

Intellectual Property Rights and Efficiency

Henry Ergas¹

DATE

6 June 2010

¹ Former Chairman, Intellectual Property and Competition Review Committee, Australia

ABSTRACT

This paper provides an overview of the economic issues involved in the design of the intellectual property laws. Its primary purpose is to explain the economic context of intellectual property, as it has shaped the manner in which issues related to investment in creative effort have been treated. The paper examines the economic characteristics of intellectual property, the economic consequences of those characteristics, and issues involved in the design of intellectual property rights. It suggests that any presumption that ‘stronger’ intellectual property rights are necessarily better than ‘weaker’ rights is invalid.

KEYWORDS

intellectual property; efficiency; economic characteristics; non-excludability; non-rivalry; creative effort; innovation

A Introduction

1. This paper provides an overview of the economic issues involved in the design of the intellectual property laws. Its primary purpose is to explain the economic context of intellectual property, as it has shaped the manner in which issues related to investment in creative effort have been treated.
2. The structure of this paper is as follows. The paper begins by considering the economic characteristics of intellectual property; it then considers the economic consequences of those characteristics before examining some of the issues involved in the design of intellectual property rights.
3. It is worth saying at the outset that there are many important aspects of intellectual property that are not covered in this paper. This reflects the fact that its purpose is not to provide a comprehensive survey of the field. Rather, it centres on a relatively simple question: are ‘stronger’ intellectual property rights necessarily better than ‘weaker’ rights? It suggests that any such presumption is invalid.

B Economic characteristics of intellectual property

4. It is generally considered that five features of creative output are especially relevant to understanding the role and design of intellectual property rights. Each of these characteristics poses challenges to the functioning of market mechanisms.

Riskiness

5. While all investments – commitments made today with a view to securing returns in future – are risky, it is widely accepted that the risk involved in investment in creative effort is especially great. Most creative outputs add little to the existing stock of knowledge, and many prove to have no commercial value. Only a few prove significant, and of those, an even smaller number ultimately have great value. The probability of success is, in other words, low.

6. Given this low overall probability of success, returns, in those instances which do prove valuable, must be very high indeed if capital is to be used for investment in creative effort. This is all the more the case as the risk involved is far from being normally distributed (funds invested are not about as likely to earn more than average as less than average) and cannot be readily eliminated by “spreading one’s bets.”

7. There is an obvious risk that the very high returns accruing to a small number of ‘winners’ will be confused for monopoly profits. In fact, to the extent to which they merely compensate for the risks initially borne, they are no more than the opportunity cost of capital. Additionally, if these returns are in some way removed, without the risk of loss being reduced, investment in creative effort will be deterred.

Non-excludability

8. Securing appropriate compensation for the results of investment in creative effort is complicated by a second feature of creative endeavours. This feature is variously described as ‘**non-excludability**’ or ‘non-appropriability’. Simply put, this means that, in the absence of specific arrangements to the contrary, the creator could not *exclude* others from sharing in the outputs at issue, or *appropriate* the benefits they generate. The supplier cannot, in other words, control who consumes or what they consume.

9. The concept of excludability is merely another form of the economist’s concept of *externality*. An externality is an unrequited service: that is, a situation in which one party provides a service to another without the recipient of the service paying for or being paid for the service at issue. A lighthouse provides external positive benefits (‘goods’) to all ships that sight it; a smoke-emitting factory provides external negative benefits (‘bads’) to those who live close to it.

10. Excludability is a matter of degree. It depends on the technology used to produce and consume the service. In practice, there are few goods, if any, that completely lack excludability, but some may come closer to lacking it than others. Excludability also depends on the legal arrangements surrounding the service, and most notably the structure of property rights: in principle, the community surrounding the factory could be given a ‘property right in clean air’, so that the factory, if it wanted to use smoke-emitting technologies, would need to compensate the residents.

11. Those who describe the outcomes of creative efforts as being characterised by non-excludability simply mean that in the absence of specific legal arrangements to the contrary, it would be especially difficult for creators to prevent others from appropriating the benefit of their work. They could not, in other words, devise arrangements for themselves that ‘internalised’ the externality.
12. This is being altered to some degree by technology. It is presumably much easier now for creators, at least in those areas where the output of creative efforts is reduced to digital expression, to use ‘self help’ means (such as encryption) to retain control over their output.
13. Additionally, even in the absence of such self-help, creators can to some extent use secrecy to prevent others from appropriating the benefits of their investment. However, though remedies such as secrecy do have a role, they come at a cost. Secrecy naturally limits the revenues the creator can obtain from the creation, say through its sale or transfer; the creator also has to invest resources in securing and maintaining secrecy; and society more widely suffers from limited access to the creation.
14. Overall, non-excludability creates three sets of social costs.
15. First, non-excludability conflicts with the conditions for efficient investment. As a general matter, efficient investment requires that resources be devoted to creating ideas up to the point where the opportunity cost of those resources (their value in the most highly valued alternative use) equals the benefit society will derive from that investment. If investment choices are made on a profit-maximising basis, efficient investment in ideas requires that investors secure a return that, at the margin of investment in creative efforts, reflects the differential efficiency their idea contributes – that is, that the anticipated return to the marginal creator equal the anticipated gain to society from having the creation relative to not having it.
16. This simply means that for efficient private investment in creative effort to occur, investors must expect a return that at the margin is equal to the gain society makes from the creation their investment will permit.
17. Note that the efficiency condition set out above is one that must hold at the *margin* of investment in creative effort. From an economic point of view, it does not mean that *on average*, investors in creative effort must be able to appropriate the full social benefit of their creation; rather, it is sufficient that they expect their costs to be covered, so long as those costs are no less than the resulting social gain. At the margin, however, the expected return to the investor and to society should be equalised. It follows that if the marginal investors in creative effort cannot appropriate the full social benefit of their creation, socially desirable investment would not occur.
18. Second, in the presence of limited appropriability it is difficult for control over the commercial exploitation of ideas to be vested in those who can put them to the best use.
19. Thus, in a situation in which creators could only protect their creations through secrecy, they would have little ability to disclose their creations to potential purchasers or to negotiate in any detail with those purchasers. Moreover, greater difficulty would attach to describing the substance of any assets being transferred. Non-excludability therefore conflicts with the efficient allocation and exchange of rights over ideas.
20. Third, in the presence of non-excludability, creators will waste resources in costly

