.3. Individual stability of a cartel

We first examine under what conditions an individual cartel is sustainable if the firms interact in only one market k. The firms can try to sustain repeated collusion by using trigger strategies in which they would return to competition the period following the deviation in case one of them deviates from the collusive outcome. In the presence of a standard amnesty policy where the first self-reporting firm pays no fine, deviating and reporting weakly dominates deviating and not reporting. A unilateral deviation is therefore always followed by an amnesty application. Hence, in the period of the deviation, the informant earns the whole cartel profit and pays no fine whereas thereafter, he gets zero profits from competition. Collusion is sustainable if the present discounted value V_k of the cartel is at least as big as the gain each firm gets from a unilateral deviation in this market, that is

$$V_k \geq 2\pi_k$$

V_k is the continuation value of the cartel in market k and is such that:

$$V_k = q(\pi_k - F_k) + (1-q)(\pi_k + \delta V_k)$$

Solving for V_k yields

$$V_k = \frac{\pi_k - qF_k}{1 - \delta(1-q)}$$

The above condition defines an **individual stability threshold** such that cartel k is individually stable if and only if

$$\delta \geq \frac{\pi_k + qF_k}{2\pi_k(1-q)} = \frac{1 + q\theta_k}{2(1-q)} \equiv \widetilde{\delta}(q, \theta_k)$$

A firm has no incentive to unilaterally deviate from the collusive equilibrium if $\delta \geq \widetilde{\delta}(q, \theta_k)$. Both firms anticipate that collusion will be sustainable and thus, they form the cartel. However, if $\delta < \widetilde{\delta}(q, \theta_k)$, the firms anticipate that, immediately after the formation of the cartel, they would both deviate and self-report. Hence, they do not form the cartel in the first place. The individual stability threshold is increasing and continuous in all its arguments. Intuitively, the higher the probability of conviction and the higher the fine-profit ratio, the more firms have to value future flows of collusive profits, and thus, the higher the δ needed to individually sustain the cartel. Note that $\widetilde{\delta}(q, \theta_k) \leq 1$ if and only if $\theta_k \leq \frac{1}{q} - 2$. Otherwise, the cartel k is individually unstable for any value $\delta \in [0, 1]$. Finally, the assumption $\theta_2 \geq \theta_1$ implies that $\widetilde{\delta}(q, \theta_1) \leq \widetilde{\delta}(q, \theta_2)$, i.e. a cartel in market 2 is equally or more difficult to sustain than a cartel in the more profitable market 1.

3.3. EC Leniency Program with standard trigger strategies

Suppose now that the firms encounter each other in the *two* markets 1 and 2, and that they use standard trigger strategies to sustain collusion. Each firm plays the collusive equilibrium in market $k = 1, 2$ as long as the partner colludes in this market. If a firm unilaterally deviates from the collusive agreement, the other firm competes from the next period on and forever after in this market. Note that a deviation in one market triggers punishment only in this specific market. Intuitively, since punishment strategies as well as detection probabilities across markets are independent, firms treat each market in isolation, and their actions in market 1 do not influence their decisions in market 2. The condition under which a cartel is formed is therefore the same as the condition under which the cartel is individually stable. We provide a formal proof of this intuitive argument and state the following proposition:

Proposition 1 *Under the EU antitrust policy, a cartel is formed if and only if it is individually stable. More precisely,*

i/ If $\delta < \widetilde{\delta}(q, \theta_1)$, *no cartel is formed.*

ii/ If $\widetilde{\delta}(q, \theta_1) \leq \delta < \widetilde{\delta}(q, \theta_2)$, *cartel* 1 *is formed whereas cartel* 2 *is not.*

iii/ If $\widetilde{\delta}(q, \theta_2) \leq \delta$, *both cartels are formed.*

Proof. See Appendix A. ■

In what follows, we assume that $\theta_2 \leq \frac{1}{q} - 2$. Without this assumption, the region in case iii/ where the firms form both cartels would be empty.

3.4. US Amnesty Program with standard trigger strategies

We now introduce an Amnesty Plus policy which allows a firm, already caught in one cartel, to benefit from a fine reduction if it is the first to report the other cartel. It has been heavily advertised that the main benefit of Amnesty Plus is its effect on cartel *desistance*: Amnesty Plus increases the firms' incentives to report a cartel after the detection of another cartel. However, we argue that, above all, Amnesty Plus may have important consequences for cartel *deterrence* in particular because it encourages reporting after a first detection through fine reductions. To see these effects, we proceed by backward induction. Suppose that the firms have formed both cartels, and the AA has detected one of them, say cartel $-k$. The remaining cartel k survives this detection only if none of the firms unveils the collusive evidence to the AA at the end of this period. The firms do not report cartel k if and only if two conditions jointly hold. First, the fine reduction a firm gets in return for the disclosure of cartel k must not exceed the discounted value of this cartel. That is

$$\delta V_k \geq R_{-k}$$

This condition defines a **robustness threshold** such that cartel k is *robust* to the detection of cartel $-k$ if and only if

$$\delta \geq \frac{\frac{R_{-k}}{\pi_k}}{1 - q\theta_k + (1-q)\frac{R_{-k}}{\pi_k}} \equiv \widehat{\delta}\left(q, \theta_k, \frac{R_{-k}}{\pi_k}\right)$$

Note that the robustness threshold is increasing and continuous in all its arguments. In particular, it increases with the fine reduction R_{-k}. The more generous the Amnesty Plus policy, the higher the robustness threshold, and the more the firms find the reporting of cartel k attractive. Hence, Amnesty Plus encourages firms to report a cartel once another cartel has been detected. Second, the firms will have to again form cartel k at the beginning of the next period. This will happen if and only if the individual stability condition holds. Hence, cartel k survives the detection of cartel $-k$ if and only if it is robust and individually stable, that is

$$V_k \geq \max\left(2\pi_k, \frac{R_{-k}}{\delta}\right)$$

If this inequality does not hold, the firms report cartel k. In particular, if cartel k is individually unstable, the firms anticipate that they cannot form this cartel next period. Reporting cartel k is then a dominant strategy for each firm. If a firm anticipates that its partner does not report cartel k, it gets a strictly positive fine reduction from reporting instead of zero from not reporting. Moreover, if a firm anticipates that its co-conspirator reports, it also prefers reporting since it gets Amnesty Plus with probability $\frac{1}{2}$ and avoids paying a fine with certainty. If cartel k is individually stable but not robust, the Nash equilibrium in dominant strategies is to report this cartel. Again, if a firm anticipates that its partner reports cartel k, it also prefers to report. However, even if a firm anticipates that its partner does not report and thus that they may form the cartel again next period, it prefers to report because the fine reduction is higher than the present discounted value the firm would get from sustaining this cartel.

Let us compare the individual stability to the robustness threshold in market k. We find that

$$\widetilde{\delta}(q, \theta_k) \geq \widehat{\delta}\left(q, \theta_k, \frac{R_{-k}}{\pi_k}\right) \Longleftrightarrow R_{-k} \leq \frac{\pi_k + qF_k}{1-q} \tag{3.1}$$

Intuitively, if the fine reduction R_{-k} is rather small, Amnesty Plus cannot induce the reporting of the cartel, and it is the individual stability and not the robustness condition which is stringent. Hence, cartel k survives the detection of cartel $-k$ if and only if $\delta \geq \widetilde{\delta}(q, \theta_k)$. However, if the fine reduction is large enough, the firms want to benefit from Amnesty Plus and therefore they report the cartel. In this case, it is the robustness condition which is stringent, and cartel k survives the detection of cartel $-k$ if and only if $\delta \geq \widehat{\delta}\left(q, \theta_k, \frac{R_{-k}}{\pi_k}\right)$.

Since we can now determine the outcome in the last stage of the game after a possible detection in one of the markets, we examine the firms' decisions in the cartel formation stage. The firms create both cartels only if they do not find the optimal unilateral deviation profitable. Hence, the joint formation of the cartels is a Nash Equilibrium if the expected present discounted value each firm gets when forming both cartels is weakly higher than the payoff from the optimal unilateral deviation, that is

$$V_{12} \geq \text{payoff from the optimal unilateral deviation}$$

This inequality defines a **joint stability condition** such that the firms form both cartels if and only if this condition is satisfied. Note that the right hand side (RHS) of the above condition does not depend on the fine reduction under Amnesty Plus since, after a unilateral deviation, at most one cartel is left, and the Amnesty Plus option is not available. The left hand side (LHS), however, is (weakly) increasing piecewise in R_1 and R_2 and thus, the joint stability condition becomes less stringent piecewise when the fine discounts increase. It is important to understand, that the effects of Amnesty Plus on cartel desistance and on cartel deterrence may go in opposite directions. On the one hand, Amnesty Plus strengthens the firms' incentives to report a cartel once the other cartel has been detected. On the other hand, it may increase the expected present discounted value of the joint cartel profits as the expected fine in case of a conviction decreases. However, note that V_{12} is discontinuous in the fine discounts. If R_1 and R_2 increase up to the point where reporting under Amnesty Plus gets so attractive that an individually stable cartel breaks down after the detection of the other cartel, V_{12} decreases drastically. However, if the fine discounts then continue to increase, V_{12} rises again. This scenario may recur if the second cartel is also individually stable and breaks down after the detection of the other cartel as the firms' reporting incentives increase. The net effect of Amnesty Plus therefore depends on the strength of its effect, first, on the firms' reporting incentives and, second, on V_{12}. It may thus be either pro- or anti-competitive. We now examine this net effect in each of the three possible constellations of cartel formation under the EU antitrust policy.

3.4.1. No cartel formed under the EU policy: a neutrality result

Proposition 2 *If both cartels are individually unstable, i.e.* $\delta < \widetilde{\delta}(q, \theta_1)$*, Amnesty Plus is neutral: the firms form no cartel both under the EU and the US policy.*

Proof. See Appendix A. ■

The intuition for the proof is as follows: Proceeding by backward induction, we take the creation of both cartels as given and examine the Nash Equilibrium after the AA has detected one of the cartels. Since both cartels are individually unstable, each firm has a dominant strategy in reporting the remaining cartel after the first detection. Thus, the only Nash equilibrium in the remaining market is the firms' simultaneous denunciation of the cartel. Each firm may be first to apply for Amnesty Plus, with a 50% chance. Examining the joint stability condition, we find that, for any possible value of R_{-k}, the expected present discounted value each firm gets when forming both cartels is always smaller than the optimal unilateral deviation which takes place in both markets. This signifies that if no cartel is individually stable under the EU Leniency Program, an ever so generous Amnesty

Plus policy cannot have any stabilizing effect. As a consequence, Amnesty Plus is neutral in this case.

3.4.2. One cartel formed under the EU policy: a potential anticompetitive effect

In Lemma 3 we give the expression for the joint stability threshold which we then use to formulate Proposition 4.

Lemma 3 *If cartel* 1 *survives the detection of cartel* 2*, but cartel* 2 *does not survive the detection of cartel* 1*, i.e.* $\max\left(\widetilde{\delta}(q,\theta_1),\widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)\right) \leq \delta < \max\left(\widetilde{\delta}(q,\theta_2),\widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right)\right)$, *the two cartels are jointly stable if and only if*

$$\delta \geq \frac{1}{2(1-q)^2} - \frac{1}{4(1-q)^2}\theta_2\left(\frac{R_1}{F_2}q(1-q) - q(3-q)\right) \equiv \breve{\delta}\left(q,\theta_2,\frac{R_1}{F_2}\right)$$

Proof. See Appendix A. ■

Proposition 4 *If cartel* 1 *is individually stable whereas cartel* 2 *is not, i.e.* $\widetilde{\delta}(q,\theta_1) \leq \delta < \widetilde{\delta}(q,\theta_2)$*, Amnesty Plus has an anticompetitive effect on cartel formation if and only if*

$$\max\left(\widetilde{\delta}(q,\theta_1),\breve{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right)\right) \leq \delta < \widetilde{\delta}(q,\theta_2)$$

This condition defines a non-empty range of values of δ *if and only if*

$$R_2 < \frac{(1+q\theta_2)(\pi_1 - qF_1)}{(1-q)(1-q\theta_2)} \qquad \textit{and} \qquad R_1 > \frac{1+q}{1-q}F_2 + \frac{2\pi_2}{1-q}$$

Amnesty Plus then encourages the firms to form the individually unstable cartel 2 *which* *they would not have formed under the EU policy.*

Proof. See Appendix A. ■

Proposition 4 suggests that Amnesty Plus may have an anticompetitive effect by stabilizing a cartel which would not have been sustainable under the EU Leniency Program in the presence of another cartel which is individually stable. Note that for this anticompetitive effect to potentially occur, the firms must form cartel 1 but not cartel 2 under the EU policy for a non-empty range of discount factor values. For this to be the case, the individual stability thresholds for markets 1 and 2 must differ. However, if $\theta_1 = \theta_2$, that is, if the fines are proportional to the collusive profits or the markets are perfectly symmetric, these thresholds are identical, and Amnesty Plus cannot have any anticompetitive effect.

We sketch the proof of Proposition 4 as follows: If the individually stable cartel 1 is detected, reporting the individually unstable cartel 2 is a dominant strategy for each firm. Hence, the firms report cartel 2 and may save part of the fine already imposed. Amnesty Plus thus decreases a firm's expected fine from a cartel conviction. The firms form both cartels if the payoff from the optimal unilateral deviation, which occurs in market 2 only, does not exceed the value of the joint creation of the cartels. Two conditions have to hold. First, cartel 1 must be not only individually stable but also robust to a detection of cartel 2. Hence, the robustness condition must hold for a non-empty range of values for δ within the interval $[\widetilde{\delta}(q,\theta_1), \widetilde{\delta}(q,\theta_2)[$. This is true if the robustness threshold for cartel 1 lies below the individual stability threshold for cartel 2. We can show that

$$\widehat{\delta}\left(q, \theta_1, \frac{R_2}{\pi_1}\right) < \widetilde{\delta}(q,\theta_2) \Longleftrightarrow R_2 < \frac{(1+q\theta_2)(\pi_1 - qF_1)}{(1-q)(1-q\theta_2)} \tag{3.2}$$

Second, the cartels must be jointly stable. For the joint stability condition to hold for a non-empty range of values for δ within the interval $[\widetilde{\delta}(q,\theta_1),\widetilde{\delta}(q,\theta_2)[$, it is necessary and sufficient that the joint stability threshold lies below the individual stability threshold for cartel 2. We find that

$$\breve{\delta}\left(q,\theta_2,\frac{R_1}{F_2}\right) < \widetilde{\delta}(q,\theta_2) \Longleftrightarrow R_1 > \frac{1+q}{1-q}F_2 + \frac{2\pi_2}{1-q} \tag{3.3}$$

Combining conditions (3.1) and (3.2), three situations are possible:

i/ $R_2 \leq \frac{\pi_1+qF_1}{1-q}$

If the fine discount in market 2 is low enough, cartel 1 survives the detection of cartel 2 for all $\delta \in [\widetilde{\delta}(q,\theta_1),\widetilde{\delta}(q,\theta_2)[$. If inequality (3.3) is satisfied, and the joint stability condition holds, the firms form both cartels whereas, in the absence of Amnesty Plus, they would have formed cartel 1 alone. Otherwise, the firms collude only in market 1, and Amnesty Plus is neutral.

ii/ $\frac{\pi_1+qF_1}{1-q} < R_2 < \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$

The formation of the two cartels is the best collusive equilibrium if and only if both cartel 1 survives the detection of cartel 2 and the cartels are jointly stable. In this case, Amnesty Plus has an anticompetitive effect. However, if the robustness condition for cartel 1 is not satisfied, the firms cannot sustain the remaining cartel after one cartel detection, and the joint stability condition never holds. As a consequence, the firms form only cartel 1.

iii/ $R_2 \geq \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$

In this case, the fine discount under Amnesty Plus in market 2 is too high such that the robustness condition for cartel 1 cannot be satisfied. Hence, no cartel survives the detection of the other cartel, and the joint stability condition does not hold. The firms form only the individually stable cartel 1, and Amnesty Plus is neutral.

Corollary 5 *Amnesty Plus has no anticompetitive effect on cartel formation if the fine discount a firm gets under Amnesty Plus for cartel* $-k$ *does not exceed the fine for the reported cartel* k*, i.e.* $R_{-k} \leq F_k$.

Proof. Note that if $R_1 \leq F_2$ the condition $R_1 > \frac{1+q}{1-q}F_2 + \frac{2\pi_2}{1-q}$ does not hold, and the joint stability condition cannot be satisfied for δ within the interval $[\widetilde{\delta}(q,\theta_1), \widetilde{\delta}(q,\theta_2)[$. It follows from Proposition 4 that Amnesty Plus cannot have any anti-competitive effect in this case. Finally, $R_2 \leq F_1$ is always true since $F_1 \geq F_2 \geq R_2$. ■

Corollary 5 suggests that, as a simple rule to avoid any stabilizing effect of the Amnesty Plus policy, the size of the fine discount granted in one market should not exceed the fine a non successful Amnesty Plus applicant would have incurred in the other market. This rule is sufficient but not necessary. Intuitively, if $R_{-k} \leq F_k$, each firm gets a negative expected payoff from reporting cartel k after the detection of cartel $-k$. This is because both firms report cartel k simultaneously and thus, with probability $\frac{1}{2}$, each firm has to pay a fine which is higher than the possible fine discount. The expected present discounted value each firm gets from the joint formation of the cartels decreases and cannot exceed the payoff from the optimal unilateral deviation. Hence, if cartel 2 is individually unstable such that the firms report it after the detection of cartel 1, the two cartels can never be jointly stable. By keeping the fine reduction low, the AA therefore can avoid any anticompetitive effect of the Amnesty Plus policy.[16]

[16] Consider the interaction of $n \geq 2$ firms on markets 1 and 2. We suppose that, if all the firms report a cartel simultaneously, each firm is first with probability $\frac{1}{n}$. As the first firm reporting is the only company eligible for the fine discount under Amnesty Plus, a firm's expected payoff from reporting cartel k when

3.4.3. *Both cartels formed under the EU policy: a potential pro-competitive effect*

Proposition 6 *If both cartels are individually stable, i.e. $\delta \geq \widetilde{\delta}(q, \theta_2)$, Amnesty Plus has a pro-competitive effect on cartel formation if and only if at least one of the cartels is not robust and the two cartels are not jointly stable. In particular:*

i/ If $R_2 \leq \frac{(1+q\theta_2)(\pi_1 - qF_1)}{(1-q)(1-q\theta_2)}$, Amnesty Plus has a pro-competitive effect for a non-empty range of values of δ if and only if $\frac{\pi_2 + qF_2}{1-q} < R_1 < \frac{2\pi_2 + (1+q)F_2}{1-q}$.

ii/ If $R_2 > \frac{(1+q\theta_2)(\pi_1 - qF_1)}{(1-q)(1-q\theta_2)}$, Amnesty Plus has a pro-competitive effect for a non-empty range of values of δ for any value of $R_1 \in]0, F_1]$.

