Tilburg University

Technological distance, growth and scale effects

Peretto, P.; Smulders, J.A.

Published in:
Economic Journal

Publication date:
2002

Document Version
Peer reviewed version

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Peretto, P., & Smulders, J. A. (2002). Technological distance, growth and scale effects. *Economic Journal*, 112(481), 603-624.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

TECHNOLOGICAL DISTANCE, GROWTH, AND SCALE EFFECTS*

Pietro Peretto and Sjak Smulders

We present an endogenous growth model in which the scale effect may be positive or negative, but vanishes asymptotically. The mechanism behind this result provides a microfoundation for models that exploit the interaction of growth and market structure to remove the scale effect. When more firms are active, the economy is more specialized: in the sense that firms are less likely to work on related problems. This increase in technological distance reduces the spillovers between firms. A larger economy with more firms accumulates more knowledge. However, the spillovers that benefit a firm do not necessarily increase because of the differentiation of the knowledge stock.

One of the puzzles in the theory and empirics of endogenous technological change is the *scale effect*. Specifically, models in which growth is driven by the accumulation of non-rival knowledge predict that larger economies (measured by a larger labour force) grow faster because (a) they have more resources to devote to knowledge creation and (b) the larger scale to which knowledge can be applied raises the returns to innovation. This prediction is difficult to reconcile with empirical evidence. In their study of the cross-sectional evidence Backus, Kehoe and Kehoe (1992) find that GDP growth is not related to the scale of the economy, although TFP growth in manufacturing is positively related to the scale of the manufacturing sector. In his study of the

time-series evidence, Jones (1995) finds that the behavior of TFP growth and R&D investment in the manufacturing sectors of OECD countries is inconsistent with the scale effect.

We present an endogenous growth model where the scale effect may be positive or negative but always vanishes asymptotically. What distinguishes our paper from several others, discussed below, is that we provide a microfoundation for a general class of models that exploit the interaction of growth and market structure to remove the scale effect. Moreover, our model is consistent with microeconomic evidence on R&D processes and knowledge spillovers within firms and industries and the role of market structure in shaping these processes.

The new feature of our framework is that we model knowledge accumulation and spillovers as two-dimensional: not only does research and development lead to improved understanding of existing problems (intensive margin), but it often implies original formulation of new problems, and new lines of research (extensive margin). Accumulation of knowledge along the intensive margin expands the public knowledge stock and gives rise to spillovers, as emphasized in the standard literature on spillovers and growth (Aghion and Howitt 1990, Romer 1990, Grossman and Helpman 1991). Accumulation along the extensive margin leads to higher specialization in firms' R&D activities and to dilution of spillovers.

To model innovation along these two dimensions, we emphasize the interaction of two elements: (a) technological differentiation, and (b) market structure dynamics. Technology is firm-specific in the sense that firms have unique knowledge of products and processes that they accumulate through R&D activity. A fraction of the knowledge that they create spills over and contributes to the aggregate stock of public knowledge that can be used to improve existing technologies or to create new ones. The extent to which a firm can take advantage of the public knowledge created by other firms decreases with the technological distance between the creator and the user of such knowledge.

We posit a mechanism whereby technological distance increases with the size of the economy. Firms operate within *spillovers-networks* determined by the technological relatedness, or proximity, of their research activities. The typical network spans only a fraction of the total population of firms because it includes only firms that work on technologically related problems. When the economy expands, there is entry of new firms and creation of new lines of research. As more firms are active, the economy becomes more specialized in the sense that each firm is less likely to work on problems that are related to those on which another firm is working, and therefore it is less likely to operate within the same network as that firm and thus contribute to that firm's spillovers. This implies that the distance between firms -- the probability that they participate in different networks -- increases and, therefore, that the networks expand at a slower pace than the economy. As a result, a larger economy accumulates a larger aggregate stock of public knowledge while the spillovers that benefit the typical firm, which expand with the size of the network within which the firm operates, do not necessarily increase and can in fact decrease. The scale effect vanishes asymptotically because the size each network becomes independent of the scale of the economy when this becomes very large.