efforts to protect their creations while those seeking to appropriate them will invest in circumventing those measures. Some part of the resources thus used could be saved were property rights over the creations properly defined.

Jointness in consumption or ‘Non Rivalry’

21. A third economic characteristic of creative endeavours can be described as ‘**jointness**’ in consumption or as ‘non rival’ consumption. This means that consumption of the outputs of creative effort does not reduce the supply available for future consumption. Given an orange, if Sam eats it, that orange is not available to John. However, Sam’s knowledge of a poem does not reduce the volume of poetry available to John. Oranges display rivalry in consumption; poems do not.

22. Analytically, jointness is merely a way of saying that the cost of serving an additional unit of consumption, given that a first unit of consumption has been served, is nil. The marginal opportunity cost of supply, for any unit other than the first, is zero.

23. Like excludability, jointness is a matter of degree. Goods for which marginal costs are very low, given that a first unit has been produced, display it to some extent.

24. The polar form of jointness, in which the marginal opportunity costs are zero once the first unit of consumption has been served, applies to the non-material results of the creative process, rather than to the physical support usually used to distribute those results. The story in a book (in the sense of the words and their arrangement) is non-rival in consumption but the physical book itself is not. Increasingly, though, the distribution means themselves are characterised by some degree of jointness. For example, the marginal costs of making another copy of an electronic document are close to zero, being essentially limited to the opportunity cost of the time it takes for the copy to be made.

25. Just as non-excludability creates difficulties for the conditions for efficient investment, jointness creates a tension between securing efficient investment and securing efficient consumption.

26. While *efficient investment* requires that each creative effort have a claim on the social benefits it yields, *efficient consumption* requires that each potential consumer faces a price for an additional unit of consumption that reflects the marginal social cost of that consumption. For efficient consumption decisions to be taken about apples, the consumer must face a price for consuming an additional apple that reflects the cost to society of producing one more apple; the same principle applies to poems and to patented ideas.

27. The marginal cost of a creation, once the first unit of demand has been served, is zero or close to it. As a result, if efficient consumption is to be attained, the price of creations at the margin of consumption would have to be very low or zero.

28. If creators could exclude those who did not pay from consuming the output of their creative efforts, and if creators could efficiently price discriminate, the conditions for efficient consumption could be met simultaneously with those for efficient investment. In this case, price discrimination simply means charging for each unit on the basis of the willingness to pay for that unit. Perfect price discrimination means being able to do this in such a way as to perfectly align price with willingness to pay. For example, a high price might be charged for those uses where the idea was highly valuable, while

the price would be set to zero at the point where the marginal cost of using the idea consumed all of the efficiency it brought.

29. Analytically, there is a close relation between the notion of perfect or fully efficient price discrimination and that of internalising externalities. When a firm reduces price from above marginal cost to marginal cost, it confers a benefit that would not otherwise accrue – that benefit being the gain to the consumers that would otherwise forego consumption (or consume less). That benefit (technically, the consumer surplus the consumer obtains) is an externality to the firm. A firm that can price discriminate perfectly captures this benefit for itself, ‘internalising’ the externality.

30. As a general matter, however, perfect price discrimination is not of this world. There is consequently a tension between resolving the non-excludability issue – which usually implies giving creators control over the consumption (and hence pricing) of their creation – and efficiently allocating the creation once it exists.

31. This does not only affect final consumption but also the use of creations as inputs to other outputs. When creations are priced above their marginal cost, less use will be made of them as inputs than would be socially desirable. There is then a productive inefficiency, in the sense that society’s output is not being produced at least cost.

32. How significant these distortions are in practice depends on the extent to which those who exercise control over ideas can in fact price discriminate and where they cannot, can use other means to reduce the distortions in consumption that flow when creations are priced at more than marginal cost.

Cumulativeness

33. The fourth economic characteristic of creative efforts is ‘**cumulativeness**’. All that is meant by this is that ideas are inputs into the production of future ideas – the ‘standing on the shoulders of giants’ effect.

34. Seen from an economic point of view, cumulativeness is a form of potential externality. Future innovators benefit from access to the existing stock of innovations. At least potentially, there is a shared benefit.

35. In a perfect world, the developer of a creation (the ‘first stage innovator’) and potential investors in a subsequent creative effort that depended on that creation (the ‘second stage innovators’), would bargain in such a way as to ensure that any worthwhile second stage innovations proceeded. The bargain would internalise the potential externality, thereby correcting it.

36. In reality, however, effecting such bargains can be difficult, to the point where the costs of securing the bargain may consume all of the gains it could bring. As a result, some worthwhile second stage innovations will not occur. Additionally, the failure may compromise some first stage innovations, to the extent to which their attractiveness depends on revenue opportunities associated with innovations at the second stage.