Proof. See Appendix A. ■

Proposition 6 suggests that Amnesty Plus may have a pro-competitive effect by destabilizing a cartel which would have been sustainable under the EU policy. The sketch of the proof is as follows: Amnesty Plus decreases the expected present discounted value of profits each firm gets when forming both cartels if, following the detection of one individually stable cartel, the firms report the other individually stable cartel to benefit from the fine discount. V_{12} may then fall below the payoff from the optimal unilateral deviation which occurs in market 2 only. As a consequence, the firms would anticipate that the cartels are not jointly stable and form only the more profitable of the cartels. To examine the exact circumstances under which the firms form both cartels, we need to find the expected discounted value each firm gets from the formation of both cartels and compare it to the payoff from the optimal deviation. Since both cartels are individually stable, each cartel

everyone else reports is $\frac{1}{n}R_{-k} - \frac{n-1}{n}F_k$. We have $\frac{1}{n}[R_{-k} - (n-1)F_k] \leq 0$ if and only if $R_{-k} \leq (n-1)F_k$. It is then straightforward that if $R_{-k} \leq (n-1)F_k$ holds for $n = 2$, it holds a fortiori for $n > 2$.

survives the detection of the other cartel if and only if the robustness condition holds. Combining conditions (3.1) and (3.2), four possible situations arise:

i/ $R_1 \leq \frac{\pi_2+qF_2}{1-q}$ and $R_2 \leq \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$

The individual stability condition in market 2 is more stringent than the robustness conditions for both cartels. Hence, each cartel survives the detection of the other cartel for all $\delta \geq \widetilde{\delta}(q,\theta_2)$. Not surprisingly, the expected present discounted value each firm gets from forming both cartels turns out to be always weakly greater than the payoff from the optimal unilateral deviation. As a consequence, the firms form both cartels. Amnesty Plus is neutral because the firms form also both cartels under the EU policy.

ii/ $R_1 > \frac{\pi_2+qF_2}{1-q}$ and $R_2 \leq \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$

The individual stability condition in market 1 is more stringent than the robustness condition in this market. Cartel 1 therefore always survives the detection of cartel 2. Cartel 2 however survives the detection of cartel 1 only if it is robust. If the robustness condition for cartel 2 is satisfied, the analysis is the same as in i/. Hence, the payoff from the optimal unilateral deviation does not exceed the expected present discounted profits from the joint creation of the cartels. The firms form both cartels, and Amnesty Plus is neutral. However, if cartel 2 is not robust, the firms form both cartels if and only if the joint stability condition holds. Amnesty Plus has a pro-competitive effect in the case where this condition is not satisfied. The firms form only cartel 1 in the US whereas they would have formed both of them in the EU.

iii/ $R_1 \leq \frac{\pi_2+qF_2}{1-q}$ and $R_2 > \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$

Cartel 2 is individually stable and robust and therefore always survives the detection of cartel 1. Cartel 1 however survives the detection of cartel 2 only if it is robust. If the robustness condition for cartel 1 holds, then the analysis is the same as in i/. The firms do not find the optimal unilateral deviation profitable and create both cartels. If cartel 1 is not robust, the firms form both cartels if and only if the joint stability condition holds. Otherwise, they form only cartel 1, and Amnesty Plus has a pro-competitive effect.

iv/ $R_1 > \frac{\pi_2+qF_2}{1-q}$ and $R_2 > \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$

Each cartel survives the detection of the other cartel only if it is robust. If both cartels are robust, the firms form both cartels, and Amnesty Plus is neutral. If either one or none of the robustness conditions holds, the firms form both cartels if and only if the joint stability condition holds. In particular, if none of the cartels is robust, the joint stability condition anyway holds for a non-empty set of discount factor values if the conviction probability q is small enough. Otherwise, the firms form only cartel 1, and Amnesty Plus has a pro-competitive effect.

3.5. Discussion

We illustrate our main findings from sections 3.4.1, 3.4.2 and 3.4.3 by means of Figures 3.1 to 3.4. Note in particular that Figures 3.1 and 3.2 depict the results only for the case where $F_1 > \frac{2\pi_2+(1+q)F_2}{1-q}$. In Figure 3.1, we show the net effect of Amnesty Plus on cartel formation as a function of the fine discount R_1 for a given $R_2 \leq \frac{\pi_1+qF_1}{1-q}$ such that cartel 1, whenever it is individually stable, always survives the detection of cartel 2. Amnesty Plus is neutral for all values of δ if R_1 is sufficiently small, i.e. $R_1 \leq \frac{\pi_2+qF_2}{1-q}$. Amnesty Plus has a pro-competitive effect on cartel formation for intermediate values of R_1, i.e.

$\frac{\pi_2+qF_2}{1-q} < R_1 < \frac{2\pi_2+(1+q)F_2}{1-q}$, and for values of δ such that both cartels are individually but not jointly stable and such that cartel 2 is not robust to a detection of cartel 1. This region is labeled with a "+". The firms form only cartel 1 in the US whereas they would have formed both cartels in the EU. As a measure of the size of the effect, we use the width of the relevant interval of values for δ on the y-axis. Hence, we can say that the pro-competitive effect increases between $\frac{\pi_2+qF_2}{1-q}$ and R_1^* where it finally reaches its maximum. Beyond R_1^*, this effect decreases and goes to zero as R_1 increases to $\frac{2\pi_2+(1+q)F_2}{1-q}$. R_1^* is determined by the intersection of the robustness threshold of cartel 2 and the joint stability threshold when cartel 1 is robust which both do not depend on R_2. As a consequence, the maximum size of the pro-competitive effect of Amnesty Plus which is the difference $\widehat{\delta}\left(q, \theta_2, \frac{R_1^*}{\pi_2}\right) - \widetilde{\delta}(q, \theta_2)$ does not involve R_2 neither. In the region labeled with a "-", Amnesty Plus has an anticompetitive effect on cartel formation. This effect occurs for higher values of R_1, i.e. $R_1 > \frac{2\pi_2+(1+q)F_2}{1-q}$, and for values of δ such that cartel 1 is individually stable and robust whereas cartel 2 is not, and the two cartels are jointly stable. The firms form both cartels in the US whereas, in the absence of Amnesty Plus, they would have formed cartel 1 alone. The size of the anticompetitive effect increases as R_1 increases from $\frac{2\pi_2+(1+q)F_2}{1-q}$ to F_1.

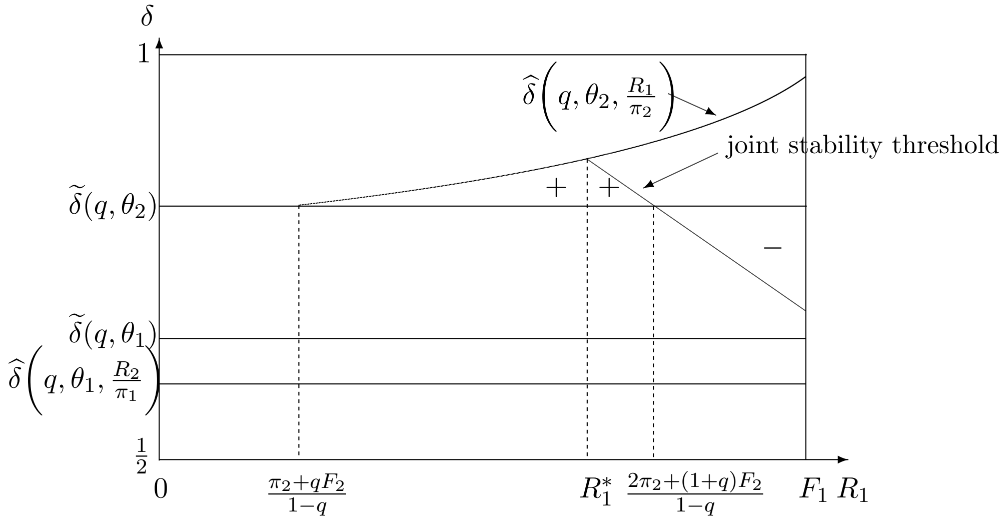

Figure 3.1: Effect of Amnesty Plus if $R_2 \leq \frac{\pi_1+qF_1}{1-q}$

Figure 3.2 depicts the net effect of Amnesty Plus for a given R_2 such that $\frac{\pi_1+qF_1}{1-q} < R_2 < \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$. The only difference with respect to Figure 3.1 is that the robustness threshold for cartel 1 is now above its individual stability threshold. The region where Amnesty Plus has an anticompetitive effect may thus be truncated at the level of the robustness threshold for cartel 1. Hence, the potential anticompetitive effect of Amnesty Plus may be more limited while its potential pro-competitive effect remains the same.

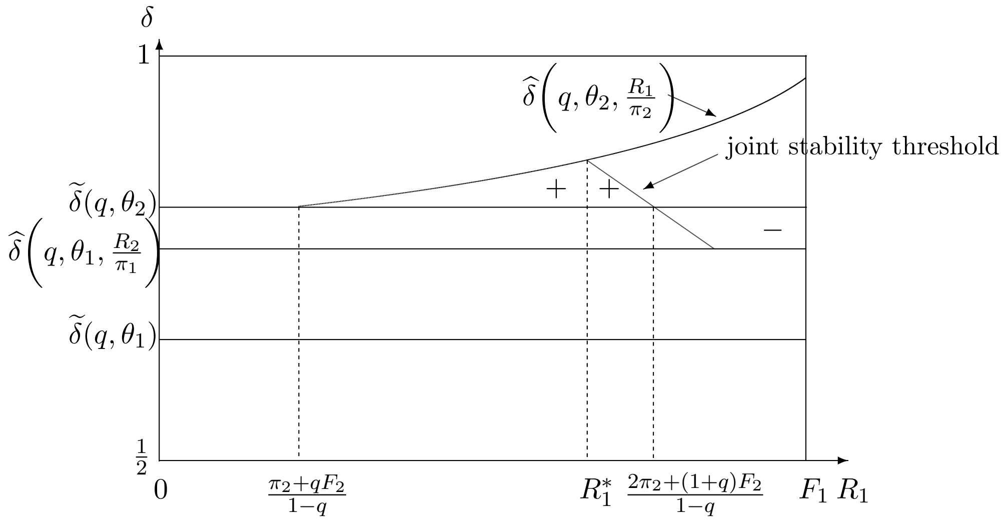

Figure 3.2: Effect of Amnesty Plus if $\frac{\pi_1+qF_1}{1-q} < R_2 < \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$

In Figures 3.3 and 3.4, we show the net effect of Amnesty Plus for a given $R_2 \geq \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$. From Proposition 4 we know that Amnesty Plus cannot have any anti-competitive effect in this case. Moreover, note that, in contrast to Figures 3.1 and 3.2, Amnesty Plus has a pro-competitive effect on cartel formation for a non-empty range of values of δ for any value of $R_1 > 0$. In Figure 3.3, the conviction probability q is very small. In this case, the highest discount factor value for which the pro-competitive effect occurs is close to the individual stability threshold of cartel 2. Hence, the size of the potential pro-competitive effect is rather small. Note in particular that Amnesty Plus cannot have any pro-competitive effect if cartel 1 is robust to a detection of cartel 2. The pro-competitive effect only occurs if cartel 1 is not robust, and both cartels are individually but not jointly stable. The interval of discount factor values where these conditions jointly hold is never empty. The value of R_1^* for which the pro-competitive effect is maximal corresponds to the intersection of the robustness threshold for cartel 2 and the joint stability threshold when both cartels are not robust. Note that, as the latter depends on both R_1 and R_2, R_1^* depends here on R_2.

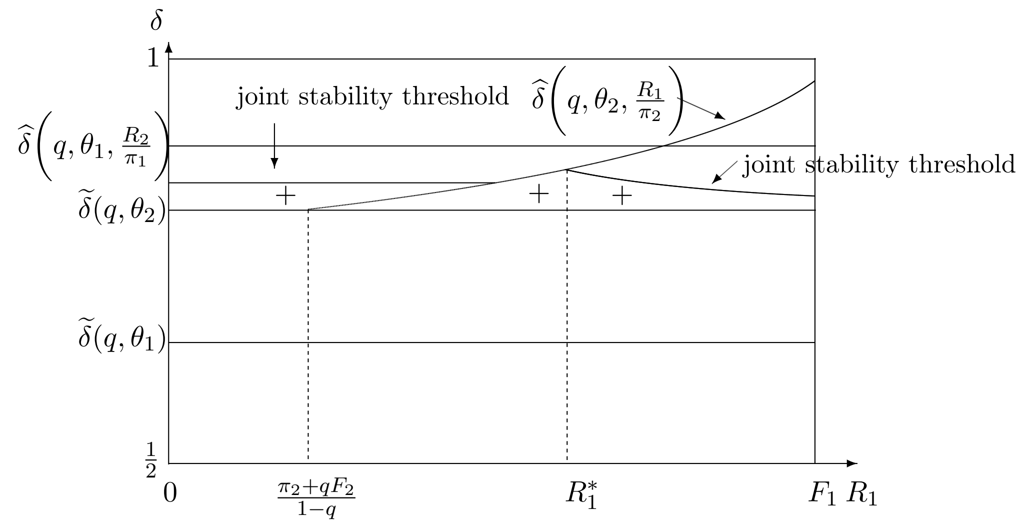

Figure 3.3: Effect of Amnesty Plus if $R_2 \geq \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$ and q is very small

Figure 3.4 displays the effect of Amnesty Plus with q not “too small”. There are two main differences with respect to Figure 3.3. First, the region where the pro-competitive effect of Amnesty Plus occurs is larger. Second, the pro-competitive effect may appear even if cartel 1 is robust to a detection of cartel 2. If this happens, the value of R_1^* is at the intersection of the robustness threshold for cartel 2 and the joint stability threshold when cartel 1 is robust. R_1^* does therefore not depend on R_2.

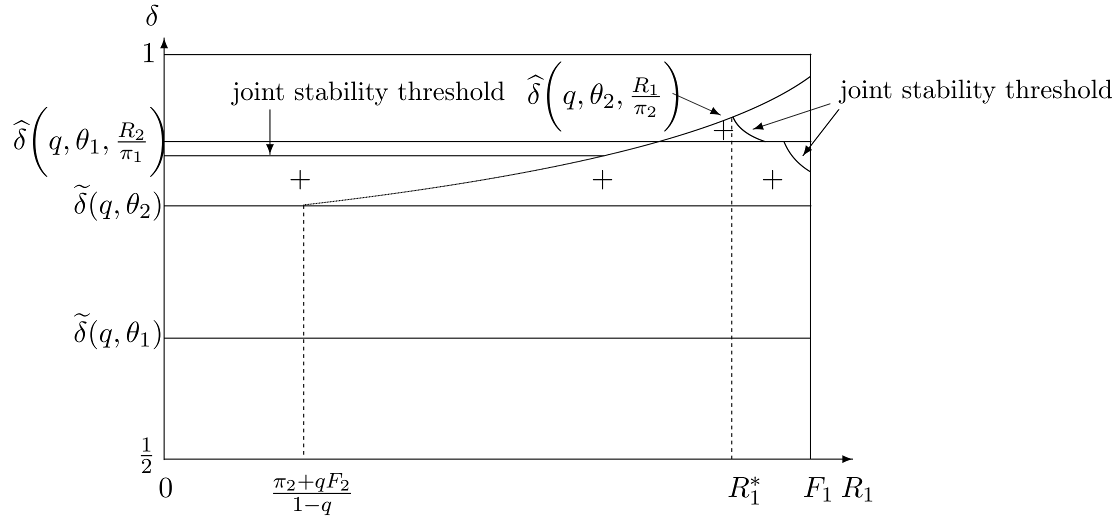

Figure 3.4: Effect of Amnesty Plus if $R_2 \geq \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$ and q is not too small

3.6. EC Leniency Program with multimarket trigger strategies

Suppose now that the firms try to sustain repeated collusion by consciously exploiting their multimarket contact and using multimarket trigger strategies. Each firm cooperates in market $k = 1, 2$ as long as its partner does. If one firm unilaterally deviates from the illegal agreement in one of the markets, the co-conspirator reacts with a reversion to competition in both markets. As a deviation in one market triggers punishment not only in the market where the deviation occurred but also in the market where the collusive agreement has been respected, the optimal deviation always takes place in both markets simultaneously. By linking the punishment across markets, firms can potentially transfer slack enforcement power from market 1 to market 2 and sustain collusion in both markets for values of δ for which only cartel 1 would have been sustainable under single-market contact. We state the formal argument in the following proposition:[17]

Proposition 7 *There exists a threshold $\widetilde{\pi}_2(q) < \pi_1$ such that:*

a- If $\pi_2 < \widetilde{\pi}_2(q)$, i.e. the asymmetry between markets 1 and 2 is sufficiently strong, the use of multimarket trigger strategies enhances the firms' ability to collude relative to the use of standard trigger strategies. In other words, there exists a threshold $\bar{\delta}(q, \theta_1, \theta_2, \lambda) \in]\widetilde{\delta}(q, \theta_1), \widetilde{\delta}(q, \theta_2)[$ such that:

i/ If $\delta < \widetilde{\delta}(q, \theta_1)$, no cartel is formed.

ii/ If $\widetilde{\delta}(q, \theta_1) \leq \delta < \bar{\delta}(q, \theta_1, \theta_2, \lambda)$, cartel 1 is formed whereas cartel 2 is not.

iii/ If $\bar{\delta}(q, \theta_1, \theta_2, \lambda) \leq \delta$, both cartels are formed.

b- If $\widetilde{\pi}_2(q) \leq \pi_2 \leq \pi_1$, i.e. the asymmetry between markets 1 and 2 is sufficiently weak, the use of multimarket strategies does not affect the firms' ability to collude relative to the use of standard trigger strategies.

Proof. See Appendix A. ■

In Bernheim and Whinston (1990), multimarket contact does not affect collusive behavior if the markets are identical (provided that firms are identical and there are constant

[17] In what follows we suppose that $\theta_1 < \theta_2$. If $\theta_1 = \theta_2$, multimarket trigger strategies cannot have any effect on the firms' ability to collude.

returns to scale). In our model, multimarket contact turns out to be irrelevant not only for identical markets but also for markets which are not too different in terms of profitability. This finding can be interpreted as a somewhat broader version of the irrelevance result of Bernheim and Whinston (1990), and it has a straightforward explanation. In the presence of an antitrust policy, the firms' may use multimarket trigger strategies to sustain cooperation on both markets just as long as the AA does not detect one of the cartels. However, if the AA detects one of the cartels, the firms cannot use multimarket punishment thereafter, and they can sustain the remaining cartel in subsequent periods only if this cartel is individually stable. Antitrust enforcement may thus shorten the time period during which the firms use multimarket trigger strategies and therefore, it may limit the effect these strategies can have on the ease to sustain cooperation on both markets.

Nonetheless, if markets 1 and 2 are sufficiently asymmetric, the use of multimarket trigger strategies *does* strengthen the firms' ability to collude. In particular, the firms form both cartels for a larger range of discount factor values.

3.7. US Amnesty Program with multimarket trigger strategies

Let us now examine whether the firms may influence the effectiveness of the Amnesty Plus policy by using multimarket trigger strategies. To do this, we need to know how the use of multimarket strategies affects the individual stability, robustness and joint stability conditions.

Consider first the individual stability and the robustness conditions: After the detection of a cartel in one market, the firms interact only in the one remaining market. Since, without multimarket contact, the firms cannot link punishment across markets, they have to use standard trigger strategies to sustain the remaining cartel. The individual stability and the robustness conditions as well as the resulting thresholds therefore are the same as in section 3.4.

Second, consider the joint stability condition: The use of multimarket trigger strategies may alter the optimal unilateral deviation and thereby affect the joint stability condition. This is because the optimal unilateral deviation occurs always in the two markets with multimarket trigger strategies whereas, with standard trigger strategies, a firm may find it optimal to deviate in one market only. More precisely, if cartel 1 is individually unstable, i.e. $\delta < \widetilde{\delta}(q, \theta_1)$, we have shown that the optimal unilateral deviation under both strategies is to deviate in both markets. Hence, the use of multimarket strategies does not affect the joint stability condition and thereby the neutrality of Amnesty Plus. However, if cartel 1 is individually sustainable, i.e. $\delta \geq \widetilde{\delta}(q, \theta_1)$, we have shown that the optimal unilateral deviation occurs only in market 2 when the firms use standard trigger strategies but takes place in both markets when they use multimarket trigger strategies. The joint stability condition $V_{12} \geq V_1 + 2\pi_2$ when trigger strategies are standard becomes $V_{12} \geq 2\pi_1 + 2\pi_2$ when strategies link markets. Since $V_1 \geq 2\pi_1$ whenever $\delta \geq \widetilde{\delta}(q, \theta_1)$, the use of multimarket strategies makes the joint stability condition less stringent. Hence, the firms form both cartels for a larger range of discount factor values if they use multimarket rather than standard trigger strategies. In particular, for values of δ such that $2\pi_1 + 2\pi_2 \leq V_{12} < V_1 + 2\pi_2$ the firms form both cartels when they use multimarket strategies whereas they would have formed only cartel 1 with standard trigger strategies.

From Proposition 7 we know that, if the markets 1 and 2 are sufficiently similar in terms of profitability, multimarket trigger strategies do not affect the set of discount factor values for which the firms create only one, respectively, two cartels, under a policy without Amnesty Plus. However, since the use of multimarket trigger strategies may lower the joint stability threshold, the firms may anyway want to use these strategies to strengthen the anticompetitive effect and weaken the pro-competitive effect of Amnesty Plus when it is included in the leniency policy. If the asymmetry between the markets is strong enough, the use of multimarket strategies enlarges the region of discount factor values for which the firms form both cartels. Hence, the pro-competitive effect of Amnesty Plus may occur for a larger range of discount factor values. At the same time, however, the use of multimarket strategies makes the joint stability condition less stringent and thereby the anticompetitive effect of Amnesty Plus more likely to occur. Similarly, the set of discount factor values for which the firms form only cartel 1 and for which Amnesty Plus may have an anticompetitive effect shrinks with the use of multimarket trigger strategies. However, the joint stability condition becomes less stringent, and the occurrence of the anticompetitive effect more likely. As a consequence, if the markets differ sufficiently in terms of profitability, the net effect of the multimarket trigger strategies is ambiguous.