The microeconomic literature on innovation, spillovers, and market structure provides evidence for the building blocks of our model. Summarizing a large number of studies, Dosi (1988) shows that technology improves in an incremental and gradual way within existing firms that build on their own history. The introduction of innovations is not generally due to entry of new firms.¹ On the other hand, while many innovations occur without entry, entry often requires innovation. If entry is costly, entrants have to introduce marketable products that are different from those that incumbents offer in order to be able to recoup the entry costs. Innovations introduced by entrants have more radical effects than incremental in-house R&D since they affect the technology structure of the economy and the pattern of specialization. This suggests that entry

and innovation should be regarded as distinct, but related, phenomena.²

While the existence of the scale effect at the aggregate level is disputed, empirical studies unambiguously find it at the firm level. Cohen and Klepper (1996) review the literature on R&D and firm size of the last 30 years. They conclude that (a) the likelihood of a firm reporting positive R&D rises with firm size and (b) R&D rises monotonically with firm size among performers of R&D. Larger firms benefit from larger output volumes to which the fruits of R&D can be applied. This *cost-spreading* mechanism determines the positive relation between firm size and returns to R&D.

The concept of technological distance is widely applied in microeconomic empirical studies on R&D and knowledge spillovers. These studies, summarized in Griliches (1992), find that a *weighted sum* of cumulated R&D expenditures within the industry or aggregated over related industries affects the productivity of a firm's own R&D. The weights measure the effective fraction of knowledge that a firm draws from another. Jaffe (1986), for example, uses the distribution of firms' patents over patent classes to characterize their relative positions and construct measures of technological distance. Evidence of knowledge dilution -- a notion similar to our concept of knowledge differentiation -- is provided in Adams and Jaffe (1996). They match firm-level R&D expenditures to plant-level TFP. They find that total firm-level R&D expenditure affects plant-level productivity, but the number of plants negatively affect these intra-firm spillovers. Also, spillovers from technologically related firms depend on R&D expenditure per plant rather than total expenditures.

Griliches's summary of the state of the art highlights some of the ideas that we capture in our model:

‘To the extent that an invention either reduces the cost of production or develops

entirely new products, it has an aspect of increasing returns to it. The same invention could produce the same proportional effect, in different size markets or economies. The public good nature of most inventions and the "multiplicative" aspect of their impact do not require their number to grow just to sustain a positive rate of productivity growth. On the other hand, economies do not grow just by replication and expansion; they also get more complex, proliferate different products and activities, and develop in different geographical and economic environments. To that extent, the "reach" of any particular invention does not expand at the same rate as the growth of the overall economy, but only at the rate of its "own" market.' (1990, p. 1698).

Several R&D-based growth models have been proposed recently that remove the scale effect through mechanisms similar to the one that we discuss here. All these models make the following two assumptions. First, the scale of the economy affects the number of firms that accumulate their own firm-specific knowledge. Second, R&D productivity depends on some measure of accumulated public knowledge that is independent of the number of firms and hence of the scale of the economy. This independence may stem from the assumption that (a) spillovers across firms are absent (Peretto, 1999), that (b) spillovers depend on average knowledge (Smulders and Van de Klundert, 1995; Peretto, 1998a, b; Dinopoulos and Thompson, 1998), or that (c) spillovers depend on the knowledge of the most advanced firm (Young, 1998; Aghion and Howitt, 1998; Howitt, 1999). All these models have the property that a large economy replicates the structure of a small economy. They thus ignore that a larger economy allows a larger degree of specialization and complexity, as emphasized by Griliches. Moreover, although they allow for spillovers, all these models assume that a larger number of firms undertaking

independent R&D projects does not support a larger aggregate stock of public knowledge. The implicit assumption is that all public knowledge is replicated.