37. The cumulative character of creation gives rise to a trade-off in which strong rights to today’s owners may limit future creation. This trade-off is obviously a complex one, as the strong rights can (1) themselves increase the availability of first stage innovations, providing more (or higher) shoulders for future innovators to stand on, and (2) raise the reward to the future creation (to the extent they too benefit from the strong rights),

while at the same time (3) increasing the costs second stage innovation involves. There are distributional consequences that flow from this (that is, some creators lose while others gain), but what matters from the perspective of society is the net impact on investment in creative effort. Recent economic journals contain a lively, but inconclusive, debate on the extent of this impact and on its determinants.

Network effects and interdependence

38. A fifth characteristic that pertains to creative effort in at least some areas is the relevance of **network effects**.

39. Network effects arise in situations in which the benefit derived from an activity increases with the number of participants that activity attracts. These effects can arise in consumption, production or both. Social networks (such as Facebook) provide a simple example of a network effect in consumption: the value of being part of such a network increases with the number of users on that network. Network effects in production arise, for example, when an operating system with a larger base of users attracts a greater number of suppliers of complementary products, such as applications software.

40. Some but not all network effects involve externalities. When Jane joins a communications network, on the basis that she now values access to that network sufficiently to cover the costs involved, she confers an external benefit on Sam, since once she is on the network, Sam can benefit from communicating with her. In a perfect world, the price Jane faces for connecting would be discounted to reflect the benefit that thus flows to Sam; if it is not, Jane may connect later than would be socially desirable (or not connect at all). When these interdependencies are not completely reflected in prices, the price mechanism is not fully playing its role of **coordinating** Sam and Jane's decision-making in such a way as to ensure that they take decisions that maximise wealth – that is, the value of the network to its users.

41. Coordination failures can also arise in the context of production. Linux may be a highly desirable operating system; however, realising its value depends on ensuring a host of complementary investments in applications software compatible with it. It would obviously be difficult for these coordinated investments to occur if those who carried them out could not appropriate and then divide the gains they permit society to obtain. As a result, the costs of non-excludability in terms of deterred efficient investment, may be especially great in situations where network effects are significant.

42. As with the other characteristics discussed above, dealing with network effects depends on devising mechanisms for 'internalising' the externalities – that is, for ensuring that those parties that are involved in a complex of interdependent transactions can each take decisions informed by the full costs and benefits, to the network as a whole, of their actions.

43. For example, network effects in demand need not distort consumption if the network owner has both an interest in and the ability to maximise the value of the network. The owner can then set price in such a way as to signal to Jane the gain the network as a whole accrues from her membership – subsidising Jane's access while taxing Sam's.

44. Equally, a network owner with both an interest in and the ability to maximise the value of the network can secure coordination in investment decisions. The owner can do this by signalling to investors in complementary assets (that is, those assets

that will enhance the network's value) the gain that will occur from their investment. This obviously requires that the network owner be able to compensate the investors in complementary assets for the gain they bring. This in turn requires some mechanism by which the network owner can secure that gain and then transfer it.

45. In that sense, solving the issues associated with network externalities generally requires, or at least can be assisted by, solutions to the problems of non-excludability.

46. A specific solution to some aspects of the problem of network externalities is provided by compatibility standards – that is, standards that allow goods and services provided by distinct suppliers to inter-work. These are essentially permissive, in the sense that they provide for inter-working, rather than securing coordinated supply. These standards can be set through negotiation, by regulation or simply through market selection leading to a *de facto* standard.

47. Where compatibility standards are set by negotiation, competition issues can arise about any horizontal and vertical agreements necessary to give effect to the standard. Where they emerge as a *de facto* standard through the operation of market forces, the concerns are more likely to involve the unilateral abuse of market power.

Conclusions on specific characteristics

48. A wide range of goods other than the results of creative effort display the characteristics set out above. For example, free-to-air TV channels are characterised by jointness in consumption and provide services that are thought to be non-excludable. However, it is reasonable to think that these characteristics are more extreme in the case of the outputs of creative efforts than in most other cases.

49. The main feature that flows from these characteristics is that, left to their own devices, it will be difficult if not impossible for investors in creative effort to align the rewards from that investment with its costs. They will not, in other words, be able to bargain their way around the externalities associated with creative effort. The result will be under-investment and distorted investment in creative effort and in activities complementary to creative effort. At the same time, control over the outputs of creative effort will not be transferred to those who can put these outputs to most productive use.

C Consequences of the economic characteristics and the role of IP

50. Intellectual property rights are generally seen as central to the means by which society addresses the characteristics set out above.

51. By allowing creators to appropriate a greater share of the social benefits associated with their creation, intellectual property rights can help support **efficient investment**.

52. Additionally, by delineating ownership, intellectual property rights facilitate the transfer by sale or assignment of the outputs of creative efforts. Because their rights are protected, creators can disclose the details of their property without losing the ability to benefit from it. They can therefore negotiate transfers more readily than they otherwise could. At the same time, the right itself can be used to define what it is that is being transferred or sold.

53. As a result, intellectual property rights support the **efficient exchange** of creations that is, their allocation to those who value them most highly. Because this increases the value of creations – both to the creator and to society – it further supports efficient investment in creative effort.

54. Efficient exchange is especially important for two reasons.

55. First, successful use of many ideas, notably in the applied sciences, requires the bringing together of the outputs of many innovators. Many products, particularly those that are ‘cutting edge’, involve combining ideas from many sources. Institutional mechanisms that reduce the transactions costs this involves can substantially increase productivity.