3.8. Conclusion

This paper examines the effect of the Amnesty Plus policy on the firms' incentives to engage in cartel activities. We develop an infinitely repeated interaction framework to highlight the mechanism through which Amnesty Plus encourages, discourages or has no effect on cartel formation when firms use standard and multimarket trigger strategies. US success stories suggest that Amnesty Plus weakens cartel stability. Our analysis shows that this intuition is not always correct.

We find that Amnesty Plus may have an anticompetitive effect by stabilizing a cartel which is individually unstable in the presence of another cartel which is individually stable. If the latter cartel is detected, the firms report the former in the hope of a discount in the fine already imposed. Hence, Amnesty Plus decreases a firm's expected fine from a cartel conviction such that, for each firm, the joint creation of the cartels may result in an expected discounted value of profits larger than the payoff from the optimal unilateral deviation. The firms would anticipate that the cartels are jointly stable and form both cartels whereas only one of them would have been created under the EU Leniency Program.

Our results also show that, Amnesty Plus may have a pro-competitive effect by destabilizing a cartel which is individually stable. Amnesty Plus decreases the expected present discounted value of profits each firm gets when forming both cartels if, following the detection of one individually stable cartel, the firms report the other individually stable cartel to benefit from the fine discount. The value of the joint creation may then fall below the payoff from the optimal unilateral deviation. The firms anticipate that the cartels are not jointly stable and form only the more profitable of the cartels whereas they would have formed both cartels under the EU policy.

We have also examined whether the firms can exploit their multimarket contact by linking punishment across markets. Amnesty Plus is implemented by the antitrust authority and inherently links the markets. We find that if the markets do not differ substantially in terms of profitability, the use of multimarket trigger strategies can partly offset the destabilizing effect of Amnesty Plus whereas it does not directly affect the firms' ability to collude. Firms may thus want to adopt multimarket trigger strategies even if their use does not directly facilitate collusion.

Our findings suggest that an antitrust policy with Amnesty Plus may help to increase cartel deterrence insofar as its potential anticompetitive effect could be avoided. We have shown that, by setting the size of the fine discount granted in one market such that it does not exceed the fine a non successful Amnesty Plus applicant would have incurred in the other market, the antitrust authority can avoid any stabilizing effect of this policy. In view of this result, we believe that, for future research, it might be particularly fruitful to elaborate on a thorough normative analysis of how an optimal Amnesty Plus policy should be designed.

3.9. References

Aubert, C., Kovacic, W., and P. Rey, 2006, "The impact of leniency and whistle-blowing programs on cartels", *International Journal of Industrial Organization*, 24, 1241-1266.

Bernheim, B.D. and M.D. Whinston, 1990, "Multimarket contact and collusive behavior", *Rand Journal of Economics*, 21, 1-26.

Chen, Z. and P. Rey, 2007, "On the design of leniency programs", IDEI working papers 452.

Connor, J.M., 2003, "International price fixing: Resurgence and deterrence", Working paper, Purdue University.

Edwards, C.D., 1955, "Conglomerate bigness as a source of power" in *Business Concentration and Price Policy*, NBER Conference Report, Princeton University Press.

Hammond, S.D., 2004, "Cornerstones of an effective leniency program", presented at the ICN Workshop on Leniency Programs, Sydney.

Hammond, S.D., 2006, "Measuring the value of second-in cooperation in corporate plea negociations", presented at the 54th Annual American Bar Association Section of Antitrust Law Spring Meeting.

Harrington, J.E., 1987, "Collusion in multiproduct oligopoly games under a finite horizon", *International Economic Review*, 28, 1-14.

Harrington, J.E., 2008, "Optimal corporate leniency programs", *Journal of Industrial Economics*, 56, 215-246.

Matsushima, H., 2001, "Multimarket contact, imperfect monitoring, and implicit collusion", *Journal of Economic Theory*, 98, 158-178.

Motta, M. and M. Polo, 2003, "Leniency programs and cartel prosecution", *International Journal of Industrial Organization*, 21, 347-379.

Spagnolo, G., 1999, "On interdependent supergames: Multimarket contact, concavity and collusion", *Journal of Economic Theory*, 89, 127-139.

Spagnolo, G., 2006, " Leniency and whistleblowers in antitrust", CEPR Discussion Papers 5794.

Thomas, C. J. and R.D. Willig, 2006, "The risk of contagion from multimarket contact", *International Journal of Industrial Organization*, 24, 1157-1184.

Wils, W.J., 2007, "Leniency in antitrust enforcement: Theory and practice", *World Competition: Law and Economics Review*, 30.

3.10. Appendix A: Proofs

Proof of Proposition 1. Note first that if the firms create both cartels at the beginning of a period, the probability that the AA detects both cartels during this period is q^2, that it detects cartel 1 (cartel 2) whereas it does not detect cartel 2 (cartel 1) is $q(1-q)$, and that it detects none of the cartels is $(1-q)^2$. These probabilities follow directly from the independence assumption on the AA's detection technology.

i/ Assume that $\delta < \widetilde{\delta}(q, \theta_1)$. In this case, both cartels are individually unstable. The firms know that, regardless of their reporting decisions right after the detection of one cartel, they will not be able to sustain the remaining cartel in the following period. Hence, there are two possible Nash Equilibria at the end of a period where the AA detects only one of the cartels: either both firms report the remaining cartel k where each firm gets an expected payoff of $-\frac{1}{2}F_k$ or both firms do not report cartel k where each firm gets a payoff of 0. Since we are looking for the best collusive subgame-perfect equilibrium, we focus on the Pareto superior equilibrium where the firms do not report the remaining cartel. The expected present discounted value each firm gets from the creation of both cartels is

$$\begin{aligned} V_{12} &= q^2(\pi_1 + \pi_2 - F_1 - F_2) + q(1-q)(\pi_1 + \pi_2 - F_1) \\ &+ q(1-q)(\pi_1 + \pi_2 - F_2) + (1-q)^2(\pi_1 + \pi_2 + \delta V_{12}) \end{aligned}$$

which we rewrite as

$$V_{12} = \underbrace{\frac{\pi_1 - qF_1}{1 - \delta(1-q)^2}}_{\leq V_1} + \underbrace{\frac{\pi_2 - qF_2}{1 - \delta(1-q)^2}}_{\leq V_2} \tag{3.4}$$

Since $V_1 < 2\pi_1$ and $V_2 < 2\pi_2$, the optimal unilateral deviation occurs in both markets. The two cartels are then jointly stable if and only if

$$V_{12} \geq 2\pi_1 + 2\pi_2$$

From expression (3.4), it follows that $V_{12} \leq V_1 + V_2 < 2\pi_1 + 2\pi_2$. Hence, the optimal unilateral deviation results in a higher payoff than the expected present discounted value each firm gets when forming both cartels. As a consequence, the two cartels are not jointly stable, and the firms do not form both cartels. Since $V_1 < 2\pi_1$ and $V_2 < 2\pi_2$, the firms do neither form one cartel alone and thus, they form no cartel at all.

ii/ Assume that $\widetilde{\delta}(q,\theta_1) \leq \delta < \widetilde{\delta}(q,\theta_2)$. In this case, if the firms form both cartels but the AA detects cartel 2, then, if cartel 1 is not reported, the firms will again form cartel 1 in the next period. However, if the AA detects cartel 1, the firms will not form cartel 2 in the next period. As in case i/, if one cartel is discovered, the firms do not report the remaining cartel in the absence of Amnesty Plus. Hence, the expected present discounted value each firm gets from the creation of both cartels is

$$\begin{aligned} V_{12} &= q^2(\pi_1+\pi_2-F_1-F_2)+q(1-q)(\pi_1+\pi_2-F_1) \\ &+ q(1-q)(\pi_1+\pi_2-F_2+\delta V_1)+(1-q)^2(\pi_1+\pi_2+\delta V_{12}) \end{aligned}$$

which we rewrite as

$$V_{12} = \frac{\pi_1 - qF_1}{1-\delta(1-q)} + \frac{\pi_2 - qF_2}{1-\delta(1-q)^2} = V_1 + \underbrace{\frac{\pi_2 - qF_2}{1-\delta(1-q)^2}}_{\leq V_2 < 2\pi_2} \tag{3.5}$$

Since $V_1 \geq 2\pi_1$ and $V_2 < 2\pi_2$, the optimal unilateral deviation is to deviate in market 2 only. This deviation results in a payoff of $V_1 + 2\pi_2$ which is greater than V_{12}. Hence, the two cartels are not jointly stable, and the firms therefore do not form both cartels. However, since cartel 1 is individually stable whereas cartel 2 is not, it is a Nash Equilibrium to form cartel 1 alone but not to form cartel 2 without cartel 1.

iii/ Assume that $\widetilde{\delta}(q,\theta_2) < \delta$. In this case, if the firms form both cartels but the AA detects one of them, they will again form the remaining cartel in the next period. Hence, the expected present discounted value each firm gets from the creation of both cartels is

$$\begin{aligned} V_{12} &= q^2(\pi_1 + \pi_2 - F_1 - F_2) + q(1-q)(\pi_1 + \pi_2 - F_1 + \delta V_2) \\ &+ q(1-q)(\pi_1 + \pi_2 - F_2 + \delta V_1) + (1-q)^2(\pi_1 + \pi_2 + \delta V_{12}) \end{aligned}$$

which we rewrite as

$$V_{12} = \frac{\pi_1 - qF_1}{1-\delta(1-q)} + \frac{\pi_2 - qF_2}{1-\delta(1-q)} = V_1 + V_2 \tag{3.6}$$

The optimal deviation occurs in market 2 only (see step 1 in Proof of Proposition 6) which results in a payoff of $V_1 + 2\pi_2$. Since $V_2 \geq 2\pi_2$ this payoff is weakly smaller than V_{12}. Hence, the creation of the two cartels is a Nash Equilibrium. Since $V_k > 0$ which implies that $V_{12} > V_1$ and $V_{12} > V_2$, the payoff from the joint creation of the cartels is higher than the payoffs from both the Nash Equilibrium where the firms form only cartel 1 and the Nash Equilibrium where the firms form only cartel 2. As a consequence, the firms form both cartels.

■

Proof of Proposition 2. We proceed in 3 steps. In step 1, we show that the optimal unilateral deviation occurs in both markets. In step 2, we determine the expected present discounted value V_{12} a firm gets from the creation of the two cartels. We then show in step 3 that the joint stability condition can never be satisfied for any value of δ.

Step 1. Since both cartels are individually unstable we know that $V_1 < 2\pi_1$ and $V_2 < 2\pi_2$. Hence, it must be true that $2\pi_1 + 2\pi_2 > V_1 + 2\pi_2$ and $2\pi_1 + 2\pi_2 > 2\pi_1 + V_2$. The optimal unilateral deviation therefore takes place in both markets.

Step 2. After the detection of one cartel, reporting the remaining cartel is a dominant strategy for each firm. This equilibrium strategy gives each firm an expected payoff of $\frac{1}{2}R_1 - \frac{1}{2}F_2$ after the detection of cartel 1 and $\frac{1}{2}R_2 - \frac{1}{2}F_1$ after the detection of cartel 2. The expected present discounted value V_{12} is then

$$\begin{aligned} V_{12} &= q^2(\pi_1 + \pi_2 - F_1 - F_2) + q(1-q)(\pi_1 + \pi_2 - F_1 + \frac{1}{2}R_1 - \frac{1}{2}F_2) \\ &+ q(1-q)(\pi_1 + \pi_2 - F_2 + \frac{1}{2}R_2 - \frac{1}{2}F_1) + (1-q)^2(\pi_1 + \pi_2 + \delta V_{12}) \end{aligned}$$

which we rewrite as

$$V_{12} = \underbrace{\frac{\pi_1 + \pi_2 - q(F_1 + F_2)}{1 - \delta(1-q)^2}}_{\leq V_1 + V_2} - \frac{q(1-q)}{2\Big(1 - \delta(1-q)^2\Big)}(F_1 + F_2 - R_1 - R_2) \tag{3.7}$$

Step 3. For the two cartels to be jointly stable, it is necessary and sufficient that the payoff from the optimal unilateral deviation does not exceed V_{12}, that is

$$V_{12} \geq 2\pi_1 + 2\pi_2$$

From equation (3.7), we know that $V_{12} \leq V_1 + V_2$ and as both cartels are individually unstable, we have $V_1 + V_2 < 2\pi_1 + 2\pi_2$. Hence, the joint stability condition never holds, and the firms do not form these cartels together. Moreover, since $V_1 < 2\pi_1$ and $V_2 < 2\pi_2$, the firms neither form one cartel alone. ■

Proof of Lemma 3. We first show that the optimal unilateral deviation takes place only in market 2 (step 1). We then determine the expected present discounted value V_{12} each firm gets when forming both cartels (step 2) and derive the joint stability condition from which we easily get the joint stability threshold (step 3).

Step 1. Since cartel 1 is individually stable whereas cartel 2 is not, we have $V_1 \geq 2\pi_1$ and $V_2 < 2\pi_2$. It follows from these two inequalities that $V_1 + 2\pi_2 \geq 2\pi_1 + 2\pi_2 > V_2 + 2\pi_1$. The optimal unilateral deviation therefore occurs in market 2 only.

Step 2. After the detection of the individually stable cartel 1, it is a dominant strategy for each firm to report the individually unstable cartel 2. The expected present discounted value V_{12} is:

$$V_{12} = q^2(\pi_1 + \pi_2 - F_1 - F_2) + q(1-q)(\pi_1 + \pi_2 - F_1 + \frac{1}{2}R_1 - \frac{1}{2}F_2)$$
$$+q(1-q)(\pi_1 + \pi_2 - F_2 + \delta V_1) + (1-q)^2(\pi_1 + \pi_2 + \delta V_{12})$$

We rewrite this expression as

$$V_{12} = V_1 + \frac{\pi_2 - qF_2}{1-\delta(1-q)^2} + \frac{q(1-q)}{2\Big(1-\delta(1-q)^2\Big)}(R_1 - F_2) \tag{3.8}$$

Step 3. The formation of both cartels constitutes a Nash Equilibrium if a firm has no incentive to unilaterally deviate from the collusive agreements in both markets. For the two cartels to be jointly stable it is thus necessary and sufficient that $V_{12} \geq V_1 + 2\pi_2$. As $V_{12} \geq V_1$, the joint stability condition also implies that whenever the formation of both cartels is a Nash Equilibrium, it leads to higher profits than the Nash Equilibrium where the firms form cartel 1 only. Rewritten on δ the joint stability condition becomes

$$\delta \geq \frac{1}{2(1-q)^2} - \frac{1}{4(1-q)^2}\theta_2\left(\frac{R_1}{F_2}q(1-q) - q(3-q)\right) \equiv \breve{\delta}\left(q, \theta_2, \frac{R_1}{F_2}\right)$$

The formation of both cartels is the best collusive equilibrium if and only if $\delta \geq \breve{\delta}\left(q, \theta_2, \frac{R_1}{F_2}\right)$.

■

Proof of Proposition 4. Depending on the fine discount in market 2, we mainly distinguish three situations:

i/ $R_2 \leq \frac{\pi_1 + qF_1}{1-q}$

As $\widetilde{\delta}(q, \theta_1) \geq \widehat{\delta}\left(q, \theta_1, \frac{R_2}{\pi_1}\right)$, cartel 1 survives a detection of cartel 2 for all $\delta \in [\widetilde{\delta}(q, \theta_1), \widetilde{\delta}(q, \theta_2)[$. The expected present discounted value V_{12} each firm gets from the creation of both cartels is given in equation (3.8). The formation of both cartels is the best collusive equilibrium if and only if $\delta \geq \breve{\delta}\left(q, \theta_2, \frac{R_1}{F_2}\right)$.

ii/ $\frac{\pi_1 + qF_1}{1-q} < R_2 < \frac{(1+q\theta_2)(\pi_1 - qF_1)}{(1-q)(1-q\theta_2)}$

As $\widetilde{\delta}(q, \theta_1) < \widehat{\delta}\left(q, \theta_1, \frac{R_2}{\pi_1}\right) < \widetilde{\delta}(q, \theta_2)$, cartel 1 survives a detection of cartel 2 if and only if $\delta \in [\widehat{\delta}\left(q, \theta_1, \frac{R_2}{\pi_1}\right), \widetilde{\delta}(q, \theta_2)[$. Hence, if $\delta < \widehat{\delta}\left(q, \theta_1, \frac{R_2}{\pi_1}\right)$ the expression for V_{12} is the same as in equation (3.7). Since $V_{12} \leq V_1 + V_2 < V_1 + 2\pi_2$, the payoff from the optimal unilateral deviation is always strictly higher than the expected present discounted value, and the formation of both cartels is not a Nash Equilibrium. However, if $\delta \geq \widehat{\delta}\left(q, \theta_1, \frac{R_2}{\pi_1}\right)$ the expression for V_{12} is given in equation (3.8), and the formation of both cartels is the best collusive equilibrium only if $\delta \geq \breve{\delta}\left(q, \theta_2, \frac{R_1}{F_2}\right)$.

iii/ $R_2 \geq \frac{(1+q\theta_2)(\pi_1 - qF_1)}{(1-q)(1-q\theta_2)}$

Under this condition we know that $\widetilde{\delta}(q,\theta_2) \leq \widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)$. Cartel 1 does never survive a detection of cartel 2 for any value of $\delta \in [\widetilde{\delta}(q,\theta_1),\widetilde{\delta}(q,\theta_2)[$, and the expression for V_{12} is given in equation (3.7). Since $V_{12} \leq V_1 + V_2 < V_1 + 2\pi_2$, the formation of both cartels is not a Nash Equilibrium.

■

Proof of Proposition 6. We proceed in 2 steps. In step 1, we show that the optimal unilateral deviation takes place only in market 2. In step 2, we determine the exact circumstances under which the pro-competitive effect of Amnesty Plus occurs.

Step 1. Since both cartels are individually stable, we have $V_1 \geq 2\pi_1$ and $V_2 \geq 2\pi_2$. It follows that $V_2 + 2\pi_1 \geq 2\pi_1 + 2\pi_2$, and it is thus sufficient to show that

$$V_1 + 2\pi_2 \geq V_2 + 2\pi_1$$

which is the same as

$$V_1 - 2\pi_1 \geq V_2 - 2\pi_2 \Longleftrightarrow \pi_1\left(\frac{1-q\theta_1}{1-\delta(1-q)} - 2\right) \geq \pi_2\left(\frac{1-q\theta_2}{1-\delta(1-q)} - 2\right)$$

The above inequality holds because $\pi_1 \geq \pi_2$, $\theta_1 \leq \theta_2$ and the expressions in the brackets are positive. The optimal unilateral deviation therefore takes place only in market 2 and yields a payoff of $V_1 + 2\pi_2$.