Jones (1999) has summarized these contributions by relating them to a reduced-form growth model with the following features. Per capita income depends on the share of the total labour force employed in production and on the productivity level of the representative firm, which increases through R&D. The output of the R&D process, and hence the rate of growth, increases with the number of researchers employed per firm and the productivity level already attained. The latter captures spillovers from the existing knowledge stock. Jones shows that an economy with larger labour force grows at an endogenous rate that is independent of the size of the labour force, only if the productivity of R&D increases linearly with the knowledge stock and if the number of firms increases linearly with the labour force. This paper shows that what seems to be a knife-edge case in Jones' representation arises naturally in a framework that takes into account the structure of knowledge. It does so because it provides a microfoundation for the spillovers function that in the aforementioned papers is posited without discussion of the mechanism that generates the properties needed to eliminate the scale effect. One of the contributions of the paper, therefore, is to spell out the empirical and theoretical foundations of models of this class.

Another contribution is to emphasize that under plausible conditions the scale effect may be *negative*. Entrants incur a sunk cost and benefit from public spillovers. We show that a larger economy supports more firms, but whether this results in a faster or slower pace of growth depends on whether it is entrants or incumbents that benefit most from public spillovers. We thus provide an alternative mechanism for the possibility of negative scale effects also present in Young (1998). More generally, our model shows that the crucial force underlying the scale effect is how an expansion of the labour force affects incentives to innovate and how it shifts the

allocation of resources over the two types of R&D activities.

Finally, we think that the incorporation of knowledge spillovers from a pool of public knowledge to which the total population of firms in the economy contributes not only relieves the knife-edge character of non-scale endogenous growth models, but does also justice to the basic property of non-rivalry of ideas and captures aspects from the microeconomic literature on innovation spillovers and market structure.

The paper is organized as follows. The model is specified in Section 1. We characterize the steady state in Section 2 and discuss our results on the scale effect in Section 3. We conclude in Section 4. The Appendix deals with the transitional dynamics and presents an extension regarding the central part of our model, *viz* spillovers among firms.

1. The Model

1.1. Households

Households inelastically supply one unit of labour each. They maximize the present discounted value of the stream of consumption of a bundle of differentiated products,³ C :

$$W = \int_0^{\infty} \ln C \cdot e^{-\rho t} dt; \quad (1)$$

$$C = \left(\int_0^{\infty} C_i^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}. \quad (2)$$

where $\varepsilon > 1$ is the elasticity of substitution between products and $\rho > 0$ is the intertemporal discount rate. The solution to the household maximization problem are the well known Dixit-

Stiglitz demand schedule and the Keynes-Ramsey rule:

$$X_i = LE \left(\int_0^N P_k^{1-\varepsilon} dk \right)^{-1} P_i^{-\varepsilon}; \quad (3)$$

$$\hat{E} = r - \rho, \quad (4)$$

where E is household expenditure, L is the number of households (and our measure of the scale of the economy), $X_i = LC_i$ is aggregate demand for product i , N is the number of products (monopolistic firms) in the market, and r is the market rate of interest. Hats denote growth rates.

1.2. Incumbents

At a moment in time, N firms are active. Each firm i controls its own stock of firm-specific knowledge Z_i that allows it to produce quantity X_i of its own variety and sell it in a monopolistic market against price P_i . Labour is the only factor of production. L_{xi} units of labour are allocated to production with productivity depending on knowledge Z_i with elasticity $0 < \theta < 1$,⁴

$$X_i = Z_i^\theta L_{xi}. \quad (5)$$

The firm employs L_{zi} units of labour to accumulate firm-specific knowledge according to

$$\dot{Z}_i = Z_i^{1-\Psi} S_i^\Psi L_{zi}. \quad (6)$$

Productivity of labour in R&D depends on own knowledge Z_i and spillovers S_i , which consist of a weighted sum of all firms' knowledge stocks; see below. This captures two important ideas.