56. Second, creative effort differs from most organised forms of market-mediated production in that it involves significant actors that are not strictly commercial in character. More specifically, public and quasi-public institutions – such as universities and research institutions – play a central role in investment in creative effort.

57. These institutions are ‘idea generators’, rather than producers of tangible goods embodying those ideas. They have limited scope to secure the benefits of creativity by diversifying into the direct commercial supply of goods and services. To secure a return on their investment, as they are increasingly required to do, they must be able to transfer the use of their ideas to third parties. Again, to the extent to which intellectual property rights facilitate this process, they can increase society’s return on the investment it makes in these institutions.

58. A dimension of efficient exchange is that between first-stage and second-stage innovators. To the extent to which clear delineation of rights also helps promote bargains between sequential investors in innovation, it can allow the innovation decisions of tomorrow to be coordinated with those already made. However, whether this occurs obviously depends on the extent to which there is, in fact, clarity about the rights vested (an issue discussed in more detail below).

59. Finally, the delineation of intellectual property rights can help address network externalities. This occurs when ownership rights to the network are defined, creating a basis for decision-making that enhances the network’s value.

60. Where rights thus created are dispersed, rights owners can, by pooling their rights, secure a joint return on investments that increase the value of the network. Where the initial rights are centralised, the overall network owner can act as an ‘externality manager’, coordinating and rewarding complementary decisions.

61. However, the benefits identified above of intellectual property rights are obtained at a cost.

62. To begin with, the rights themselves are typically a reasonably blunt instrument, and there can be no assurance that the compensation they provide is ‘just right’ in inducing creative effort. There is, in particular, a risk that the rights provide too much compensation to infra-marginal creators – so that society would ‘pay too much’ for the innovations it secures. From the point of view of an individual country, ‘paying too much’ may be a cost if the resulting payments accrue to foreigners. Additionally, to the extent to which these excess payments are associated with unnecessarily high prices for goods (including ideas) covered by the rights, the result can be to reduce economic efficiency (a point discussed in more detail below).

63. Relatedly, it is conceivable that intellectual property rights induce *too much* investment. This is mainly because the rights, unless their value is exactly matched to the cost society necessarily incurs so as to secure the creation, can provoke a ‘race’ in which competing investors duplicate each others’ efforts in an attempt to be the first to secure the prize. Whether this happens depends to a degree on the design of the rights and of their allocation mechanism; whether it is harmful from a social perspective, and if so by how much, depends on how the costs of the duplication compare to any benefits society derives by the greater speed of creative progress.

64. The rights themselves, and competition to secure them, can therefore lead to waste in the process of creation.

65. Additionally, the rights can distort consumption decisions once the process of creation has occurred. This is for four reasons.

66. First, in the absence of perfect or near-perfect price discrimination, the price of creations protected by intellectual property rights will not be set, at the margin of consumption, to the relevant marginal cost – that is, zero. As a result, too little use will be made of creations.

67. The inability of creators to engage in perfect price discrimination can lead to a second distortion. This occurs when potential users of, or competitors to, the creation devote resources to ‘inventing around’ the rights. From society’s point of view, the resources thus used are wasted, at least to the extent to which they do not result in improvements on the original creation.

68. Third, creators, because they cannot perfectly price discriminate, cannot be expected to most efficiently support cumulative innovation. Rather, they may increase costs to future creators, discouraging some socially worthwhile investment in creative effort.

69. Fourth, for the same reason creators will not always succeed in securing the bargains needed to coordinate the multiple decisions involved in coping with network effects and other instances of interdependence in innovation. Indeed, to the extent to which the rights fragment or disperse control over interdependent creations (that is, over creations whose value is maximised when they are worked together), they may create a ‘tragedy of the anticommons’ in which innovations that would otherwise be brought together are not.

70. The costs associated with these distortions may be especially high when network effects are significant.

71. Where network effects are significant, pricing above marginal cost has the potential to impose greater social losses than it usually would. This is because the above cost pricing does not only lead the marginal consumer to forego some consumer surplus, but also imposes a loss on all network users – as the user base as a whole would have benefited to some extent from the foregone network expansion. In the jargon of economics, there is an infra-marginal loss, as well as the more standard marginal loss.

72. Additionally, intellectual property rights may affect competition between networks. More specifically, networks characterised by weak rights, or by a decision not to exercise rights, may not be able to compete on the merits with networks characterised by strong rights. For example, it may be difficult for a non-proprietary operating system such as Linux to displace a proprietary system, despite technical superiority. Given that the non-proprietary nature of Linux confers some gains on those users it does obtain, it

is not an answer to this issue to suggest that Linux too should adopt the proprietary model.

73. Three points summarise these considerations. The first is that intellectual property rights do address the central consequences of the features of creative outputs identified above. The second is that a key means by which they do this is by facilitating bargaining and contracting over rights and their allocations. The third is that they achieve their benefits at a cost. It is therefore certainly not true as a matter of economics that ‘stronger’ rights are always preferable to ‘weaker’ rights.

D Intellectual Property Rights and Efficiency: Balancing

74. As a general matter, intellectual property rights, as with other forms of property, are a right to exclude. The specific form of this right to exclude is a temporary right to prevent others from replicating the protected features of the creation covered by the right.

75. The scope of that exclusion is defined by the type of intellectual property right at issue. That said, all IP rights provide at the very least a right to prevent the supply of perfect substitutes. In some cases, the right extends also to preventing the supply of any form of economically relevant substitutes – that is, it involves conferring a right to be the sole supplier in a relevant market for a specified period of time.