Step 2.

i/ $R_1 \leq \frac{\pi_2+qF_2}{1-q}$ and $R_2 \leq \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$

As $\widetilde{\delta}(q,\theta_2) \geq \widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right)$ and $\widetilde{\delta}(q,\theta_2) \geq \widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)$, each cartel survives the detection of the other cartel for all $\delta \geq \widetilde{\delta}(q, \theta_2)$. Thus, the expected present discounted value V_{12} each firm gets when forming both cartels is:

$$V_{12} = q^2(\pi_1 + \pi_2 - F_1 - F_2) + q(1-q)(\pi_1 + \pi_2 - F_1 + \delta V_2)$$
$$+q(1-q)(\pi_1 + \pi_2 - F_2 + \delta V_1) + (1-q)^2(\pi_1 + \pi_2 + \delta V_{12})$$

We can rewrite this expression as

$$V_{12} = \frac{\pi_1 - qF_1}{1 - \delta(1-q)} + \frac{\pi_2 - qF_2}{1 - \delta(1-q)} = V_1 + V_2$$

Since $V_2 \geq 2\pi_2$, the payoff from the unilateral optimal deviation does not exceed V_{12}, and the formation of both cartels is the best collusive equilibrium. The firms create both cartels, and Amnesty Plus is neutral.

ii/ $R_1 > \frac{\pi_2 + qF_2}{1-q}$ and $R_2 \leq \frac{(1+q\theta_2)(\pi_1 - qF_1)}{(1-q)(1-q\theta_2)}$

Cartel 1 survives the detection of cartel 2 for all $\delta \geq \widetilde{\delta}(q, \theta_2)$ whereas cartel 2 survives the detection of cartel 1 if and only if $\delta \geq \widehat{\delta}\left(q, \theta_2, \frac{R_1}{\pi_2}\right)$. If $\delta \geq \widehat{\delta}\left(q, \theta_2, \frac{R_1}{\pi_2}\right)$, the analysis is the same as in i/ and leads to the result that the firms form both cartels, and Amnesty Plus is neutral. However, if $\widetilde{\delta}(q, \theta_2) \leq \delta < \widehat{\delta}\left(q, \theta_2, \frac{R_1}{\pi_2}\right)$, the expression for V_{12} is given in equation (3.8). The firms form both cartels only if the joint stability condition holds, that is $\delta \geq \breve{\delta}\left(q, \theta_2, \frac{R_1}{F_2}\right)$. From Proposition 4, we know that $\widetilde{\delta}(q, \theta_2) > \breve{\delta}\left(q, \theta_2, \frac{R_1}{F_2}\right)$ if $R_1 > \frac{1+q}{1-q}F_2 + \frac{2\pi_2}{1-q}$. We thus distinguish between two subcases:

a/ $R_1 > \frac{1+q}{1-q}F_2 + \frac{2\pi_2}{1-q}$

Since $\widetilde{\delta}(q, \theta_2) > \breve{\delta}\left(q, \theta_2, \frac{R_1}{F_2}\right)$, we have $\delta > \breve{\delta}\left(q, \theta_2, \frac{R_1}{F_2}\right)$ for all $\delta \in \left[\widetilde{\delta}(q, \theta_2), \widehat{\delta}\left(q, \theta_2, \frac{R_1}{\pi_2}\right)\right[$.

The firms form both cartels for any δ in this interval.

b/ $R_1 \leq \frac{1+q}{1-q}F_2 + \frac{2\pi_2}{1-q}$

Since $\widetilde{\delta}(q,\theta_2) \leq \breve{\delta}\Big(q,\theta_2,\frac{R_1}{F_2}\Big)$, the firms form both cartels if $\delta \geq \breve{\delta}\Big(q,\theta_2,\frac{R_1}{F_2}\Big)$ whereas they form only cartel 1 if $\delta \in \Big[\widetilde{\delta}(q,\theta_2),\breve{\delta}\Big(q,\theta_2,\frac{R_1}{F_2}\Big)\Big[$. In the latter case, Amnesty Plus has a pro-competitive effect on cartel formation.

iii/ $R_1 \leq \frac{\pi_2+qF_2}{1-q}$ and $R_2 > \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$

Cartel 2 survives the detection of cartel 1 for all $\delta \geq \widetilde{\delta}(q,\theta_2)$, whereas cartel 1 survives the detection of cartel 2 if and only if $\delta \geq \widehat{\delta}\Big(q,\theta_1,\frac{R_2}{\pi_1}\Big)$. If $\delta \geq \widehat{\delta}\Big(q,\theta_1,\frac{R_2}{\pi_1}\Big)$, the analysis is the same as in i/ and leads to the result that the firms form both cartels. However, suppose now that $\widetilde{\delta}(q,\theta_2) \leq \delta < \widehat{\delta}\Big(q,\theta_1,\frac{R_2}{\pi_1}\Big)$. For this range of discount factor values, we derive the expression for V_{12} from equation (3.8) by swapping 1 and 2, that is

$$V_{12} = V_2 + \frac{\pi_1 - qF_1}{1-\delta(1-q)^2} + \frac{q(1-q)}{2\Big(1-\delta(1-q)^2\Big)}(R_2 - F_1) \tag{3.9}$$

The formation of both cartels is the best collusive equilibrium if and only if the cartels are jointly stable. This is equivalent to

$$\underbrace{\left(\frac{\pi_1 - qF_1}{1-\delta(1-q)^2} - \frac{\pi_1 - qF_1}{1-\delta(1-q)}\right)}_{A\leq 0} + \underbrace{\left(\frac{\pi_2 - qF_2}{1-\delta(1-q)} - 2\pi_2\right)}_{B\geq 0} + \underbrace{\frac{q(1-q)}{2(1-\delta(1-q)^2)}(R_2 - F_1)}_{C\leq 0} \geq 0 \tag{3.10}$$

Note that inequality (3.10) does not depend on R_1. Let us show that the set of values $\delta \in \Big[\widetilde{\delta}(q,\theta_2),\widehat{\delta}\Big(q,\theta_1,\frac{R_2}{\pi_1}\Big)\Big[$ which satisfies inequality (3.10) is not empty if the detection probability is small enough. Setting δ to its upper bound $\widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)^-$, the terms A and C go to 0 as $q \to 0$ whereas the term B goes to $\frac{\pi_2}{1-\frac{R_2}{\pi_1+R_2}} - 2\pi_2$. This expression is strictly positive since $R_2 > \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)} > \pi_1$. Hence, we can conclude that for q sufficiently small and δ close enough to $\widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)$, condition (3.10) holds. Moreover, we can say that the set of values $\delta \in \left[\widetilde{\delta}(q,\theta_2),\widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)\right[$ which does not satisfy condition (3.10) is never empty whatever the value of $q > 0$. This is so because inequality (3.10) does not hold for $\delta = \widetilde{\delta}(q,\theta_2)$ and thus, due to the continuity of the LHS of (3.10) with respect to δ, it does not hold for δ sufficiently close to $\widetilde{\delta}(q,\theta_2)$. Therefore, the set over which Amnesty Plus has a pro-competitive effect within the interval $\left[\widetilde{\delta}(q,\theta_2),\widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)\right[$ is never empty.

iv/ $R_1 > \frac{\pi_2+qF_2}{1-q}$ and $R_2 > \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)}$

We have $\widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right) \geq \widetilde{\delta}(q,\theta_2)$ and $\widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right) \geq \widetilde{\delta}(q,\theta_2)$.

If $\delta \geq \max\left(\widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right),\widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)\right)$ the firms form both cartels, and Amnesty Plus is neutral. If $\widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right) \leq \delta < \widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right)$ we get V_{12} from equation (3.8), and the analysis is the same as in ii/. If $\widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right) \leq \delta < \widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)$ we get V_{12} from equation (3.9), and the analysis from iii/ applies. If $\widetilde{\delta}(q,\theta_2) \leq \delta < \min\left(\widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right),\widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)\right)$ the expression for V_{12} is given in (3.7), and the firms form both cartels if and only if the joint stability condition holds, that is

$$\underbrace{\left(\frac{\pi_1-qF_1}{1-\delta(1-q)^2}-\frac{\pi_1-qF_1}{1-\delta(1-q)}\right)}_{D\leq 0}+\underbrace{\left(\frac{\pi_2-qF_2}{1-\delta(1-q)^2}-2\pi_2\right)}_{E\geq 0 \text{ or } E\leq 0} + \underbrace{\frac{q(1-q)}{2(1-\delta(1-q)^2)}(R_1+R_2-F_1-F_2)}_{F\leq 0} \geq 0 \tag{3.11}$$

Let us show that the set of values $\delta \in \left[\widetilde{\delta}(q,\theta_2), \min\left(\widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right), \widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)\right)\right[$ which satisfies inequality (3.11) is not empty if q is sufficiently small. Setting δ to its upper bound $\min\left(\widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right), \widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)\right)^-$, the terms D and F go to 0 as $q \to 0$ whereas the term E goes to $\frac{\pi_2}{1-\min\left(\frac{R_2}{\pi_1+R_2},\frac{R_1}{\pi_2+R_1}\right)} - 2\pi_2$ which is strictly positive since $R_2 > \frac{(1+q\theta_2)(\pi_1-qF_1)}{(1-q)(1-q\theta_2)} > \pi_1$ and $R_1 > \frac{\pi_2+qF_2}{1-q} > \pi_2$. Therefore, for q sufficiently small and for δ close enough to $\min\left(\widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right), \widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)\right)$ condition (3.11) holds. Moreover, the set of values $\delta \in \left[\widetilde{\delta}(q,\theta_2), \min\left(\widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right), \widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)\right)\right[$ which does not satisfy condition (3.11) is never empty for any value of $q > 0$. This is because condition (3.11) does not hold for $\delta = \widetilde{\delta}(q,\theta_2)$ and, due to the continuity of its LHS with respect to δ, it does not hold for δ sufficiently close to $\widetilde{\delta}(q,\theta_2)$. Therefore, the set over which Amnesty Plus has a pro-competitive effect within the interval $\left[\widetilde{\delta}(q,\theta_2), \min\left(\widehat{\delta}\left(q,\theta_2,\frac{R_1}{\pi_2}\right), \widehat{\delta}\left(q,\theta_1,\frac{R_2}{\pi_1}\right)\right)\right[$ is never empty.

■

Proof of Proposition 7. Note first that when firms use multimarket strategies, the optimal unilateral deviation is to deviate in both markets since punishment occurs in both markets. A firm's payoff from such an optimal deviation is $2\pi_1 + 2\pi_2$.

i/ Assume that $\delta < \widetilde{\delta}(q,\theta_1)$. From the analysis of the EU antitrust policy when firms use standard trigger strategies and, especially, from the proof of Proposition 1, we know that the expected present discounted value V_{12} each firm gets from the creation of both cartels is equal to expression (3.4):

$$V_{12} = \underbrace{\frac{\pi_1 - qF_1}{1-\delta(1-q)^2}}_{<V_1} + \underbrace{\frac{\pi_2 - qF_2}{1-\delta(1-q)^2}}_{<V_2} \leq V_1 + V_2 < 2\pi_1 + 2\pi_2$$

A unilateral deviation in both markets is therefore always profitable, and the two cartels are not jointly stable. Hence, the firms do not form both cartels. Moreover, since $V_1 < 2\pi_1$ and $V_2 < 2\pi_2$, they do not form only one of the cartels either.

ii/ Assume that $\widetilde{\delta}(q,\theta_1) \leq \delta < \widetilde{\delta}(q,\theta_2)$. From (3.5) we get the expression for V_{12} which is:

$$V_{12} = V_1 + \underbrace{\frac{\pi_2 - qF_2}{1-\delta(1-q)^2}}_{\leq V_2 < 2\pi_2}$$

Since cartel 1 is individually stable whereas cartel 2 is not, it is a Nash Equilibrium to form cartel 1 alone, whereas it is not a Nash Equilibrium to form cartel 2 without cartel 1. The formation of both cartels is the best collusive equilibrium if and only if $V_{12} \geq 2\pi_1 + 2\pi_2$. Note that this condition is less stringent than its counterpart when firms use standard trigger strategies, i.e. $V_{12} \geq V_1 + 2\pi_2$. The joint stability condition is therefore more easily satisfied with multimarket strategies which supports the intuition that multimarket contact is basically pro-collusive. We provide, however, a more detailed analysis of this argument: Note that, since V_{12} is increasing in δ over $[\widetilde{\delta}(q,\theta_1), \widetilde{\delta}(q,\theta_2)[$, a necessary and sufficient condition for the inequality $V_{12} \geq 2\pi_1 + 2\pi_2$ to hold over a non-empty range of values of $\delta \in [\widetilde{\delta}(q,\theta_1), \widetilde{\delta}(q,\theta_2)[$ is that $V_{12}\left(\widetilde{\delta}(q,\theta_2)^-\right) > 2\pi_1 + 2\pi_2$. After some algebraic manipulations, we can write the following:

$$V_{12}\left(\widetilde{\delta}(q,\theta_2)^-\right) - (2\pi_1 + 2\pi_2) = 2\pi_1 \frac{q(\theta_2-\theta_1)}{1-q\theta_2} - 2\pi_2 \frac{q(1+q\theta_2)}{1-q\theta_2+q(1+q\theta_2)}$$

Therefore,

$$\begin{aligned} V_{12}\left(\widetilde{\delta}(q,\theta_2)^-\right) &> 2\pi_1 + 2\pi_2 \qquad (3.12) \\ \Longleftrightarrow & \underbrace{\frac{(\theta_2-\theta_1)\left(1-q\theta_2+q(1+q\theta_2)\right)}{1-q^2\theta_2^2}}_{\text{term A}} \underbrace{\frac{\pi_1}{\pi_2}}_{\text{term B}} > 1 \end{aligned}$$

We show that term A is increasing in θ_2 over $[\theta_1, \frac{1}{q} - 2]$ by differentiating it with respect to θ_2. Since $\theta_2 = \frac{F(\pi_2)}{\pi_2}$ is decreasing in π_2, term A is also decreasing in π_2 over $[0, \pi_1]$. Term B, which is the same as $\frac{1}{\lambda}$, is clearly decreasing in π_2. Then, the LHS of inequality (3.12) which is the multiplication of terms A and B is decreasing in π_2 over $[0, \pi_1]$. Moreover, it is continuous in π_2, goes to $+\infty$ as $\pi_2 \longrightarrow 0$ and takes the value 0 for $\pi_2 = \pi_1^-$. Using the intermediate value theorem, we can say that there exists a threshold $\widetilde{\pi}_2(q)$ such that inequality (3.12) is satisfied if and only if $\pi_2 < \widetilde{\pi}_2(q)$. Hence, we distinguish two subcases:

– If $\pi_2 < \widetilde{\pi}_2(q)$ then $V_{12} > 2\pi_1 + 2\pi_2$ for $\delta = \widetilde{\delta}(q,\theta_2)^-$, whereas $V_{12} < 2\pi_1 + 2\pi_2$ for $\delta = \widetilde{\delta}(q,\theta_1)$. Since V_{12} is continuous and increasing in δ, we can again use the intermediate value theorem to conclude that there exists a threshold $\bar{\delta}(q,\theta_1,\theta_2,\lambda) = \bar{\delta}$ such that $V_{12} < 2\pi_1 + 2\pi_2$ for $\delta \in [\widetilde{\delta}(q,\theta_1), \bar{\delta}[$ and $V_{12} \geq 2\pi_1 + 2\pi_2$ for $\delta \in [\bar{\delta}, \widetilde{\delta}(q,\theta_2)[$. Hence, the formation of both cartels is the best collusive equilibrium for $\delta \in [\bar{\delta}, \widetilde{\delta}(q,\theta_2)[$ but is not an equilibrium for $\delta \in [\widetilde{\delta}(q,\theta_1), \bar{\delta}[$.

– If $\pi_2 \geq \widetilde{\pi}_2(q)$ then $V_{12} \leq 2\pi_1 + 2\pi_2$ for $\delta = \widetilde{\delta}(q, \theta_2)^-$. It follows that the inequality $V_{12} < 2\pi_1 + 2\pi_2$ holds for all $\delta \in [\widetilde{\delta}(q, \theta_1), \widetilde{\delta}(q, \theta_2)[$ which implies that forming both cartels is never a Nash Equilibrium for $\delta \in [\widetilde{\delta}(q, \theta_1), \widetilde{\delta}(q, \theta_2)[$.

iii/ Assume that $\delta \geq \widetilde{\delta}(q, \theta_2)$. In this case, we get the expression for V_{12} from (3.6), that is:

$$V_{12} = \frac{\pi_1 - qF_1}{1 - \delta(1 - q)} + \frac{\pi_2 - qF_2}{1 - \delta(1 - q)} = V_1 + V_2 \geq 2\pi_1 + 2\pi_2$$

From the above inequality, it is straightforward that the formation of both cartels is the best collusive equilibrium for all $\delta > \widetilde{\delta}(q, \theta_2)$.

■

3.11. Appendix B: Extensions

3.11.1. Appendix B1: Relaxation of the assumption $\theta_2 \geq \theta_1$

Suppose now that the opposite assumption $\theta_1 > \theta_2$ holds. The direct implication, albeit somewhat counterintuitive, is that cartel 2 is easier to sustain than the more profitable cartel 1, i.e. $\widetilde{\delta}(q, \theta_2) < \widetilde{\delta}(q, \theta_1)$.

It is straightforward that Proposition 1 remains valid, provided that we reverse the subscripts 1 and 2. Hence, under the EU antitrust policy, the firms form a cartel if and only if it is individually stable.

Proposition 2 remains true, but we have to substitute $\widetilde{\delta}(q, \theta_2)$ for $\widetilde{\delta}(q, \theta_1)$, or more generally, if we do not make any assumption on the size of θ_1 relative to θ_2, we substitute

$\min\left(\widetilde{\delta}(q,\theta_1),\widetilde{\delta}(q,\theta_2)\right)$ for $\widetilde{\delta}(q,\theta_1)$. Hence, if no cartel is individually stable, the firms do not form any of the cartels under the US antitrust policy, and Amnesty Plus is still neutral.

It is easy to show that by reversing the subscripts 1 and 2, the first part of Proposition 4 still holds. However, whereas the necessary and sufficient conditions for Amnesty Plus to have an anticompetitive effect which we provide in the second part of Proposition 4 may be satisfied under the initial assumption $\theta_2 \geq \theta_1$, the reverse is not true. One of the new conditions defining a non-empty range of values of δ for which Amnesty Plus has an anticompetitive effect would be $R_2 > \frac{1+q}{1-q}F_1 + \frac{2\pi_1}{1-q}$. Hence, since we have $R_2 \leq F_2 \leq F_1 < \frac{1+q}{1-q}F_1 + \frac{2\pi_1}{1-q}$, the latter condition cannot be satisfied, and the potential anticompetitive effect of Amnesty Plus cannot occur.

Proposition 6 remains valid, although we have to substitute $\widetilde{\delta}(q,\theta_1)$ for $\widetilde{\delta}(q,\theta_2)$. More generally, if we do not want to make any particular assumption on the relative size of θ_1 and θ_2, we substitute $\max\left(\widetilde{\delta}(q,\theta_1),\widetilde{\delta}(q,\theta_2)\right)$ for $\widetilde{\delta}(q,\theta_2)$. Hence, Amnesty Plus may still have a pro-competitive effect on cartel formation.

3.11.2. Appendix B2: Unrestricted strategy choice

We relax the assumption that the firms can form a cartel in a period $t > 0$ only if this cartel has been formed in the previous period. The key difference with respect to our initial time structure is that, if the firms form only one cartel in some period, and the AA detects this cartel during this period, they still have to possibility to form the other cartel in the following period. More precisely, we modify the timing within a period $t \geq 1$ as follows: If no cartel has been convicted in the previous period, the time structure of the latter applies to the current period. If both cartels have been formed and convicted in the previous period, the firms compete in both markets. If either one cartel has been formed and not convicted or both cartels have been formed and only one has been convicted, the timing is the same as the one presented in stages 0 to 3 of our initial set-up.

The modification of the time structure does not affect our results under the EC Leniency Program. In particular, Proposition 1 remains valid. Neither does the modification affect the results under the US Amnesty Program with Amnesty Plus for $\delta < \widetilde{\delta}(q, \theta_1)$ and $\widetilde{\delta}(q, \theta_1) \leq \delta < \widetilde{\delta}(q, \theta_2)$. Both the neutrality of Amnesty Plus result in Proposition 2 and the result on the potential anticompetitive effect of Amnesty Plus stated in Proposition 4 still hold.

Allowing for an unrestricted strategy choice, however, alters our results in the region where Amnesty Plus may have a pro-competitive effect, i.e. for $\delta \geq \widetilde{\delta}(q, \theta_2)$. The fact that the firms may now *start* forming a cartel in a period $t \geq 1$ gives rise to a new equilibrium where they form cartel 1 until it is detected and then form cartel 2. This is an equilibrium because for $\delta \geq \widetilde{\delta}(q, \theta_2)$, both cartels are individually stable. Amnesty Plus cannot prevent such an outcome. Each firm gets an expected payoff V_1^2 such that

$$V_1^2 = q(\pi_1 - F_1 + \delta V_2) + (1 - q)(\pi_1 + \delta V_1^2)$$

The AA detects cartel 1 with probability q in which case the firms form cartel 2 in the following period. With probability $(1 - q)$ the AA does not detect cartel 1, and the firms form it again in the following period. Solving the above equation for V_1^2 we get

$$V_1^2 = V_1 + q\frac{\delta}{1 - \delta(1 - q)}V_2$$

For the joint formation of both cartels to be the most profitable equilibrium, two conditions must hold: First, $V_{12} \geq V_1 + 2\pi_2$, i.e. the optimal unilateral deviation must not be profitable, second, $V_{12} \geq V_1^2$, i.e. the equilibrium where the firms form both cartels must be more profitable than the equilibrium where the firms form first cartel 1 and then only, if detected, cartel 2. We can combine these two conditions as follows:

$$V_{12} \geq \max(V_1 + 2\pi_2, V_1^2) = \max\left(V_1 + 2\pi_2, V_1 + q\frac{\delta}{1-\delta(1-q)}V_2\right)$$

It is straightforward that the above condition is weakly more stringent than the condition $V_{12} \geq V_1 + 2\pi_2$. Moreover, it is strictly more stringent than the latter for at least some values of the parameters δ and q. This is because $q\frac{\delta}{1-\delta(1-q)}V_2 = q\frac{\delta}{1-\delta(1-q)}\frac{\pi_2 - qF_2}{1-\delta(1-q)}$ and $\lim_{\delta\to 1}\left(\lim_{q\to 0} q\frac{\delta}{1-\delta(1-q)}\frac{\pi_2-qF_2}{1-\delta(1-q)}\right) = +\infty$. Hence, for small values of q the expression $q\frac{\delta}{1-\delta(1-q)}V_2$ can be greater than $2\pi_2$ for values of δ sufficiently close to 1.