76. It is a matter of philosophy whether there is here a difference of degree, if not of kind, relative to simpler property rights, such as those in real property. The right to real property is a right to possession and its benefits; exclusion is merely the natural corollary of the right to possess. My right to my bedroom gives me a right to exclude you in so far as that is needed to ensure that I can possess my bedroom; it does not give me a right to prevent you from enjoying your own bedroom, if you can do so without undermining my possession of the bedroom I own. In contrast, an intellectual property right allows me to exclude you from having an idea more or less like mine (depending on the right involved) even though the fact that you have that idea in no way reduces my ability to have that idea myself.

77. This right to exclude acquires its legitimacy as a social institution from the effects it has on the supply of creations, and through that supply on living standards. More specifically, it reflects the view that creators must be able to appropriate some part of the social gain associated with their creation if investment in creative effort is to occur. Exclusion of more or less perfect substitutes is the means by which that appropriation occurs.

78. This mechanism has an inherent balancing in it. The greater the social benefit from the creation, the more valuable will be the right to be its sole supplier and the greater the degree of exclusion effected by the right. Economically significant monopolies are only awarded to those who provide great benefits.

79. A system that rewarded creators through state grants or subsidies could not hope to be as effective in matching the reward with the contribution. Such a subsidy scheme would assign the costs of innovation to taxpayers, with government deciding which creative efforts should or should not proceed. These types of decisions are difficult to take even for basic science; using collective decision-making to choose among alternative commercial inventions is plainly fraught with difficulties, including problems of adverse

selection (the state scheme is offered the bad ideas, while creators keep the good ones for themselves) and of moral hazard (having government as an investor in an inventive project may lead the creator to slacken his or her own efforts). In contrast, in a system where inventors compete to secure intellectual property rights, and are primarily rewarded for success through the ability to exercise those rights, the costs of invention are borne in the first instance by inventors (and their financiers), creating strong incentives for success.

80. From a societal point of view, the incentive provided by this mechanism gives a significant stimulus to dynamic competition – that is, to competition through innovation. As this is the form of competition that over the longer term is the most powerful, intellectual property rights make for greater and more potent rivalry. But, as explained above, that gain is not secured without a cost.

81. In seeking to the balance between the undoubted benefits of IP rights and their equally undoubted costs, individual countries (and groupings of countries such as the EU) have taken somewhat differing approaches. Despite that diversity, three broad sets of balancing mechanisms can be identified: the weight given to IP rights relative to other rights; the design of IP rights; and the administration of the intellectual property system.

The weight given to IP rights

82. IP rights are one form of right among others, with those others qualifying IP rights to a greater or lesser degree. Especially relevant here are the rights and obligations associated with the competition laws.

83. The right to exclude granted by IP rights can be used to distort competition in ways that go beyond the mere exclusion provided for in the right itself. The issue then arises of the extent to which competition policy is capable of being applied (and, as a matter of tradition and precedent, is in fact applied) when these instances occur.

84. In the United States, it has long been clear that conduct involving intellectual property is fully within the reach of the anti-trust laws. For example, in 1995, the US Department of Justice (DoJ) and the Federal Trade Commission (FTC) issued *Antitrust Guidelines for the Licensing of Intellectual Property* including copyright, patent, trade secret and know-how. The Guidelines are designed to provide guidance for parties and to help them predict whether the DoJ or the FTC will challenge a particular practice as anti-competitive. The Guidelines embody the general principle that “for the purposes of antitrust analysis, intellectual property is comparable to other forms of property”.

85. This principle is reflected in a long tradition of granting compulsory patent licenses under the anti-trust laws. Its application to copyright was highlighted in the successive Microsoft decisions, with the trial judge in *United States v. Microsoft Corp.*, (D.C. Circuit, 2001) finding that “Microsoft’s primary copyright argument [that the right protects conduct, even if anti-competitive] borders upon the frivolous. The company claims an absolute and unfettered right to use its intellectual property as it wishes. . . . That is no more correct than the proposition that use of one’s personal property, such as a baseball bat, cannot give rise to tort liability.”

86. This capacity of the US competition laws to over-ride the intellectual property rights, or at least to hem in the use of IP rights, has to some degree offset the wide scope of

the IP rights granted in that country. At the very least, the interaction between the intellectual property laws and the competition laws creates a corrective mechanism, which can, and historically has, come into play in the United States when excessively broad or strong IP rights unduly constrain dynamic competition.

87. In contrast, in Australia and New Zealand, considerations as to the appropriate balance to be struck in the definition and enforcement of IP rights have very largely, if not entirely, been given effect by embodying specific provisions, aimed at achieving that balance, in the IP system itself; the bundle of rights thus defined has then been largely protected from the more general statutes aimed at promoting competition. There has, in other words, been a marked preference for relying on the definition of the rights, rather than on *ex post* constraints on the exercise of those rights, as the primary means of striking a balance between incentives to creative effort and the public interest in wide access to ideas.

88. It is useful to consider why this has been the case. At the highest level of abstraction, three factors seem relevant.

89. The first is the legal tradition, which despite some significant differences relative to the UK, is probably closer to the formalism Atiyah and Summers (1987), in their classic comparison of the English and American legal traditions, ascribe to the UK, than to the instrumentalist approach common in American law. In the IP area especially, this translates into a tendency to interpret rights fairly narrowly, and hence to hew closely to the precise definition of the extent of the rights as given in legislation.