We have shown that allowing for an unrestricted strategy choice does not affect the potential anticompetitive effect of Amnesty Plus but has an ambiguous impact on its pro-competitive effect. On the one hand, the region where the pro-competitive effect occurs may be larger because it may be more difficult to achieve the joint stability of the cartels. On the other hand, since Amnesty Plus can only deter the formation of cartel 2 as long as cartel 1 goes undetected, the pro-competitive effect of Amnesty Plus, if it occurs, is weaker relative to our previous findings.

Chapter 4

Regulation Stringency and Efficiency: Individual versus Relative Regulation[1]

4.1. Introduction

Various regulatory schemes are used in practice. One dimension along which regulation contracts may differ is whether they are based on the regulated firm's own performances or on other firms' performances. More specifically, when a sector consists of regional monopolies, firms can be regulated on the basis of their own efficiency but they can also be regulated by means of a relative performance regulatory scheme, such as *yardstick competition* (Shleifer 1985). The underlying idea of yardstick competition is the regulation of these firms using constructed benchmarks that are based on costs of all other firms in the considered industry, but are independent of the costs of the firm the benchmark is created

[1] In this chapter we explore in a more general framework an issue studied in Ebel and Lefouili (2008), except for subsection 4.4.3. which is a reprint of subsection 5.2 of the latter paper.

for. Due to this independence, firms are expected to have higher incentives to reduce and reveal their costs (Shleifer 1985, Dalen, 1998, Tangerås 2002). Yardstick competition has been implemented in utility industries in many countries, such as the electricity industry in the UK, Switzerland, Chile and Germany,[2] the water industry in the UK and Italy and the gas distribution sector in Japan[3]. It has also been used in the bus industry in Norway and the telecommunications sector in the US.[4]

This paper examines how cost reduction incentives are affected by the stringency of regulation when *ex post* price-cap schemes are used.[5] We analyze the cost reduction efforts of symmetric local monopolies under two types of regulation. First, we assume the ceiling price set for each firm to depend on its own *realized* cost. We refer to this regulation regime as *individual regulation.* Second, we deal with the case where the ceiling price for a given firm depends on the average of the realized marginal costs of the other firms in the industry. We refer to this regulation regime as *relative regulation.* We do not allow for transfers on the part of the regulator, thus departing from most of the literature on *yardstick competition.*[6] Our approach is motivated by the empirical observation that most regulators set prices using some regulatory rule without making any transfer to the regulated firms. Though it is clear that making side payments leads to less deadweight loss than setting prices higher than marginal costs, there are some rationales for not using transfers. First,

[2]See Jamasb and Politt (2001) for an extensive survey of different regulation practices in electricity markets around the world.

[3]See Suzuki (2008)

[4]See FCC (1997).

[5]*Ex post* regulation is used in Sweden, Finland and the Netherlands, see Jamasb & Pollitt[2001] and Farsi *et al.*[2007].

[6]To the best of our knowledge, the only paper that examines yardstick competition without transfers on the part of the regulator is Shleifer (1985). In a short extension to his main analysis, he characterizes the optimal regulatory scheme when the only regulatory tool is the price. However, he does not show how higher regulated prices would affect the firms' cost reduction incentives, which is the focus of this paper. Moreover, he does not establish that the second-best outcome the regulator is willing to implement through the proposed scheme is the unique equilibrium whereas the regulatory contracts studied in this paper are shown to induce a unique equilibrium in terms of cost reductions.

collecting public funds is costly and, second, making transfers may increase the likelihood of regulatory capture since it is more likely there will be a lack of transparency relative to a situation where the only regulatory tool is the price.

We model the stringency of regulation through a mark-up parameter, which we suppose to be exogenously given. Since we do not specify the regulator's objective function, we argue that it is possible to embed our analysis in a more general framework than the standard normative one where a regulator maximizes the usual utilitarian social welfare function. In particular, our analysis could be used in a framework where regulators' decisions are influenced by interest groups such as lobbies[7] and therefore need not result in a socially desirable value of the mark-up. Furthermore, the only informational assumption made in our analysis is that realized costs are observable to the regulator. Hence, as in Shleifer (1985), we do not discard the possibility of a lack of information on the part of the regulator, in particular about the cost reduction technology and the regulated firms' initial efficiency.

We show that the cost reduction level is decreasing in the mark-up under relative regulation while it is increasing in the mark-up under individual regulation. Thus a more stringent regulation encourages cost reduction under relative regulation while it undermines the incentives to reduce costs under individual regulation. We also show that a more stringent regulation leads to a decrease in the regulated price under relative regulation while its effect on the price under individual regulation is ambiguous. Due to those differences, there may be a tension between encouraging cost reductions and minimizing prices under individual regulation while this is not true under relative regulation. Note that in the case of regulation with transfers, prices are set equal to marginal costs by regulators who wish to achieve allocative efficiency, and consequently they are negatively related to cost

[7]For an overview of lobbyism activities in the European Union, see Svendsen(2002).

reduction. This makes it possible to derive the comparative statics on prices from those on cost reductions. In our setting, prices are affected by the regulatory tool (the mark-up) in a *direct* way and not only indirectly through cost reductions, which makes the price analysis not just a corollary of the cost reduction analysis.

We show that relative regulation cannot be implemented if the mark-up is too small. Indeed, the artificial competition induced by relative regulation can be so intense that it does not allow regulated firms to break-even. Furthermore, we extend our model by allowing firms to undertake quality-improving investments. We show that our central result that cost reduction investment is undermined by a more lenient relative regulation holds in a setting where firms can conduct quality investments as well. However it is couterbalanced by the positive effect of a more lenient relative regulation on the incentives to enhance quality.

Dalen (1998) also compares individual regulation to *yardstick competition*. He considers an asymmetric information framework and investigates firms' informational rents (granted through transfers) and their incentives for conducting industry-specific and firm-specific investments. He finds that the optimal regime for encouraging investments depends crucially on their nature: *yardstick competition* leads to more firm-specific investments and individual regulation entails more industry-specific investments. In his paper, the advantages of *yardstick competition* are strongly related to the reduction of the regulator's informational problem and her ability to make firms reveal their costs with less distorsion under this regime. This a major difference with our paper since we rather focus on the artificial competition between regulated firms induced by relative regulation (as in Shleifer, 1985). Our paper has two other important differences with Dalen (1998). First, the latter allows for lump-sum transfers on the part of the regulator while we do not. Second, our focus is not on comparing the investment levels under individual regulation and yarsdstick competition when both schemes are socially optimal but rather on comparing how firms react to less stringent (and potentially sub-optimal) regulation under the two regimes.

The remainder of this paper is organized as follows. The basic set-up is presented in Section 4.2. In this section we also characterize the optimal cost reduction and price of an unregulated monopoly, which can be seen as a benchmark to which the outcomes under regulation may be compared. In Section 4.3, we study how the incentives to reduce costs and the prices are affected by the stringency of regulation when an individual scheme is used by the regulator. Section 4.4 is devoted to relative regulation and more specifically to the determination of the equilibrium cost reductions and prices and their comparative statics with respect to the mark-up parameter. Section 4.5 concludes.

4.2. The basic set-up

Consider N symmetric local monopolies.[8] We suppose that all markets are characterized by the same demand function $D(.)$. We assume that fixed production costs are zero and we denote by c the initial (constant) marginal cost of all firms. Firm $i = 1, 2, ..., n$ can reduce its marginal cost from c to $c - u_i$ by spending an amount $C(u_i)$[9]

The assumptions we will set on the demand function $D(.)$ and the investment cost function $C(.)$ are closely regulated to the existence and unicity of an unregulated monopoly's optimal cost reduction and price. Therefore, we start by writing the maximization program of the firms if they were not regulated, where we drop the subscript i because all firms and

[8] E.g. electricity network providers, railway system operators, hospitals, etc.

[9] $C(.)$ can also be interpreted as the manager's disutility function from realizing efficiency gains. Note that all firms share the same investment cost function $C(.)$.

markets are symmetric:

$$\max_{u\in[0,c],p\in[0,+\infty[}\pi(u,p)=D(p)(p-(c-u))-C(u) \tag{4.1}$$

We suppose that the following assumptions hold:

A1 : $D(c)>0$ and there exists $a\in\,]c,+\infty[\,\cup\{+\infty\}$ such that $D(a)=0$. If $a<+\infty$ then D is continuous over $[0,a]$, twice differentiable over $[0,a[$ and strictly decreasing over $[0,a[$ while $D(p)=0$ over $[a,+\infty[$. If $a=+\infty$ then D is twice differentiable and strictly decreasing over $[0,+\infty[$.

This assumption is fulfilled by the vast majority of usual demand functions (linear, iso-elastic,...). Note that we allow for demand functions that reach zero for a finite price as well as demand functions that reach zero only for an infinite price.

A2 : The function $p\rightarrow pD(p)$ is strictly concave over $[0,a[$.

This assumption means that a firm's gross revenue is strictly concave in its price (over the set of prices for which the demand is not zero). This is the case whenever the demand function is either concave or not too convex. Indeed, assumption A2 can be rewritten as the inequality $D"(p)<\frac{-2D'(p)}{p}$ where the right-hand side is strictly positive over $[0,a[$. Assumptions A1 and A2 ensure the existence and uniqueness of the monopoly price when the firm produces with its initial marginal cost c. Anticipating on a notation we will use in lemma 1 (see below), we denote $p^M(0)$ the monopoly price when the firm produces with its marginal cost c.

A3 : The function C is three times differentiable over $[0,c[$. In case $C(c)<+\infty$, this property extends to $[0,c]$. Furthermore $C'\geq 0$, $C">0$ and $C'''\geq 0$.

The concavity assumption captures the idea that it becomes costlier to reduce the marginal cost as the firm becomes more efficient. The assumption about the third derivative of C is a technical assumption (fulfilled by a wide range of functions) which is quite standard in the regulation literature.[10]

A4 : $D(0) < C'(c)$ and $D\left(p^M(0)\right) > C'(0)$.

As it will be made clear below this assumption ensures that corner solutions do not arise under the unregulated monopoly regime. More specifically, the first inequality ensures that an unregulated monopoly does not find it optimal to reduce its marginal cost to zero (which would be quite unrealistic), while the second inequality implies that an unregulated monopoly finds it optimal to invest a strictly positive amount to decrease its marginal cost. Note that the first inequality is automatically satisfied if $C'(c) = +\infty$ and the second one always holds if $C'(0) = 0$.

A5 : $(C')^{-1} oD$ is either strictly concave or linear over $[0, a[$.[11]

Given that C' is weakly convex and strictly increasing (see A3), we can state that $h \equiv (C')^{-1}$ is weakly concave and strictly increasing. Hence A5 holds whenever the demand function D is concave or not too convex (relative to h) . Indeed A5 can be rewritten as $D" \leq -\frac{(h" oD)\times D'^2}{h' oD}$ where the right-hand side is a weakly positive function.

A6 : $\frac{D}{D'}$ is either strictly convex or linear over $[0, a[$.

This technical assumption is fulfilled by a wide range of demand functions including linear and iso-elastic demands as well as polynomial demands having the form $D(p) = (a-p)^n$.[12] Assumption A6, along with A5, ensures the uniqueness of the solution to the maximization

[10]See chapter 4 of Laffont and Tirole (1993) and Tangerås (2002).

[11]Note that this assumption is a bit stronger than assuming $(C')^{-1} oD$ to be weakly convex. We do not make the latter assumption because it will allow the function $(C')^{-1} oD$ to be strictly convex over a subset of $[0, a[$ and linear over the remaining subset which can raise problems about the unicity of the solution to the different maximization programs examined in this paper. The same applies for assumption A6.

[12]Note that, in all the mentioned examples, $\frac{D}{D'}$ is linear.

program of an unregulated monopoly.

We first present some results regarding the behavior of an unregulated monopoly in terms of cost reduction and pricing.

Lemma 1 *For any* $u \in [0, c]$*, there exists a unique price* $p^M(u)$ *that maximizes* $\pi(p, u)$*. Moreover, the monopoly price* $p^M(u)$ *is strictly decreasing and either strictly convex or linear in* u*.*

Proof. See Appendix ■

This lemma states that a decrease in the monopoly's marginal cost yields at most a proportional decrease in the monopoly price. This is used to establish the unicity result in the following proposition.

Proposition 2 *The unregulated monopoly's maximization program (4.1) has a unique solution* (u^M, p^M) *and the optimal cost reduction* u^M *is in the interval* $]0, c[$.

Proof. See Appendix ■

Now that we have characterized the firms' behavior in case they are not regulated, we examine how their decisions are affected by regulation.

4.3. Individual regulation

Under *individual regulation*, firms are regulated on the basis of their own production costs. The regulator does not make transfers to the firm and sets for each firm i a ceiling price depending on its realized marginal cost $c - u_i$.[13] More specifically, we assume that

[13]Such a cost based regulation method can be found in the regulation of the german electricity industry for example.

the ceiling price is given by:

$$p_i^{\max} = \mu(c - u_i),$$

where $\mu \geq 1$ is a mark-up parameter that can be interpreted as an inverse measure of regulation stringency.

Since the firms are symmetric and are treated independently, we can again drop the subscript i when writing the maximization program of an individually regulated firm:

$$\max_{u\in[0,c],p\in[0,+\infty[} \pi(u,p) = D(p)(p - (c - u)) - C(u) \tag{4.2}$$

subject to the regulatory constraint:

$$p \leq \mu(c - u).$$

Before stating the formal result on the optimal cost reduction of an individually regulated firm, let us identify the effects of a cost reduction on its profit. First, there is an obvious negative *cost effect*. Second, there is a negative *margin effect* since cost reduction will decrease the price the firm will be allowed to set . Third, there is a positive *demand effect* since cost reduction will result in lower price and therefore in higher demand.

Proposition 3 *For any $\mu \geq 1$, there exists a unique pair $\left(u^I(\mu), p^I(\mu)\right)$ that maximizes the regulated firm's profit subject to the individual regulatory constraint. Denoting $\mu^M = \frac{p^M}{c-u^M}$, there exists a unique threshold $\tilde{\mu} \in \left[1, \mu^M\right[$ such that*

1. *If $1 \leq \mu \leq \tilde{\mu}$, the regulated firm does not invest in cost reduction and the regulatory constraint is binding: $u^I(\mu) = 0$ and $p^I(\mu) = \mu c$.*

2. *If $\tilde{\mu} < \mu < \mu^M$, the regulated firm invests a strictly positive amount in cost reduction*

and the regulatory constraint is binding: $u^I(\mu) > 0$ *and* $p^I(\mu) = \mu(c - u^I(\mu))$.

3. If $\mu \geq \mu^M$*, the regulated firm behaves as an unregulated monopoly:* $u^I(\mu) = u^M$ *and* $p^I(\mu) = p^M$

Moreover, $\tilde{\mu} = 1$ *if and only if* $C'(0) = 0$ *and* $D(c) + cD'(c) \leq 0$.

Proof. See Appendix ■

Proposition (3) shows that the *demand effect* outweights the *cost effect* and the *margin effect* for sufficiently lenient regulation constraints (i.e. sufficiently high values of the mark-up μ), hence resulting in a strictly positive investment in cost reduction. If the regulation constraint is too stringent, the *cost effect* and the *margin effect* dominate the *demand effect* which yields no investment. However it is shown that whenever c is high enough[14] and the marginal investment cost is very small for weak cost reductions, the regulated firm has an incentive to reduce its marginal cost even if the mark-up is arbitrarly close to 1 (but different from 1). The proposition also shows that there exists a *monopoly-replicating mark-up*, that is, a value of the mark-up above which the regulatory constraints are not binding which means the regulated firm will be able to implement the monopoly outcome.

Let us now see how the stringency of the regulation policy affects each of the three identified effects. An increase in the mark-up does not affect the *cost effect* but it positively affects both the *margin effect* and the *demand effect.* Therefore, the overall effect of a higher mark-up on the investment incentives is *a priori* ambiguous. The next proposition provides a clear-cut result.

Proposition 4 *The cost reduction under individual regulation* $u^I(\mu)$ *is weakly increasing in the mark-up parameter* μ *. More specifically, it is constant over* $[1, \tilde{\mu}]$*, strictly increasing*

[14]The condition $D(c) + cD'(c) \leq 0$ has this interpretation because A2 implies that $D(c) + cD'(c)$ is strictly decreasing in c.

over $\left[\tilde{\mu}, \mu^M\right]$ *and constant over* $\left[\mu^M, +\infty\right[$.

Proof. See Appendix ■

This proposition states that a decrease in the stringency of the regulatory constraint yields (weakly) more investment in cost reduction. A regulator that is mainly concerned with productive efficiency could then encourage cost reductions by granting relatively high mark-ups. The question that arises at this stage is whether such increase of productive efficiency comes at the expense of a decrease in allocative efficiency. The next proposition provides an answer to this question.

Proposition 5 *The price* $p^I(\mu)$ *is strictly increasing in* μ *over* $[1, \tilde{\mu}]$ *and constant over* $\left[\mu^M, +\infty\right[$ *but the effect of a higher mark-up* μ *on the price* $p^I(\mu)$ *is ambiguous over* $\left]\tilde{\mu}, \mu^M\right[$.

A sufficient condition for $p^I(\mu)$ *to be strictly decreasing in* μ *over* $\left]\tilde{\mu}, \mu^M\right[$ *is that* $(c-u)\,C"(u) \leq C'(u)$ *for all* $u \in [0, c]$.

If this condition does not hold then $p^I(\mu)$ *may be increasing in* μ *over* $\left]\tilde{\mu}, \mu^M\right[$. *This is for instance true if the demand is linear, i.e.* $D(p) = a - p$ *and the investment cost function is quadratic, i.e.* $C(u) = \frac{1}{2}\gamma u^2$.

Proof. See Appendix . ■

Relaxing the regulatory constraint, i.e. granting a higher mark-up, affects the regulated price through two effects: a direct effet that captures the way the mark-up affects the price for an unchanged level of marginal cost, and an indirect one that captures how the mark-up affects the price through its effect on cost reduction. The two effects are opposite since the former increases the price whereas the second decreases the price. More formally, for

$\mu \in \left]\tilde{\mu}, \mu^M\right[$:

$$\frac{dp^I(\mu)}{d\mu} = \underbrace{c - u^I(\mu)}_{>0 \text{ (direct effect)}} \underbrace{-\mu \frac{du^I(\mu)}{d\mu}}_{<0 \text{ (indirect effect)}}$$

In the case of $\mu \in [1, \tilde{\mu}]$, the indirect effect does not exist since there is no investment in cost reduction. But whenever the mark-up is in the range $\left]\tilde{\mu}, \mu^M\right[$, the overall effect depends on which of the two effects outweighs the other one. If the direct effect outweighs the indirect one then the regulated price increases if the regulatory constraint is relaxed (i.e. if the mark-up increases), which is quite intuitive. However, proposition (5) suggests that the former result does not always hold: the indirect effect may outweigh the direct effect leading to a decrease of the price in the mark-up parameter. This means that the regulated price decreases if the regulatory constraint is made more lenient, which is rather counter-intuitive. The following examples illustrate the two possibilities:

Example 1 : $D(p) = a - p$ and $C(u) = \gamma \ln\left(\frac{c}{c-u}\right)$

Under this specification, the condition $(c-u)\, C"(u) \leq C'(u)$ is fulfilled. Denoting $\eta = \frac{1}{\mu}$ and assuming that $\mu \in \left]\tilde{\mu}, \mu^M\right[$, straightforward calculations lead to $\frac{dp^I(\eta)}{d\eta} = \frac{p^I(\eta)}{1-\eta} > 0$ which implies that $p^I(\eta)$ is strictly increasing in η or equivalently that $p^I(\mu)$ is strictly decreasing in μ.

Example 2: $D(p) = a - p$ and $C(u) = \frac{1}{2}\gamma u^2$

Under this specification, the condition $(c-u)\, C"(u) \leq C'(u)$ is not fulfilled. Assuming that $\mu \in \left]\tilde{\mu}, \mu^M\right[$, we find that $p^I(\mu) = \frac{a\mu(\mu-1)+\mu\gamma c}{2\mu(\mu-1)+\gamma}$ which is shown to be strictly increasing in μ in the proof of proposition (5).

Hence, there exists a tension between productive efficiency and allocative efficiency when the direct effect of the mark-up on the price outweighs the indirect one: a less stringent individual regulation policy increases the cost reduction incentives but at the expense of a higher price. In this case, allocative efficiency is improved under individual regulation with respect to the unregulated monopoly regime but this comes at the expense of a loss in productive efficiency. If the indirect effect of the mark-up on the price outweighs its direct effect, there is no such tension: a less stringent individual regulation induces more investment and weaker prices. In the latter case, the interesting following results holds: both allocative and productive efficiency are higher if the firms are not regulated at all rather than regulated under an individual scheme based on a mark-up $\mu \in \left]\tilde{\mu}, \mu^M\right[$. Thus, in this case, there is no justification for an individual regulation based on such a mark-up because, it will not only decrease cost reduction relative to the no regulation regime, but it will lead to higher prices as well. However, individual regulation can still be relevant to the regulator if she is mainly concerned with minimizing prices and the initial cost c is weaker than the monopoly price p^M after cost reduction. In this case, the regulator may still prefer setting $\mu = 1$, hence inducing $u = 0$ but $p = c$ rather than leaving the monopoly unregulated.

4.4. Relative regulation

The fundamental idea of yardstick competition is the construction of a benchmark, on which firms are regulated. This benchmark is generally constructed in such a way that the cost reduction of firm i does not *directly* affect the price firm i is regulated on.[15] In our model, we assume, as in Shleifer (1985), that the benchmark for firm i is based on the

[15] Hence, if there was no strategic interaction, firms would fully benefit from cost reduction.

average of realized marginal costs of all other $N-1$ firms, that is,

$$c-\bar{u}_i = c-\frac{1}{N-1}\sum_{j\neq i} u_j.$$

where $\bar{u}_i = \frac{1}{N-1}\sum\limits_{j\neq i} u_j$.

Here again, we do not allow for lump-sum transfers. The regulator can only set a *ceiling price* for each firm i, which is based on $c-\bar{u}_i$ and guarantees additionally a mark-up captured by a parameter $\mu \geq 1$. More specifically, the *ceiling price* for firm firm i is given by

$$p_i^{\max} = \mu\left(c-\bar{u}_i\right)$$

The timing of the game is as follows:

The cost reduction stage: Once informed of the value of the mark-up parameter μ, all firms decide simultaneously and independently of their level of cost reduction.

The pricing stage: The regulator observes the realized costs $c-u_i$, $i=1,2,...,n$ and informs each firm i of its ceiling price $\mu\left(c-\bar{u}_i\right)$. Then each firm $i=1,2,...,n$ sets its price p_i subject to the regulatory constraint $p_i \leq \mu\left(c-\bar{u}_i\right)$. We assume that the firms commit to serve the demand addressed to them (hence quantity is not a strategic variable).

We determine now the game's subgame perfect Nash equilibria.

4.4.1. Cost reduction and pricing under relative regulation

The following lemma characterizes the unique equilibrium of the pricing stage given the cost reductions realized by the firms at the cost reduction stage.

Lemma 6 *The pricing stage has a unique equilibrium, which is characterized by each firm* *i playing the dominant strategy that consists of setting its price to:*

$$p_i^*(u_i, \bar{u}_i) = \min\left(\mu(c - \bar{u}_i), p^M(u_i)\right)$$

Moreover, $((u_1, p_1^*(u_1, .)), (u_2, p_2^*(u_2, .)), ..., (u_n, p_n^*(u_n, .)))$ *is a subgame perfect equilibrium of the two-stage game if and only if for each* $i = 1, 2, ..., n$ *the pair* $((u_i, p_i^*(u_i, \bar{u}_i)))$ *is a solution to the maximization program*

$$\max_{u_i \in [0,c], p_i \in [0,a]} \pi(u_i, p_i) = (p_i - c + u_i)D(p_i) - C(u_i), \tag{4.3}$$

subject to the regulatory constraint

$$p_i \leq \mu(c - \bar{u}_i),$$

Proof. See Appendix ■

Since there is no strategic interaction between the firms at the pricing stage, it is possible to describe the subgame perfect equilibrium(a) of the sequential game in a simultaneous-like way as shown in lemma (6). This makes the determination of the equilibrium(a) easier relative to the standard way of solving for SPEs using backward induction, in particular because the maximization program in this lemma is a constrained version of the unregulated monopoly's program (4.1) which has been already solved. The following lemma gives the best response function of a regulated firm when the relative regulation scheme is used.

Lemma 7 *The best response of firm i in terms of cost reduction depends on* $\bar{u}_i$ *as follows:*

$$u_i\left(\bar{u}_i\right)=\begin{cases}\left(C'\right)^{-1}oD\left(\mu\left(c-\bar{u}_i\right)\right) & \text{if} \quad \bar{u}_i>c-\frac{p^M}{\mu}\\ u^M & \text{if} \quad \bar{u}_i\leq c-\frac{p^M}{\mu}\end{cases}$$

and the optimal price of firm i given $\bar{u}_i$ *is:*

$$p_i^*\left(u_i\left(\bar{u}_i\right),\bar{u}_i\right)=\begin{cases}\mu\left(c-\bar{u}_i\right) & \text{if} \quad \bar{u}_i>c-\frac{p^M}{\mu}\\ u^M & \text{if} \quad \bar{u}_i\leq c-\frac{p^M}{\mu}\end{cases}$$

Proof. See Appendix ■

Unsurprisingly the best response of a firm i depends on the other firms' cost reductions only through the ceiling price computed on the basis of these cost reductions. Since C' is strictly increasing (see A3) then $\left(C'\right)^{-1}$ is strictly increasing which results in the strict increasingness of $u_i\left(\bar{u}_i\right)$ over$\left]c-\frac{p^M}{\mu},c\right]$. It follows that $u_i\left(\bar{u}_i\right)$ is weakly increasing over $[0,c]$. Since for any (i,j) such that $i\neq j$ an increase in u_j yields an increase of $\bar{u}_i$, it follows that an increase in u_j entails (weakly) an increase in $u_i\left(\bar{u}_i\right)$. This means that the variables u_i are strategic complements. This is an important feature of the "artificial competition" induced by relative regulation. Note that the best response function $u_i\left(\bar{u}_i\right)$ shifts downwards as the mark-up increases. The intuition behind this result is that a higher mark-up has a negative effect on demand, which decreases the marginal positive effect of cost reduction on the gross profit.

The following proposition establishes the existence and uniqueness of a subgame-perfect Nash equilibrium of the two-stage regulation game for any mark-up parameter μ.

Proposition 8 *1. If* $1 \leq \mu < \mu^M$ *then there is a unique subgame perfect Nash equilibrium to the relative regulation game. This equilibrim is symmetric and characterized by all firms reducing their costs by* $u^* (\mu)$ *defined as the unique solution in* u *to the equation* $u = (C')^{-1} oD (\mu (c - u))$ *over the interval* $\left] c - \frac{p^M}{\mu}, c \right]$, *and setting their prices to* $p^* (\mu) = \mu (c - u^* (\mu))$.

2. If $\mu \geq \mu^M$ *then there is a unique subgame perfect Nash equilibrium to the relative regulation game. This equilibrim is symmetric and chacterized by all firms behaving as unregulated monopolies, that is, reducing their costs by* u^M *and setting their prices to* p^M.

Proof. See Appendix. ■

The following proposition states that the reaction of regulated firms (in terms of cost reduction) to a decrease in the stringency of relative regulation is completely opposite to their reaction under individual regulation.

Proposition 9 *The equilibrium cost reduction under relative regulation* $u^* (\mu)$ *is strictly decreasing in the mark-up parameter* μ *over* $\left[1, \mu^M\right]$.

Proof. See Appendix. ■

The intuition behind this quite surprising result is a follows. A higher mark-up makes a firm anticipate it will be able to set a higher price. Since this will result in a lower demand, the marginal effect of cost reduction on its profit will be weaker, which decreases its incentives to carry out cost-reducing investments. The key difference with the individual regulation case is that firm's cost reduction has no direct effect on its demand which rules out the (positive) *demand effect* that appears under individual regulation. However the two other (negative) effects are still at work under relative regulation and both of them are reinforced by a higher mark-up.

Let us now turn to the effet of regulation stringency on the equilibrium price. In contrast to individual regulation, the direct and indirect effet of a higher mark-up on the price are not opposite under relative regulation. Both lead to a higher price which yields the clear-cut result stated in the following proposition.

Proposition 10 *The equilibrium price under relative regulation is strictly increasing in the mark-up parameter* μ *over* $\left[1, \mu^M\right]$.

Proof. Follows immediately from $p^*(\mu) = \mu(c - u^*(\mu))$ and proposition (9). ■

Hence, there is no tension between productive efficiency and allocative efficiency under relative regulation: a more stringent regulation policy improves both of them. This suggests that regulators should grant low mark-ups when they use relative schemes similar to the one described in this paper. However, this result has to be mitigated for at least two reasons that we present below.

4.4.2. The participation constraint

Proposition 11 *The equilibrium profit under relative regulation* $\pi(u^*(\mu), p^*(\mu))$ *is strictly increasing in the mark-up* μ *over* $\left[1, \mu^M\right]$. *Moreover, there exists a threshold* $\mu_0 \in \left]1, \mu^M\right[$ *such that* $\pi(u^*(\mu), p^*(\mu)) \geq 0$ *if and only if* $\mu \geq \mu_0$.

Proof. See Appendix ■

The regulation stringency affects negatively the regulated firms' equilibrium profits through two channels. The direct effect is that a smaller mark-up results in a smaller price which yields a decrease of the profit. But there is also an indirect effect: a smaller mark-up increases the over-investment of the regulated firms relative to their private optimal investment u^M. In this sense a smaller mark-up makes the artificial competition between the regulated firms more intense.

Proposition (11) suggests that a regulator cannot set too small a mark-up without risking to make firms not accept the regulatory contract in the first place or shut down if they were active on the market previously. Such problem would not occur if the regulator were able to make lump-sum transfers to the regulated firms.

4.4.3. Quality investment

In addition to efficiency, improving (or at least maintaining) quality is commonly seen as an important objective of regulation. To analyze the effect of a more stringent relative regulation on the incentives to invest in quality, we assume now that the regulated firms can undertake, simultaneously with their cost reduction, a quality-enhancing investment that shifts the demand upwards.

Focusing on the case of a linear demand $D(p) = a - p$, we suppose that a regulated firm can increase the demand in its market from $D(p) = a - p$ to $D(p, \theta) = a + \theta - p$ if it incurs an investment cost $\Psi(\theta) = \frac{1}{2}\lambda\theta^2$. This means that from the firms' point of view, a quality-enhancing investment "enlarges" the market they operate in: the market size increases from a to $a + \theta$. Thus, for a given $\theta \geq 0$, the analysis in the basic set-up without quality investments remains true, whenever we take into account the fact the new market size is $a + \theta$. It is then important to look at how the relevant thresholds are affected by the market size. Most importantly, it is straightforward to show that under the assumption $a < \gamma c$ (this is $D(0) < C'(c)$ in the linear-quadratic setting considered here), the monopoly replicating mark-up is given by $\mu^M(a) = \frac{(a+c)\gamma - a}{2\gamma c - a}$, which is increasing in a. This entails that for any given $\theta \geq 0$, it holds that $\mu^M(a) \leq \mu^M(a+\theta)$.

The maximization program of firm i, given the cost reduction efforts of the other firms, is the following:

$$\max_{u_i \in [0,c], p_i \in [0,a], \theta_i \in [0,+\infty[} \pi_i(u_i, p_i, \theta_i) = (p_i - c + u_i)(a + \theta_i - p_i) - \frac{1}{2}\gamma u_i^2 - \frac{1}{2}\lambda \theta_i^2$$

under the regulatory constraint

$$p_i \leq \mu \left(c - \bar{u}_i\right).$$

We focus on the values of the mark-up parameter in the range $\left[\mu_0, \mu^M(a)\right[$ and we assume hereafter that $c > 1$ and $\lambda > 1$. The latter assumption ensures that there exists a threshold $\bar{\theta}$ (depending only on a) such that the net profit $\pi_i(p_i, u_i, \theta_i)$ is smaller than $\pi_i(p_i, u_i, 0)$ for any $\theta_i \geq \bar{\theta}$, independent of the values of $\mu \in \left[\mu_0, \mu^M(a)\right[, p_i \in [c, a], u_i \in [0, c]$, $\bar{u}_i \in [0, c]$. This implies that maximizing the function $\theta_i \longrightarrow \pi_i(p_i, u_i, \theta_i)$ over the interval $\left[0, \bar{\theta}\right]$ is equivalent to maximizing it over the interval $[0, +\infty[$. Consequently, by assuming that $a + \bar{\theta} \leq \gamma c$, we are sure that we can conduct the same analysis as in the case without quality investment without risking to discard a potential maximizing value of θ_i. Since $\mu^M(a) \leq \mu^M(a+\theta)$, the assumption $\mu \in \left[\mu_0, \mu^M(a)\right[$ entails that for any given $\theta \geq 0$, we have: $\mu \leq \mu^M(a+\theta)$. Then we derive from the analysis of the case without quality investment that the regulatory constraint is binding and that the solution to the program $\max_{u_i \in [0,c], p_i \in [0,a]} \pi_i(p_i, u_i, \theta_i)$ is given by:

$$u_i\left(\bar{u}_i, \theta_i\right) = \frac{1}{\gamma}(a + \theta_i - \mu c + \mu \bar{u}_i)$$

$$p_i\left(\bar{u}_i\right) = \mu\left(c - \bar{u}_i\right).$$

At this point we need to maximize $\pi_i(p_i(\theta_i, \bar{u}_i), u_i(\theta_i, \bar{u}_i), \theta_i)$ with respect to θ_i. Using the first-order condition, we get:

$$\theta_i(\bar{u}_i) = \frac{\gamma c(\mu - 1) + a - \mu c - (\gamma - 1)\mu\bar{u}_i}{\lambda\gamma - 1},$$

which leads to the investment best response function of firm i :

$$u_i(\bar{u}_i) = u_i(\bar{u}_i, \theta_i(\bar{u}_i)) = \frac{a\lambda - c - \mu c(\lambda - 1) + \mu(\lambda - 1)\bar{u}_i}{\lambda\gamma - 1}.$$

Solving these N equations, it is straightforward to show that the unique Nash equilibrium is symmetric and characterized by:

$$\hat{u}(\mu, \gamma, \lambda) = \frac{\lambda(a - \mu c) + c(\mu - 1)}{\lambda(\gamma - \mu) + \mu - 1}$$

$$\hat{\theta}(\mu, \gamma, \lambda) = \frac{(\mu - 1)(\gamma c - a)}{\lambda(\gamma - \mu) + \mu - 1}$$

$$\hat{p}(\mu, \gamma, \lambda) = \mu\frac{\lambda(\gamma c - a) + c - 1}{\lambda(\gamma - \mu) + \mu - 1}.$$

From the above expressions of $\hat{u}(\mu, \gamma, \lambda)$, $\hat{\theta}(\mu, \gamma, \lambda)$ and $\hat{p}(\mu, \gamma, \lambda)$, it is straightforward to derive the following proposition that shows how the equilibrium cost reducion, price and quality are affected by the mark-up parameter μ and the cost parameters γ and λ.

Proposition 12 *1. The equilibrium cost reduction $\hat{u}(\mu, \gamma, \lambda)$ is decreasing in the mark-up parameter μ, and in the investment cost parameters γ and λ.*

2. The equilibrium quality $\hat{\theta}(\mu, \gamma, \lambda)$ is increasing in the mark-up parameter μ and the cost-reducing investment cost parameter γ and is decreasing in the quality investment cost parameter λ.

3. The equilibrium price $\hat{p}(\mu, \gamma, \lambda)$ is increasing in the the mark-up parameter μ and in the investment cost parameters γ and λ.

This proposition shows that the cost-reduction (respectively the price) is decreasing (respectively increasing) in the mark-up like in the case without quality investment. Hence the result that a more lenient regulation policy, i.e. a higher mark-up, affects negatively the investment in cost reduction holds. However, an increase in the mark-up value affects positively the investment in quality. This is partly due to the fact that the firms are regulated only with respect to their costs. Under this assumption, the marginal benefit of enlarging the demand through higher quality is clearly increasing in the mark-up.
The previous proposition also states that the more costly the investment in quality the lower the cost reduction realized by the firms. The logic behind this result is as follows: a more costly investment in quality induces less quality enhancement which entails a lower market size, which leads to lower cost reduction incentives. Furthermore, we showed that the investment effort in quality is increasing in the cost of efficiency-improving investments. Hence, when cost reduction becomes costlier, the regulated firms have an incentive to shift some of their investment from cost reduction to quality improvement. Finally, note that the previous analysis can be applied to any demand-enhancing investment and not only to quality investments.

4.5. Conclusion

This paper aims to shed some light on how a more lenient regulation policy could affect some potentially conflicting objectives of a regulator. More specifically, it investigates the effect of regulation stringency on productive and allocative efficiency under individual and relative regulation when the regulator cannot make transfers to firms.

One of the messages of this paper is about the consequences of granting high mark-ups to regulated firms. In the case of individual regulation, the conventional wisdom that a higher mark-up leads to higher investment is true and may serve as a justification for granting high mark-ups, particularly if the regulator is mainly concerned with encouraging cost reduction investments. This justification cannot be used when firms are regulated using relative performance schemes since in this case, high mark-ups entail not only high prices but weak cost reductions as well. This is due to the fact that the potential tension between allocative and productive efficiency under individual regulation does not appear under its relative counterpart, making it socially desirable to grant relatively small mark-ups, which means using a quite stringent relative regulation.

However, the regulator has to take into account the effect of regulation stringency on the regulated firms' participation constraint and incentives to conduct quality-improving investments. Indeed, a very stringent regulation makes the artificial competition under relative performance schemes very intense, which may result in negative profits for the regulated firms. Anticipating that they will not be able to break-even, the firms will not accept the regulatory contract in the first place. Moreover, quality-improving investments, and more generally any demand-enhancing investments, are negatively affected by the stringency of relative regulation. Therefore, if improving quality has an important weight in the regulator's objective, she might want to provide more lenient regulatory schemes in order to encourage investments in quality even if this comes at the expense of lower investments in cost reduction.

4.6. References

Auriol, E., 2000, "Concurrence par comparaison: Un point de vue normatif", *Revue Economique*, 51, 621-634.

Buehler, S., Gaertner, D., and Halbheer, D., 2006, "Deregulating network industries: Dealing with price-quality tradeoffs", *Journal of Regulatory Economics*, 30, 99-115.

Buehler, S., Schmutzler, A., and benz, M. A., 2004 : "Infrastructure quality in deregulated industries: Is there an underinvestment problem?", *International Journal of Industrial Organization*, 22, 253-267.

Dalen, D. M., 1998, "Yardstick competition and investment incentives", *Journal of Economics & Management Strategy* , 7, 105-126.

Dranove, D., 1987, "Rate-setting by diagnosis related groups and hospital specialization", *Rand Journal of Economics*, 18, 417-27.

Ebel, N. and Lefouili, Y., 2008, "Individual vs relative regulation: an investigation of investment incentives under a cost-plus approach", mimeo.

Farsi, M., Fetz, A., and Filippini, M., 2007, "Benchmarking and regulation in the electricity distribution sector", *CEPE Working Paper*, No. 54.

Federal Communications Commission (FCC), 1997, "Price cap performance review for local exchange carriers", FCC.

Jamasb, T., and Pollitt, M., 2001, "Benchmarking and regulation: international electricity experience", *Utilities Policy*, 9, 107-130.

Joskow, P. L., 2006, "Regulation of the electricity market", *CESifo DICE Report*, 2, 3-9.