90. A second factor is the economic context. Australia and New Zealand have long relied heavily on imported technology, and well-defined IP rights may have been valuable in reducing the transactions costs of the contracts required. Moreover, a substantial share of the R&D undertaken domestically has been publicly funded. Traditionally, this publicly funded R&D was carried out largely in government agencies; much of this research involved projects where individual end-users, for example dairy farmers, were price-takers in world markets, and hence did not perceive each other as rivals. As a result, these agencies tended to distribute the results of their research widely, especially domestically, with IP rights placing little restriction on end-users' access to research results. More recently, government has placed considerable emphasis on cooperative research (involving both public and private sector participants) in emerging technologies, and clear IP rights may have been viewed as useful in facilitating both the initial investments and the commercialisation of any results.

91. Last but not least, competition policy – both in the sense of ‘antitrust’ and in the sense of policy aimed at removing statutory obstacles to competition – is a relatively recent arrival on the scene (dating, in Australia, to the early 1970s and in New Zealand to the 1980s). In contrast, IP law has roots that stretch back to initial colonisation, reflected in a long legacy of both statute and case law. Concerns about preserving the integrity and clarity of that precedent have weighed on policy-makers whenever questions have arisen of whether and how competition issues relating to the IP laws should be addressed.

92. In short, in contrast to the United States, other rights, and notably those associated with the competition laws, cannot be counted upon to temper any adverse societal consequences of changes in the intellectual property laws.

The design of IP rights

94. This is a response the IP laws lend themselves to, as they are rich in instruments that can be used to ‘fence in’ or more generally ‘fine tune’ the exclusivity provided. These instruments include through changes to coverage, originality requirements, duration, the precise content of the reserved rights and the mechanisms for the exhaustion of those rights, statutory exemptions and exceptions, the allocation of burdens in the context of infringement proceedings and the nature and extent of remedies. This multiplicity of defining variables offers (or at least has been thought to offer) the scope to deal in a targeted, well-defined way with the competing considerations of exclusivity and access, without introducing the uncertainties that would arise were an approach based on general concepts and principles – as is largely a characteristic of competition law – superimposed on, and allowed to over-ride or redefine, the rights granted in the IP statutes.

95. Although this approach is not per se inefficient (and indeed, has notable strengths relative to that prevailing in the US), its implementation places considerable weight on a careful understanding of the efficiency consequences of individual design elements and their interaction. While substantial economic research has been undertaken on this question in recent years, the sheer complexity of the rights, and of the effects arising from the interaction between their component elements (such as breadth, duration and requisite inventive step), mean that research is necessarily based on relatively abstract, often highly simplified, models. Nonetheless, the results are at least suggestive. They can be highlighted by reference to three important elements in the design of patents: duration, inventive step and scope.

96. Modelling the **optimal duration** of patents involves calibrating the trade-off between the stimulus to invention provided by longer duration and the cost to society of the exclusive right. Because technological circumstances differ between industries (it is easier to innovate in industries where the technological base is in any event experiencing rapid progress), that trade-off will differ across industries, as will the optimal duration of patents. Optimal duration, in turn, is closely connected with the optimal structure of renewal fees, i.e. the amount patentees should periodically be charged for maintaining the registration of their patent during its life. A shorter optimal duration is equivalent to a more steeply rising renewal charge.

97. Using a model that seeks to capture these trade-offs, Cornelli and Shankerman (1999) examine both the optimal differentiated patent lives and the optimal uniform patent life. They find an optimal uniform length of 15~19 years. As regards the optimal differentiated length, the minimum is 7 years, most optimal lives are between 8 and 15 years, and for a few very high value projects, optimal lives exceed 20 years. They also compute the welfare gains from moving from an optimal uniform life to an optimal differentiated life for France, Germany and the UK. They estimate gains in the order of 5 percent.

98. A result of Cornelli and Shankerman (1999) is that optimal renewal charges should rise far more steeply than do current charges, so as to encourage low value patents to lapse.

99. More recent models take account of sequential innovation. For example, Llobet, Hopenhayn and Mitchell (2000) devise a mechanism in which the innovator proposes a price at which it will sell rights to the patent, and pays a fee that increases with that price. Absent such a mechanism, the inference is again that renewal fees should rise

more steeply, which is consistent with an optimal patent life that is shorter than the current standard.

100. Turning to the height of the required **inventive step**, there is an inverse U-shaped relation between that step and the rate of innovation (Hunt 2003). For moderate levels of the inventive step requirement, increasing (reducing) it, will increase (reduce) the rate of innovation. This is because an increase in the height of the test increases effective patent life, which rewards investment in major innovations (Scotchmer, 2004). The inventive step requirement should be especially high where or when the exogenous rate of technical progress is high. In other words, in industries where the underlying rate of progress is high, allowing minor inventions to qualify for protection reduces the rate of invention, and hence is socially harmful.

101. Finally, **patent breadth** refers to the scope of the exclusivity accorded the rights-holder. The greater is patent breadth, the greater is the range of related inventions a rights-holder can claim as falling within (and hence potentially infringing) the scope of his or her right. As a result, greater patent breadth increases the ‘winner takes all’ aspect of patents, and hence increases the likelihood of inefficient patent races.

102. Greater breadth is more dangerous from a social point of view than greater duration. This is because greater breadth makes it more difficult to imitate around, or on top of, the patent, while longer duration increases the incentive to imitate or improve on the invention. When breadth and duration are treated as the relevant design parameters, and tradeoffs are allowed between them, optimal patents prove to be ‘long and narrow’, rather than ‘short and wide’. In other words, it is better for duration to be set slightly too high, than for the scope of exclusivity to be over-broad: The fact that most rights are in any event made obsolescent relatively quickly underscores this conclusion.

103. While the formal literature in respect of each of these design attributes primarily relates to patents, the underlying models are typically very general. This means that attributes such as duration, inventiveness and breadth can be reasonably transposed from the patent context to that of other IP rights. Seen in this perspective, the overall result of the literature is that the balancing, though never simple or unambiguous, does not support rights that are very broad, can be acquired for a low inventive step, or have long durations. Consideration of the sequential nature of innovation tends to (again noting that the results are never unambiguous) increase the desirability of taking great care about the duration and breadth of rights, especially when transactions costs associated with licensing are high.

Administration of the intellectual property regime

104. The formal literature discussed above typically assumes that administration of IP regimes is costless and perfect. These assumptions imply that no errors are made in attributing the rights to right-holders and that no costs are incurred by parties in identifying what rights are owned by whom and to what. This is obviously a far cry from reality, where substantial difficulties arise in the allocation and enforcement of rights, and where parties must incur high costs so as to identify the structure and coverage of the rights that have been granted.

105. Thus, Jaffe and Lerner (2004) examine the US patent system and conclude that there has been an explosive rise in the number of poor quality patents issued. While some part of this growth is due to growth in the patenting of business methods, their

results suggest the problem goes much further.

106. Before turning to this issue in greater detail, it is worth considering the respects in which rights can be of “poor quality”. Four such respects can be identified. First, a right may be of uncertain validity, in the sense that it may not meet the statutory hurdles. Second, the boundaries of a right may be unclear, making it difficult for parties to know just how far its exclusivity reaches. Third, information about the right may not be publicly available, or may be unduly costly to obtain, creating a risk of inadvertent infringement. Fourth (and somewhat different in kind from the other defects noted), the allocation of rights may be inefficient, notably in being excessively fragmented, creating high costs for parties that need to assemble a complex or bundle of rights if they are to operate – this is the ‘anti-commons’ problem referred to above.

107. A large scale empirical study by Bessen and Meurer (2008) examines the US patent system in terms of these potential quality problems. Bessen and Meurer conclude that other than in chemicals and pharmaceuticals, the system performs poorly. This is mainly because there has been a trend to allowing ever more abstract and broader claims, in some cases of low quality and with a low degree of specificity. These claims cover technologies that are in fact not known to the inventor at the time the patent is granted. They hence confer rights on patentees who have not in fact made the innovation those claims protect. This can only reduce the incentives for those innovations to be undertaken, and to that extent, represents a net social loss.

108. Bessen and Meurer argue that over-broad, too abstract, claims undermine the ‘notice’ function of the intellectual property system. By this they mean the function of delineating the ownership pattern, allowing potential users to trace that pattern, contact and then contract with rights-owners, and thus obtain the rights they require for non-infringing operation. This ‘notice’ function of the patent system can be analogised to the similar function undertaken by the cadastral system with respect to rights over land. Together with the US patent system’s tolerance (if not positive encouragement) of ‘submarine’ patents, the result, according to Bessen and Meurer, is an intellectual property landscape in which there is a high risk of inadvertent infringement, with little or no scope for users to insure themselves against that risk. This is in contrast to the situation with respect to property in land, where the delineation of rights is relatively clear and where there is an active, low cost, market for title insurance.

109. The resulting inefficiencies are accentuated by the ever greater fragmentation and overlapping of IP rights. As stressed in Heller (2008), such excessive fragmentation – often referred to as creating a “rights thicket” – undermines efficiency in at least two respects: it prevents *efficient use*, as the costs of assembling the minimum bundle of rights needed for non-infringing use may exceed the value of that bundle; and it deters *efficient investment*, because it becomes too costly for spill-over effects (in which the efforts of one innovator benefit another) to be internalised.

110. These costs are to some extent reduced by the ever-greater reliance on complex cross-licensing arrangements, in which parties essentially pool IP rights in return for protection against infringement claims. However, these arrangements can impose significant costs of their own. There are, to begin with, potentially high transactions costs involved in the set-up and operation of networks of cross-licenses. Second, the increasing importance of cross-licensing arrangements encourages purely defensive use of the IP system, in which parties seek IP rights merely for the sake of having something to trade. In so doing, they can make existing ‘thickets’ even more tangled and hence so-

cially costly. Third, such cross-licensing arrangements may become vehicles for tacit or even explicit collusion, with all its costs. Finally, cross-licensing is of little help to new and small producers, who will typically have few rights to trade, with the result that networks of cross-licenses may create or increase entry barriers.

111. Bessen and Meurer use a range of methods to quantify the costs associated with the performance of the US patent system. They estimate that the patent system provides a significant positive net subsidy to invention in the chemical and pharmaceutical industries but acts as a negative net subsidy to (i.e. a tax on) invention in other industries. This difference between industries arises mainly from the fact that prior art is more readily established in the chemical and pharmaceutical industries than in other industries, with the result that patent claims tend to be more readily tested, more precise in their formulation and hence better bounded.

112. Bessen and Meurer's overall conclusion is that "during the late 1990s, the aggregate cost of patents exceeded the aggregate benefit of patents for United States public firms outside the chemical and pharmaceutical industries. This implies that patents very likely provided a net disincentive for innovation."

113. The concerns about system administration raised by Jaffe and Lerner (2004) and Bessen and Meurer (2008) have a number of implications. At the most general level, they highlight the dangers involved in defining strong rights in the hope that parties can "negotiate around" any overly great breadth or coverage. Rather, the transactions costs that would have to be incurred in seeking, organising and enforcing such bargains may swamp the benefits to the parties involved. This is especially the case when rights are poorly defined and inherently uncertain. The scope for parties to bargain does not, in other words, make errors in the design and administration of rights self-correcting.