Laffont, J.-J., and Martimort, D., 2000, "Mechanism design with collusion and correlation", *Econometrica*, 68, 309-342.

Laffont, J.-J., and Tirole, J., 1993, *A Theory of incentives in procurement and regulation*, MIT Press, Cambridge.

Powell, K., and Szymanski, S., 1997, "Regulation through comparative performance evaluation", *Utilities Policy*, 6, 293-301.

Shleifer, A., 1985, "A theory of yardstick competition", *Rand Journal of Economics*, 16,319–327.

Sobel, J. 1999, "A reexamination of yardstick competition," *Journal of Economics & Management Strategy*, 8, 33-60.

Suzuki, A., 2008, "Yardstick competition to elicit private information: an empirical analysis of the Japanese gas distribution industry", ISER Discussion Paper No. 709.

Tangerås, T. P., 2002, "Collusion-proof yardstick competition", *Journal of Public Economics*, 83, 231–254.

4.7. Appendix

Proof of lemma 1:

Note first that for any $u \in [0, c]$ and $p \in [c-u, a[$, we have $\frac{\partial^2 \pi}{\partial p^2} = 2D'(p) + (p-c+u)\, D"(p) < 0$. This is obviously true if $D"(p) < 0$ but it also holds if $D"(p) \geq 0$. Indeed, in the latter case rewritting $\frac{\partial^2 \pi}{\partial p^2}$ as $\frac{\partial^2 \pi}{\partial p^2} = [2D'(p) + pD"(p)] - (c-u)\, D"(p)$ allows to see that $\frac{\partial^2 \pi}{\partial p^2}$ remains negative in this case as well because $2D'(p) + pD"(p) \leq 0$ (see A2). Furthermore, the first derivative $\frac{\partial \pi}{\partial p} = D'(p)\,(p-c+u) + D(p)$ goes to a strictly negative value as p goes to a whereas it is strictly positive for $p = c-u$. This ensures that for any $u \in [0, c]$

there exists a unique price $p^M(u)$ that maximizes in p the profit function $\pi^M(u,p)$ and this price is the unique solution in p of the FOC $\frac{\partial \pi}{\partial p}=0$. Let us now show that $p^M(u)$ is strictly decreasing in u. Using the implicit function theorem we can state that $p^M(u)$ is differentiable and $\frac{d}{du}p^M(u)=\frac{-\left(1+p^M(u)\right)D'\left(p^M(u)\right)}{\frac{\partial^2\pi}{\partial p^2}(p^M(u),u)}$ which is strictly negative since the numerator is strictly positive and the denominator is strictly negative. Moreover we can show that $p^M(u)$ is either strictly convex or linear in u. Indeed the FOC definining $p^M(u)$ can be rewritten as $u=c-p+\frac{D(p)}{D'(p)}$ which yields $\left(p^M\right)^{-1}(p)=c-p+\frac{D(p)}{D'(p)}$. Then assumption A6 entails that $\left(p^M\right)^{-1}$ is either linear or strictly convex. Combining this with the fact that p^M is decreasing, we conclude that p^M is either strictly convex or linear.

Proof of proposition 2:

The result that $p^M(u)$ is decreasing in u (see lemma 1) allows to state that $p^M(u)\leq p^M(0)$ for all $u\in[0,c]$. Thus maximizing $\pi^M(u,p)$ over $[0,c]\times[0,+\infty[$ is equivalent to maximizing it over $[0,c]\times\left[0,p^M(0)\right]$. Now we are sure that $\pi^M(p,u)$ has at least one global maximum because $[0,c]\times\left[0,p^M(0)\right]$ is a compact set and the function $\pi^M(.,.)$ is continuous over this set. Let us show that such a global maximum is necessarily reached at an interior point. It is clear that it cannot be reached at a point (u,p) such that $p=0$. Furthermore, $\frac{\partial \pi}{\partial u}(c,p)=D(p)-C'(c)\leq D(0)-C'(c)<0$ which entails that a maximum cannot be reached at a point (u,p) such that $u=c$. Finally, a maximum cannot be reached at a point (u,p) such that $u=0$: otherwise, it would be reached at the point $\left(0,p^M(0)\right)$ but this is impossible because $\frac{\partial \pi}{\partial u}\left(0,p^M(0)\right)=D\left(p^M(0)\right)-C'(0)>0$ (by assumption A4) which shows that for sufficiently small positive values of u the value of $\pi^M(u,p^M(0))$ is greater than $\pi^M(0,p^M(0))$. We now know that the global maximum(a) is (are) necessarily interior and thus characterized by the FOCs

$$\begin{cases} D(p) + (p - c + u)D'(p) = 0 \\ D(p) - C'(u) = 0 \end{cases}$$

which we rewrite as:

$$\begin{cases} u = \left(p^M\right)^{-1}(p) \\ u = (C')^{-1} oD(p) \end{cases}$$

We know that the curves of the two functions $p \to \left(p^M\right)^{-1}(p)$ and $p \to (C')^{-1} oD(p)$ meet at least once (since we showed the existence of an interior maximum). Furthermore note that both functions are decreasing in p, the first one is either strictly convex or linear (see the proof of proposition 1) and the second one is either strictly concave or linear (see A5). If we had strict convexity and strict concavity we could directly state that their curves meet at most twice. However, even with the weaker properties we have on those functions, this will hold. Indeed, the only case where they can meet in more than two points is the special case where they are both linear and identical. The following analysis will make it clear that this is impossible due to A4 and will show that the curves cannot meet twice neither in the (relevant) domain defined by $u \geq 0$ and $p \geq 0$. The assuption A4 states that $D\left(p^M(0)\right) > C'(0)$. Along with A1, this yields $p^M(0) < D^{-1}oC'(0)$ which means that the curve of the first (either linear or strictly convex) function meets the horizontal axis defined by $u = 0$ in a point on the left of the point where the curve of the second (strictly concave or linear) function meets this axis. Given the shapes of the two functions, this would not hold if the curves met twice in the domain defined by $u \geq 0$ and $p \geq 0$, or if they were identical. We can then rule out both those cases and state that the curves of the functions $p \to \left(p^M\right)^{-1}(p)$ and $p \to (C')^{-1} oD(p)$ meet only once. Thus, we can state the uniqueness of the solution to the maximization program (4.1) which we denote by $\left(u^M, p^M\right)$ instead of $\left(u^M, p^M\left(u^M\right)\right)$. Note that we also showed that $\pi(u,p)$ has no local maximum over $[0,c] \times [0,+\infty[$ but its unique global maximum.

Proof of proposition 3:

Note that the existence of at least one solution to the maximization program (4.2) is easily derived from the fact that the profit function is continuous in both its arguments and the (constrained) domain over which it is maximized is compact.

We first deal with the simpler case, that is $\mu \geq \mu^M$. Under this condition, the pair $\left(u^M, p^M\right)$ satisfies the pricing constraint $p \leq \mu(c-u)$. Since $\left(u^M, p^M\right)$ is the unique solution to the unconstrained maximization program (4.1), we can state that in case $\mu \geq \mu^M$ it is as well the unique solution to the maximization program (4.2) (the constraint is not binding except for the special case $\mu = \mu^M$).

Assume now that $1 \leq \mu < \mu^M$. In this case, $\left(u^M, p^M\right)$ does not fulfill the pricing constraint $p \leq \mu(c-u)$. Since we know from the proof of proposition 2 that $\pi(u,p)$ has no local maximum over $[0,c] \times [0,+\infty[$ but its unique global maximum, the pricing constraint is necessarily binding. The maximization program (4.2) amounts then to maximizing $\pi(u,\mu(c-u)) = (\mu-1)(c-u)D(\mu(c-u)) - C(u)$ with respect to u over $[0,c]$. The strict concavity of this function with respect to u results from the strict concavity of the function $p \to pD(p)$ (asumption A2), the linearity of $u \to \mu(c-u)$ and the strict convexity of C (assumption A3). Thus we can state that $u \to \pi(u,\mu(c-u))$ reaches its maximum over $[0,c]$ at a unique point $u^I(\mu)$ which results in a unique optimal price $p^I(\mu) = \mu(c-u^I(\mu))$. To determine whether $u^I(\mu) = 0$ or $u^I(\mu) \neq 0$, it is sufficient to compare the (total) derivative of $\pi(u,\mu(c-u))$ with respect to u computed for $u = 0$ with 0. It is straightforward to show that this derivative is $-(\mu-1)[D(\mu c)+\mu c D'(\mu c)]-C'(0)$. Therefore $u^I(\mu)=0$ if and only if $C'(0)\geq-(\mu-1)[D(\mu c)+\mu c D'(\mu c)]$. Consider the set:

$$A=\left\{\mu\in\left[1,\mu^M\right]\ /\ u^I(\mu)=0\right\}=\left\{\mu\in\left[1,\mu^M\right]\ /\ C'(0)\geq-(\mu-1)\left[D(\mu c)+\mu c D'(\mu c)\right]\right\}$$

and denote $\tilde{\mu}=\sup A$ (A is not empty since $1\in A$). Note first that $\tilde{\mu}\in A$ (due to the continuity of D) and that $\tilde{\mu}<\mu^M$ since $u^I(\mu^M)=u^M>0$. To establish that $A=[1,\tilde{\mu}]$, it is sufficient to show that if μ_1 is in the interval $\left[1,\mu^M\right]$ but does not belong to A, then all elements of the interval $\left[\mu_{1,}\mu^M\right]$ do not belong to A. To show this we use the following result: in each point where the function $\mu\rightarrow-(\mu-1)[D(\mu c)+\mu c D'(\mu c)]$ is positive, its derivative is positive. Indeed the derivative of this function with respect to μ is given by $-[D(\mu c)+\mu c D'(\mu c)]-(\mu-1)c[2D'(\mu c)+\mu c D"(\mu c)]$ which is positive whenever the term $-[D(\mu c)+\mu c D'(\mu c)]$ is positive since the remaining term $-(\mu-1)c[2D'(\mu c)+\mu c D"(\mu c)]$ is always positive due to assumption A2 (that entails that $2D'(\mu c)+\mu c D"(\mu c)<0$ because $2D'(p)+pD''(p)$ is the second derivative of $pD(p)$). This result has the following straightforward implication: whenever the function $\mu\rightarrow-(\mu-1)[D(\mu c)+\mu c D'(\mu c)]$ reaches a positive value at some point, say μ_1, it remains above this positive value for all $\mu\geq\mu_1$. Consider now $\mu_1$such that $\mu_1\in\left[1,\mu^M\right]$ and $\mu_1\notin A$. The latter condition implies that the function $-(\mu-1)[D(\mu c)+\mu c D'(\mu c)]$ is above $C'(0)\geq 0$ at the point $\mu=\mu_1$. Using the previous result it follows that $-(\mu-1)[D(\mu c)+\mu c D'(\mu c)]>C'(0)$ for all $\mu\in\left[\mu_{1,}\mu^M\right]$ which means that all the elements of the interval $\left[\mu_{1,}\mu^M\right]$ do not belong to A. This is sufficient to state that μ_1 cannot be weakly smaller than $\tilde{\mu}$ since this would entail that $\tilde{\mu}\notin A$. Hence $[1,\tilde{\mu}]\subset A$. Since the reverse inclusion holds as well, it follows that $A=[1,\tilde{\mu}]$ which means that $u^I(\mu)=0\Leftrightarrow\mu\in[1,\tilde{\mu}]$.

Let us now show that $\tilde{\mu} = 1$ if and only if $C'(0) = 0$ and $D(c) + cD'(c) \leq 0$.

Suppose that $\tilde{\mu} = 1$.Then $C'(0) < -(\mu - 1)[D(\mu c) + \mu c D'(\mu c)]$ for any $\mu \in \left]1, \mu^M\right]$. Since the right-hand side of the latter inequality goes to 0 as μ goes to 1, it must hold that $C'(0) \leq 0$ which results in $C'(0) = 0$ since the reverse (weak) inequality is true as well (see A3). Moreover from the inequality $C'(0) < -(\mu - 1)[D(\mu c) + \mu c D'(\mu c)]$ we derive that $D(\mu c) + \mu c D'(\mu c) < 0$ for any $\mu \in \left]1, \mu^M\right]$ which entails that $D(c) + cD'(c) \leq 0$ since the left-hand side of the latter inequality goes to $D(c) + cD'(c)$ as μ goes to 1.

Conversely suppose that $C'(0) = 0$ and $D(c) + cD'(c) \leq 0$. Since $p \to D(p) + pD'(p)$ is decreasing over $[0, a[$ (because of A2) and $c < a$, it is true that $D(\mu c) + \mu c D'(\mu c) < D(c) + cD'(c) \leq 0$ for all $\mu \in \left]1, \mu^M\right]$ which yields $-(\mu - 1)[D(\mu c) + \mu c D'(\mu c)] < 0 = C'(0)$ for all $\mu \in \left]1, \mu^M\right]$. It follows that $u^I(\mu) > 0$ for all $\mu \in \left]1, \mu^M\right]$. This entails that $\tilde{\mu} = 1$.

Proof of proposition 4:

Denote $g(u, \mu) = \pi(u, \mu(c - u))$ and let $\mu \in \left]\tilde{\mu}, \mu^M\right[$. For such a μ the optimal cost reduction $u^I(\mu)$ is defined as the unique solution in u to the FOC, that is $\frac{\partial g}{\partial u}(u^I(\mu), \mu) = 0$. Differentiating this equation with respect to μ we get that:

$$\frac{\partial g}{\partial \mu \partial u}(u^I(\mu), \mu) + \frac{\partial^2 g}{\partial u^2}(u^I(\mu), \mu).\frac{du^I}{d\mu} = 0$$

which yields

$$\frac{du^I}{d\mu} = -\frac{\frac{\partial g}{\partial \mu \partial u}(u^I(\mu), \mu)}{\frac{\partial^2 g}{\partial u^2}(u^I(\mu), \mu)}$$

We have:

$$\frac{\partial g}{\partial \mu \partial u} = -\left[D\left(\mu\left(c-u\right)\right)+\mu\left(c-u\right)D'\left(\mu\left(c-u\right)\right)\right] \\ -(\mu-1)\left(c-u\right)\left[2D'\left(\mu\left(c-u\right)\right)+\mu\left(c-u\right)D''\left(\mu\left(c-u\right)\right)\right]$$

Using the FOC $\frac{\partial g}{\partial u}(u^I(\mu),\mu)=0=-(\mu-1)\left[D\left(\mu\left(c-u\right)\right)+\mu\left(c-u\right)D'\left(\mu\left(c-u\right)\right)\right]-C'(u)$ we find that

$$\frac{\partial g}{\partial \mu \partial u}(u^I(\mu),\mu) = \underbrace{\frac{C'(u^I(\mu))}{\mu-1}}_{>0} \\ -\underbrace{(\mu-1)\left(c-u^I(\mu)\right)}_{>0}\underbrace{\left[2D'\left(\mu\left(c-u^I(\mu)\right)\right)+\mu\left(c-u^I(\mu)\right)D''\left(\mu\left(c-u^I(\mu)\right)\right)\right]}_{<0 \text{ because this is the value of } \frac{d^2}{dp^2}(pD(p)) \text{ at } p=\mu(c-u^I(\mu))}$$

Thus, the cross derivative $\frac{\partial g}{\partial \mu \partial u}(u^I(\mu),\mu)$ is strictly positive. Furthermore, $\frac{\partial^2 g}{\partial u^2}(u^I(\mu),\mu)<0$ because g is strictly concave with respect to u (even if g were not concave the inequality would hold because the function $g(.,\mu)$ reaches an interior maximum at $u^I(\mu)$). We now can state that $\frac{du^I}{d\mu}>0$ which entails that $u^I(\mu)$ is strictly increasing in μ over $\left]\tilde{\mu},\mu^M\right[$.

Proof of proposition 5:

The results that $p^I(\mu)$ is strictly increasing in μ over $[1,\tilde{\mu}]$ and constant over $\left[\mu^M,+\infty\right[$ are straighforward corollaries of proposition (4).

Assume now that $\mu\in\left]\tilde{\mu},\mu^M\right[$. We showed in the proof of proposition (3) that in this case the regulatory constraint is binding and that solving the maximization program (4.2) (in the two variables u and p) amounts to solving a maximization program in the variable u only by replacing p in the profit function by its constrained value $\mu\left(c-u\right)$. Likewise, it also amounts to solving the following maximization program in p only by replacing u in the profit function by $c - \frac{p}{\mu}$:

$$\max_{p\in[0,\mu c]} \pi\left(c - \frac{p}{\mu}, p\right) = \left(1 - \frac{1}{\mu}\right) pD(p) - C\left(c - \frac{p}{\mu}\right)$$

Since $u^I(\mu)$ is the unique solution in u to the FOC of the maximization program written as a function of u, then $p^I(\mu) = \mu\left(c - u^I(\mu)\right)$ is the unique solution to the FOC of the maximization program written as a function of p:

$$\left(1 - \frac{1}{\mu}\right)\left[pD'(p) + D(p)\right] + \frac{1}{\mu}C'\left(c - \frac{p}{\mu}\right) = 0$$

Denoting $\eta = \frac{1}{\mu}$ and $p^I(\eta)$ instead of $p^I(\mu)$, it follows that:

$$(1-\eta)\left[p^I(\eta) D'\left(p^I(\eta)\right) + D\left(p^I(\eta)\right)\right] + \eta C'\left(c - \eta p^I(\eta)\right) = 0$$

Differentiating this equation with respect to η and writing p^I instead of $p^I(\eta)$ we get that:

$$\frac{dp^I}{d\eta} = \frac{\left[p^I D'\left(p^I\right) + D\left(p^I\right)\right] + \left[\eta p^I C\text{''}(c - \eta p^I) - C'\left(c - \eta p^I\right)\right]}{2D'(p^I) + p^I D''(p^I) - (p^I)^2 C\text{''}(c - \eta p^I)}$$

The denominator is strictly negative since it is the value at $p = p^I$ of the second derivative of a strictly concave function. The first term of the numerator $p^I D'\left(p^I\right) + D\left(p^I\right)$ is strictly negative. Indeed using the FOC, it is equal to $-\frac{\eta C'\left(c - \eta p^I\right)}{(1-\eta)} < 0$. Thus a sufficient condition for p^I to be strictly increasing in η and thus decreasing in μ is that the remaining term in the numerator is weakly negative. This is true if $uC\text{''}(c-u) \leq C'(c-u)$ for all $u \in [0, c]$ or equivalently if $(c-u)\,C\text{''}(u) \leq C'(u)$ for all $u \in [0, c]$.

If the investment cost function is quadratic, i.e. $C(u) = \frac{1}{2}\gamma u^2$, the latter condition does not hold. In this case, under a linear demand $D(p) = a - p$ we find that the cost reduction under individual regulation is $u^I(\mu) = \frac{(2\mu c-a)(\mu-1)}{2\mu(\mu-1)+\gamma}$ (for $\mu \in \left]\tilde{\mu}, \mu^M\right[$) which yields $p^I(\mu) = \frac{a\mu(\mu-1)+\mu\gamma c}{2\mu(\mu-1)+\gamma}$. Using again the variable $\eta = \frac{1}{\mu}$, the latter expression can rewritten as $p^I(\eta) = \frac{a-\eta(a-\gamma c)}{2-2\eta+\gamma\eta^2}$. Some tedious computations lead to $\frac{dp^I}{d\eta} = \frac{\eta^2\gamma(a-\gamma c)+2\gamma(c-a\eta)}{(2-2\eta+\gamma\eta^2)^2}$. Note first that the monopoly-replicating mark-up in this case is given by $\mu^M = \frac{(a+c)\gamma-a}{2\gamma c-a}$ which results in $\mu^M - \frac{a}{c} = \frac{(a-c)(a-\gamma c)}{c(2\gamma c-a)}$. Note also that the assumption $D(0) < C'(c)$ is equivalent in the linear-quadratic model to $a < \gamma c$ which allows to state that $\mu^M < \frac{a}{c}$. This entails that $c - \frac{a}{\mu} < 0$ (or equivalently that $c - a\eta < 0$) for any $\mu \in \left]\tilde{\mu}, \mu^M\right[$. Using this inequality and $a < \gamma c$ we can now state that $\frac{dp^I}{d\eta} < 0$ for any $\eta \in \left]\frac{1}{\mu^M}, \frac{1}{\tilde{\mu}}\right[$. Hence, p^I is decreasing in η and thus strictly increasing in μ.