114. Nor does litigation make such errors self-correcting. The difficulty is that the successful applicant in any invalidity proceeding confers an externality on his or her rivals and on consumers; in contrast, that applicant is exposed to very high costs should the application fail. As a result, there are (from the standpoint of efficiency) too few proceedings for invalidity. At the same time, infringement proceedings can be used strategically to prevent competition. This may lead to an inefficiently high number of infringement suits.

115. Bessen and Meurer test whether the defendants in infringement proceedings are likely to be 'copy cats', who are merely seeking to free ride on others' investment in creative effort. To the contrary, they find that alleged infringers are likely to invest more heavily in R&D than are the plaintiffs in infringement suits. As a result, with uncertain, potentially over-broad, IPRs, litigation may increase rather than reduce inefficiency.

116. Considered as a whole, these concerns suggest policy makers need to strike a very careful balance between Type I errors (granting an IP right where it is not merited) relative to the corresponding Type II errors (not granting a right where it is merited). At least on the analyses reviewed, the cost of the former can be high relative to the costs of the latter.

117. In terms of how that balance could be achieved, one possibility is more demanding criteria for the grant of rights. Options in this respect include better defining the scope of the subject matter that can be protected (for instance, so as to better enforce the distinction between inventions and discoveries) and increasing the height of originality or inventive step requirements. A shorter duration of granted rights would then act as

an insurance mechanism against errors in the grant of those rights, as the harm would be less durable. Another possibility involves requiring more stringent examination of applications, notably for patents, and weakening presumptions of validity. Finally, there may be scope to alleviate some of the concerns about the IP system by increasing the constraints imposed on exercise of IP rights by the competition laws.

E Conclusions

118. As rights to exclude, IPRs sit uneasily with the conventional notion of the primacy of competition, which centres on the ability of several, possibly many, parties to act as rivals in striving for economic rewards.

119. However, the uneasy nature of the relationship should not suggest that there is any contradiction between IP rights and competition: rather, it is a truism that IP rights, by allowing creators to secure a greater share of the social gain from their creation than they would otherwise, can promote investment in creative effort; and that it is this investment that underpins the development of new processes and products which not only contributes directly to increased wellbeing but also, in Schumpeter's famous phrase, is a form of competition "as much more effective than [conventional price competition] as a bombardment is in comparison with forcing a door".

120. This is clearly a substantial gain to society; but it is bought at a cost. Once made, ideas are relatively readily transmitted and used; they are, in economic terms, non-rivalrous. From an economic point of view, therefore, they ought to be used as widely as possible – which implies a price of or close to zero (for the idea itself, though not for its material form). The exclusionary right granted by IP rights, however, enables the right-owner to set a positive price for the protected material, thereby reducing the flow of and output from ideas. Even more important, as ideas themselves serve as the basis for generating further ideas, the social cost of the reduced flow can take the form not only of less income today but also of less growth in income in periods to come. If owners of IP rights can use their exclusionary rights to extend their control even further than merited, securing an income stream in excess of the social gain arising from their creation, the costs to society will be greater yet.

121. The balancing this imposes has been the subject of an enormous, and still rapidly growing, economic literature. While hardly free of differences and controversies, it is clear from that literature that there is no presumption that 'stronger' rights are necessarily more socially desirable than 'weaker' rights, or that systems in which rights are granted more liberally and enforced more strictly are necessarily preferable to those in which rights are only granted cautiously and enforced prudently. Granting rights that are 'strong' (in the sense of allowing substantial exclusion) but poorly specified and not well based on criteria of significant inventiveness is especially problematic. The trend towards ever broader, more poorly defined rights, with ever more draconian penalties for infringement, is therefore as likely to reduce welfare as to increase it.

References

Atiyah, P.S and Summers, R.S., 1987, *Form and substance in Anglo-American law: A comparative study of legal reasoning, legal theory, and legal institutions*, Clarendon Press,

Oxford.

Bessen, James and Michael J Meurer, 2008, *Patent Failure: How Judges, Bureaucrats and Lawyers put Innovators at Risk*, Princeton University Press, Princeton and Oxford.

Cornelli, F. and M. Schankerman, 1999, 'Patents Renewals and R&D Incentives', *RAND Journal of Economics*, 30, 197-213.

Heller, Michael, 2008, *The Gridlock Economy: How Too Much Ownership Wrecks Markets, Stops Innovation, and Costs Lives*, Basic Books, New York.

Hunt, Robert M., 2003, 'When do more patents reduce R&D?', *Working Paper No. 06-6*, Federal Reserve Bank of Philadelphia.

Jaffe, Adam B., and Josh Lerner, 2004, *Innovation and its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It*, Princeton University Press, Princeton and Oxford.

Llobet, Gerard, Hugo Hopenhayn and Matthew F. Mitchell, 2000, 'Rewarding sequential innovators: prizes, patents and buyouts' *Staff Report 273*, Federal Reserve Bank of Minneapolis.

Scotchmer, Suzanne, 2004, *Innovation and Incentives*, MIT Press, Cambridge and London.

U.S. Department of Justice and the Federal Trade Commission, 1995, *Antitrust Guidelines for the Licensing of Intellectual Property*, available at <http://www.justice.gov/atr/public/guidelines/0558.htm>.

United States Court of Appeals for the District of Columbia Circuit, 2001, *United States of America, Appellee v. Microsoft Corporation, Appellant*, available at http://cyber.law.harvard.edu/msdoj/msft_ruling.html