Proof of lemma 6:

Note first that a firm's profit is not affected by the price set by the other firms. The maximization program of firm i in the pricing stage is:

$$\max_{p_i \in [0,a]} \pi(u_i, p_i) = (p_i - c + u_i)D(p_i) - C(u_i) \tag{4.4}$$

subject to the regulatory constraint

$$p_i \leq \mu(c - \bar{u}_i)$$

We showed in the proof of lemma 1 that $\pi(p_i, u_i)$ is strictly concave in p_i over $[0, a]$. Then the solution to this maximization program is $\min\left(\mu(c - \bar{u}_i), p^M(u_i)\right)$. Therefore we can state that the unique equilibrium of the pricing stage is characterized by each firm i setting its price to $p_i^*(u_i, \bar{u}_i) = \min\left(\mu(c - \bar{u}_i), p^M(u_i)\right)$, which is a dominant strategy.

Proof of lemma 7:

By definition, $p_i^*(u_i, \bar{u}_i)$ is the equilibrium price of firm i given its own cost reduction and the average cost reduction of the other firms in the first stage. Then $((u_1, p_1^*(u_1, .)), (u_2, p_2^*(u_1, .)), ..., (u_n, p_n^*(u_n, .)))$ is a subgame perfect equilibrium if and only if for each i, u_i is a solution to the maximization program $\max_{u_i \in [0,c]} \pi(p_i^*(u_i, \bar{u}_i), u_i)$. Moreover $p_i^*(u_i, \bar{u}_i)$ is a solution to the maximization program (4.4) then u_i is a solution to the maximization program $\max_{u_i \in [0,c]} \pi(p_i^*(u_i, \bar{u}_i), u_i)$ if and only if $((u_i, p_i^*(u_i, \bar{u}_i)))$ is a solution to the maximization program (4.3).

Proof of proposition 8:

If $\bar{u}_i \leq c - \frac{p^M}{\mu}$ then $p^M \leq \mu(c - \bar{u}_i)$, that is, (u^M, p^M) satisfies the relative regulatory constraint. Given this, the solution to the constrained maximization program (4.3) is the same as the the solution to the unconstrained maximization program (4.1), that is (u^M, p^M).

If $\bar{u}_i > c - \frac{p^M}{\mu}$ then $p^M > \mu(c - \bar{u}_i)$, that is, (u^M, p^M) does not satisfy the regulatory constraint. First, the maximization program (4.3) has a solution because the objective function is continuous and the domain of maximization is compact. Second, we showed in the proof of proposition 1 that the function $\pi(u, p)$ has no local maximum over $[0, c] \times [0, +\infty[$ but its unique global maximum. Therefore we are sure that the constraint is binding. It follows that solving for the maximization program (4.3) amounts to solving the maximization program:

$$\max_{u_i \in [0,c]} \pi_i(u_i, \mu(c - \bar{u}_i)) = (\mu(c - \bar{u}_i) - c + u_i) D(\mu(c - \bar{u}_i)) - C(u_i)$$

The function $w : u_i \to \pi_i(u_i, \mu(c - \bar{u}_i))$ is strictly concave (its second derivative is $-C''(u_i) <$

0). Futhermore, $w'(0) = D(\mu(c - \bar{u}_i)) - C'(0) > D(p^M) - C'(0) > 0$ (the first inequality results from $p^M > \mu(c - \bar{u}_i)$ and the second one from assumption A4), and $w'(c) = D(0) - C'(c) < 0$ (see A4). We can conclude that the unique solution to the previous maximization program is interior and thus characterized by the FOC: $D(\mu(c - \bar{u}_i)) = C'(u_i)$ which we rewrite as $u_i = (C')^{-1} oD(\mu(c - \bar{u}_i))$

Proof of proposition 9:

Let us first establish the following preliminary result:

Preliminary lemma: Let $\alpha \in]0, c]$ and let $g : [0, c] \to [0, c]$ be a strictly increasing and differentiable function. Assume further that g is either strictly concave or linear, and $g(c) < c$. The following holds:

1- If $g(\alpha) \leq \alpha$ then $g(u) < u$ for all $u \in]\alpha, c]$.

2- If $g(\alpha) > \alpha$ then there exists a unique solution $u^* \in]\alpha, c[$ to the equation $g(u) = u$. Moreover $g(u) > u$ for all $u \in [\alpha, u^*[$ and $g(u) < u$ for all $u \in]u^*, c]$.

Proof of the preliminary lemma:

Let us first deal with a linear g. The linearity of g implies that for all $u \in]\alpha, c]$:

$$g(u) - u = g(\alpha) + (u - \alpha) . \frac{g(c) - g(0)}{c} - u$$

which is strictly decreasing in u because $\frac{g(c)-g(0)}{c} \leq \frac{g(c)}{c} < 1$. Thus,

- if $g(\alpha) - \alpha \leq 0$, then $g(u) - u < 0$ for all $u \in]\alpha, c]$.

- if $g(\alpha) - \alpha > 0$, then using the fact that $u \to g(u) - u$ is continuous and strictly decreasing, $g(c) - c < 0$, we state that: first, there exists a unique $u^* \in]\alpha, c[$ to the equation $g(u) = u$ (we apply the intermediate value theorem) and second, $g(u) > u$ for all $u \in [\alpha, u^*[$ and $g(u) < u$ for all $u \in]u^*, c]$.

We now tackle the case of a strictly concave g.

-If $g(\alpha) \leq \alpha$, we derive from the strict concavity of g that $g'(\alpha) < \frac{g(\alpha)-g(0)}{\alpha} \leq \frac{g(\alpha)}{\alpha} \leq 1$. Using again the strict concavity of g we state that $g'(t) \leq g'(\alpha) < 1$ for any $t \in [\alpha, c]$ which yields $g(u) < g(\alpha) + (u - \alpha) \leq u$ for all $u \in]\alpha, c]$.

-If $g(\alpha) > \alpha$, we state that there exists at least one solution to the equation $g(u) = u$ over $]\alpha, c[$. Indeed the function $g(u) - u$ is continuous, takes a strictly negative value at $u = \alpha$ and a strictly positive value at $u = c$, which entails that it is equal to 0 at least in one point of the interval $]\alpha, c[$ (we apply the intermediate value theorem). To show that this point is unique, it is sufficient to show that in a point u^*such that $g(u^*) = u^*$, we necessarily have $g'(u^*) < 1$, which will ensure that $g(u) - u$ remains strictly negative after it reaches 0 (by a similar reasoning to that of the case $g(\alpha) \leq \alpha$). As g is strictly concave, we have: $g'(u^*) < \frac{g(u^*)-g(\alpha)}{u^*-\alpha} = \frac{u^*-g(\alpha)}{u^*-\alpha} < 1$ which establishes the unicity of u^*. Since g is continuous, $g(\alpha) > \alpha$ (resp. $g(c) < c$) and $g(u) \neq u$ for all $u \in]\alpha, u^*[$ (resp. $u \in]u^*, c]$), then $g(u) > u$ (resp. $g(u) < u$) for all $u \in]\alpha, u^*[$ (resp. $u \in]u^*, c]$). QED.

Let us apply the previous lemma to $g : u \to (C')^{-1} oD(\mu(c-u))$ and $\alpha = c - \frac{p^M}{\mu}$. Note that $g\left(c - \frac{p^M}{\mu}\right) = u^M$ which yields $g\left(c - \frac{p^M}{\mu}\right) \leq c - \frac{p^M}{\mu} \Leftrightarrow \mu \geq \frac{p^M}{c-u^M} = \mu^M$. Hence, we have to distinguish two cases:

Case 1: $\mu < \mu^M$

In this case, the equation $u = (C')^{-1} oD(\mu(c-u))$ has a unique solution over $\left]c - \frac{p^M}{\mu}, c\right]$ that we denote by $u^*(\mu)$. Since the best response of firm i to all other firms reducing their costs by u^* is $u_i(u^*) = u^*$, all firms reducing their costs by $u^*(\mu)$ and setting their prices to $p^*(\mu) = \mu(c - u^*(\mu))$ is an equilibrium. Let us show that this is the unique equilibrium. To do so, consider a Nash equilibrium and denote $(u_1^*, u_2^*, ..., u_n^*)$ the cost reductions realized by the firms $i = 1, 2, ..., n$ in this equilibrium.

First note that $u_i^* > c - \frac{p^M}{\mu}$ for all i. Indeed the best response function $u_i(.)$ takes values in the interval $\left[u^M, (C')^{-1} oD(0)\right]$ which entails that $u_i^* \geq u^M$. Under the assumption $\mu < \mu^M$, it holds that $u^M > c - \frac{p^M}{\mu}$ which yields $u_i^* > c - \frac{p^M}{\mu}$. Hence, we can state that $u_i^* = (C')^{-1} oD(\mu(c - \bar{u}_i^*)) = g(\bar{u}_i^*)$.

Second let us show that $u_1^* = u_2^* = ... = u_n^* = u^*(\mu)$. To establish this result it is sufficient to show that $u_{\max}^* \leq u^*(\mu) \leq u_{\min}^*$ where $u_{\min}^* = \min u_i^*$ and $u_{\max}^* = \max u_i^*$. On the one hand, since $u_{\max}^* \geq \bar{u}_{\max}^*$, $u_{\max}^* = g(\bar{u}_{\max}^*)$, and g is strictly increasing, we have: $g(u_{\max}^*) \geq u_{\max}^*$. Using the preliminary lemma, we can state then that $u_{\max}^* \leq u^*(\mu)$. On the other hand, since $u_{\min}^* \leq \bar{u}_{\min}^*$, $u_{\min}^* = g(\bar{u}_{\min}^*)$, and g is strictly increasing, it follows that $g(u_{\min}^*) \leq u_{\min}^*$. Using here again the preliminary lemma, we derive that $u_{\min}^* \geq u^*(\mu)$. Hence we obtain $u_{\max}^* \leq u^*(\mu) \leq u_{\min}^*$. This conludes the proof under case 1.

Case 2: $\mu \geq \mu^M$

In this case, $u^M \leq c - \frac{p^M}{\mu}$. Using lemma (7), we get that $u_i\left(u^M\right) = u^M$ which is sufficient to state all firms investing reducing their costs by u^M and setting their prices to p^M is a Nash equilibrium. Let us now show that this is the unique Nash equilibrium of the relative regulation game.

Note first that it is obviously the unique equilibrium in which firms reduce their costs by $u_i \leq c - \frac{p^M}{\mu}$.

Now let us show *ad absurdum* that there is no equilibrium in which firms reduce their cost by $u_i > c - \frac{p^M}{\mu}$. Suppose that such an equilibrium exists and denote $(u_1^*, u_2^*, ..., u_n^*)$ the cost reductions realized in this equilibrium, and $u_{\min}^* = \min u_i^*$ and $u_{\max}^* = \max u_i^*$. Since $g\left(c - \frac{p^M}{\mu}\right) \leq c - \frac{p^M}{\mu}$ because $\mu \geq \mu^M$, then applying part 1 of the preliminary lemma we obtain that $g(u_i^*) < u_i^*$ for all $i = 1, 2, ..., n$. In particular this implies that $g(u_{\max}^*) < u_{\max}^*$.

Since g is strictly increasing and $\bar{u}^*_{\max} \leq u^*_{\max}$ we know that $g\left(\bar{u}^*_{\max}\right) \leq g\left(u^*_{\max}\right)$. This yields $g\left(\bar{u}^*_{\max}\right) < u^*_{\max}$ which contradicts the fact that $u^*_{\max} = g\left(\bar{u}^*_{\max}\right)$. This concludes the proof under case 2.

Proof of proposition 11:

For any $\mu < \mu^M$, the equilibrium cost reduction is defined by the condition $u^*(\mu) = g(\mu, u^*(\mu))$ where $g(\mu, u) = (C')^{-1} oD(\mu(c-u))$. Differentiating with respect to μ we obtain:

$$\frac{du^*(\mu)}{d\mu}\left(1 - \frac{\partial g}{\partial u}(\mu, u^*(\mu))\right) = \frac{\partial g}{\partial \mu}(\mu, u^*(\mu))$$

Since D is decreasing and $(C')^{-1}$ is strictly increasing (because C' is strictly increasing), $g(\mu, u)$ is strictly decreasing in μ for any $u < c$. In particular this is true for $u = u^*(\mu)$ which leads to $\frac{\partial g}{\partial \mu}(\mu, u^*(\mu)) < 0$. Furthermore, we showed in the proof of proposition (10) (more specifically in the proof of the preliminary lemma used to establish this proposition) that the derivative of g with respect to u is strictly smaller than 1 at the fixed point. In a setting where g depends upon μ as well this result writes as: $\frac{\partial g}{\partial u}(\mu, u^*(\mu)) < 1$. Hence, we can conclude that $\frac{du^*(\mu)}{d\mu} = \frac{\frac{\partial g}{\partial \mu}(\mu, u^*(\mu))}{1 - \frac{\partial g}{\partial u}(\mu, u^*(\mu))} < 0$ which means that $u^*(\mu)$ is strictly decreasing in the mark-up parameter μ over $\left[1, \mu^M\right[$. This property extends to $\left[1, \mu^M\right]$ since the equality $g\left(\mu^M, u^M\right) = u^M$ ensures that $u^*(\mu)$ is continuous at point $u = u^M$.

Proof of proposition 12:

Let $\mu \in \left[1, \mu^M\right[$. All firms reducing their costs by $u^*(\mu)$ and setting their prices to $p^*(\mu)$ is a subgame-perfect Nash equilibrium. According to lemma (6), it follows that $(u^*(\mu), p^*(\mu))$ is the solution to the maximization program:

$$\max_{u_i \in [0,c], p_i \in [0,a]} \pi(u_i, p_i) = (p_i - c + u_i) D(p_i) - C(u_i),$$

subject to the regulatory constraint

$$p_i \leq \mu \left(c - u^* \left(\mu\right)\right),$$

This shows that the maximization domain expands if μ increases, which ensures that $\pi \left(u^* \left(\mu\right), p^* \left(\mu\right)\right)$ is weakly increasing in μ. Moreover, it is strictly increasing in μ because the constraint (which depends on μ) is binding.

We have: $\pi \left(u^* \left(1\right), p^* \left(1\right)\right) = -C \left(u^* \left(\mu\right)\right) < -C \left(u^M\right) < 0$ and $\pi \left(u^* \left(\mu^M\right), p^* \left(\mu^M\right)\right) = \pi \left(u^M, p^M\right) > 0$. Since $\pi \left(u^* \left(\mu\right), p^* \left(\mu\right)\right)$ is continuous and stricly increasing in the mark-up μ over $\left[1, \mu^M\right]$, we conclude (using the intermediate value theorem) that there exists a (unique) threshold $\mu_0 \in \left]1, \mu^M\right[$ such that $\pi \left(u^* \left(\mu\right), p^* \left(\mu\right)\right) \geq 0$ if and only if $\mu \geq \mu_0$.

Conclusion

Cette thèse, constituée de quatre contributions en économie de l'innovation et de la concurrence, a essentiellement cherché à répondre aux questions suivantes:

1. Dans un environnement où la protection par le brevet est incertaine, quel est l'effet de la taille de l'innovation sur le choix du régime de protection de la propriété intellectuelle adopté par l'innovateur?

2. Comment le pouvoir de marché d'un innovateur sur le marché des licences est-il affecté par l'incertitude portant sur la validité du brevet?

3. Le programme de clémence Amnesty Plus a-t-il comme effet de dissuader la création de cartels ou au contraire d'encourager leur formation?

4. Une politique de régulation des prix plus contraignante diminue-t-elle les incitations d'une entreprise régulée à réduire ses coûts?

En réponse à la première question, et dans un cadre d'innovation de procédé dont la taille est mesurée par la réduction de coût qu'elle permet, nous établissons que la taille de l'innovation est un élément crucial dans la décision d'un innovateur quant au choix entre brevet et secret. D'une part, les innovations mineures sont toujours brevetées tandis que les innovations intermédiaires ne le sont que si la probabilité que le brevet soit valide et enfreint est suffisamment grande. D'autre part, un innovateur peut trouver optimal de protéger son innovation par le secret industriel lorsque son innovation est drastique. Par ailleurs, nous montrons que les accords à l'amiable pour éviter, ou résoudre, un conflit juridique entre le détenteur d'un brevet et un infracteur potentiel peuvent prendre diverses formes, dont celle d'un contrat de licence permettant de répliquer les profits espérés que les deux parties obtiendraient si le litige avait lieu. Des accords collusifs, dont le cas extrême est celui d'un accord permettant de répliquer le profit de monopole, sont également possibles et appellent à une vigilance de la part des autorités de la concurrence. Un paiement forfaitaire accordé par le détenteur du brevet à un tiers peut aussi être une manière d'éviter le litige, comme cela a été parfois observé dans le secteur pharmaceutique. Nous mettons en évidence dans ce chapitre une première conséquence de la prise en compte du caractère incertain du brevet, puisqu' un titre de protection parfaite serait toujours préférable au secret, dans la mesure où une infraction ne donnerait lieu à aucun dommage dans le cas du secret. Le travail a été prolongé par l'analyse d'une forme de dommages alternative à la notion d'enrichissement indû, à savoir celle afférente au manque à gagner (*lost profit*). Les résultats de ce prolongement, non présentés dans cette thèse, corroborent ceux du premier chapitre.

En réponse à la deuxième question, nous distinguons deux types de licences, celles avec redevance fixe et celles avec redevance unitaire (i.e. une redevance proportionnelle à la production). Nous montrons que le schéma avec redevance unitaire peut permettre au détenteur d'un brevet incertain de réaliser un profit supérieur au profit espéré qu'il aurait réalisé si la validité du brevet était testée avant que l'accord de licence ne soit conclu. Par contre, lorsque la licence prend la forme d'une redevance fixe, ceci n'est plus possible. L'incertitude qui porte sur la validité du brevet affecte donc le choix du régime optimal de licence du point de vue du détenteur du brevet, et c'est là une deuxième conséquence du caractère probabiliste de la propriété intellectuelle. Le détenteur d'un brevet incertain peut préférer un régime de licence avec redevance unitaire alors qu'il aurait préféré un régime de redevance fixe si la validité du brevet était certaine.

En réponse à la troisième question, nous étudions les effets potentiels du programme "Amnesty Plus" qui permet à une firme d'obtenir une réduction d'amende pour participation à un cartel si elle dénonce un second cartel. Nous montrons que sous certaines circonstances, ce programme peut avoir un effet anti-concurrentiel consistant à stabiliser un cartel qui n'aurait pu être soutenu sans le lien entre marchés créé par Amnesty Plus. Nous montrons néanmoins qu'il est possible d'éviter cet effet négatif potentiel en respectant une règle simple lors de la conception de ce programme. Nous montrons également que les entreprises peuvent être amenées à utiliser des stratégies de déclenchement multi-marchés sous Amnesty Plus pour contrer (partiellement) l'effet pro-concurrentiel d'Amnesty Plus ou renforcer son effet anti-concurrentiel alors même que ces stratégies n'auraient eu aucun effet collusif en l'absence d'Amnesty Plus.

Enfin, en réponse à la quatrième question, nous observons qu'une politique de régulation qui ne contraint pas trop le niveau des prix régulés peut encourager les entreprises à réduire leurs coûts lorsque la régulation se fait sur les performances propres de chaque entreprise régulée. Il s'ensuit une tension potentielle entre les deux objectifs d'efficacité allocative (niveau du prix régulé) et d'efficacité productive (réduction des coûts) dans le cas d'une régulation individuelle. Cette tension disparaît lorsque les entreprises sont régulées selon un schéma de performance relative: les deux objectifs sont parfaitement alignés dans ce cas. Autrement dit, une politique de régulation relative plus contraignante augmente l'investissement en gains d'efficacité et induit des prix moins élevés. Ce résultat est néanmoins nuancé par deux forces qui pourraient dissuader un régulateur d'appliquer une politique trop stricte: la contrainte de participation des entreprises peut ne plus être satisfaite si le *mark-up* autorisé par le régulateur est faible, et les investissements en amélioration de la qualité sont pénalisés par une régulation excessivement contraignante.

Au total, nos travaux nous auront permis de contribuer à améliorer la compréhension d'un certain nombre de problèmes complexes qui sont au coeur de la réalité économique. Certes, bien des questions posées appellent des prolongements multiples. De même, les améliorations potentielles souhaitables restent nombreuses. Enfin, les tests empiriques, absents de cette thèse, offrent des opportunités importantes. Mais il nous semble que le propos d'une thèse est au mieux de servir d'étape dans un processus d'innovation cumulative. L'auteur ne peut qu'espérer que ce modeste objectif a été atteint.