First, technology is path-dependent and outside firms with no practical experience of producing good i cannot improve productivity of good i 's production process. In other words, firms with zero Z_i cannot develop knowledge that applies to this activity. (This is the rationale for assuming that R&D is undertaken in-house.) Second, opportunities to exploit spillovers depend on the firm's own knowledge stock. The parameter $0 < \psi < 1$ measures to what extent R&D relies on spillovers as opposed to own knowledge. The assumption of constant returns to scale with respect to the two sources of knowledge ensures endogenous growth.

The firm maximizes the net present value of profits subject to demand (3) and technology (5) and (6). Instantaneous profit is

$$\Pi_i = P_i X_i - L_{Xi} - L_{Zi},$$

where the wage rate is the numeraire. Using (3), (5) and (6) to eliminate X_i , L_{Xi} and L_{Zi} respectively, we write the profit function as:

$$\Pi_i = \pi(P_i, Z_i, \dot{Z}_i; \cdot) \equiv (P_i - Z_i^{-\theta}) \left(\frac{LE}{\int P_k^{1-\varepsilon} dk} \right) P_i^{-\varepsilon} - \dot{Z}_i Z_i^{\psi-1} S_i^{-\psi}, \quad (7)$$

where π is the profit function and the dot represents the variables that the firm takes as given. From the first order condition $\partial \pi / \partial P_i = 0$ we find that the firm's optimal pricing strategy is given by the mark-up rule

$$P_i = \frac{\varepsilon}{\varepsilon - 1} Z_i^{-\theta}. \quad (8)$$

The optimal R&D strategy can be derived in the following intuitive way. Suppose the firm

finances R&D by issuing ownership claims on the flow of profits generated by the marginal innovation. Let the market value of such financial assets be q_i . The firm is willing to undertake R&D up to the point at which the value of the innovation, q_i , equals its cost, $L_{Z_i}/\dot{Z}_i$. Using (6), we obtain

$$q_i = \left(Z_i^{1-\Psi} S_i^\Psi \right)^{-1}. \quad (9)$$

Since the innovation is used in-house, its benefits are determined by the marginal profit it generates and by the reduction in the cost of innovation due to the accumulation of knowledge.

Thus, the value of an innovation must satisfy the arbitrage condition

$$r q_i = \frac{\partial \pi(P, Z_i, \dot{Z}_i; \cdot)}{\partial Z_i} + \frac{d \left(Z_i^{1-\Psi} S_i^\Psi \right)^{-1}}{dt}. \quad (10)$$

Differentiating the profit function in (7) with respect to Z_i , and then substituting (5), we find

$$\frac{\partial \Pi_i}{\partial Z_i} = \frac{\theta L_{X_i}}{Z_i} + (1-\Psi) \frac{\dot{Z}_i}{Z_i} \frac{(Z_i/S_i)^\Psi}{Z_i}. \quad (11)$$

This expression defines the marginal profit from R&D as the contribution of the marginal innovation to reducing production costs and future R&D costs. Taking logs and time derivatives of (9) and substituting (9) and (11) into (10), we find

$$r = \theta (S_i/Z_i)^\Psi L_{X_i} - \Psi \hat{S}_i \equiv r_Z. \quad (12)$$

The right-hand side of this expression defines the rate of return to cost reduction, r_Z . Three

determinants of this return are worth noting. First, r_Z increases with the size of the firm L_{Xi} . This is the cost-spreading effect due to the non-rivalry of knowledge: a larger firm can apply a new idea to a larger volume of production. Second, r_Z decreases with Z_i because the R&D technology exhibits diminishing returns to firm-specific knowledge. Third, r_Z increases with spillovers, as captured by the term $(S_i/Z_i)^\psi$, but decreases with its rate of change, as captured by the term $-\psi \hat{S}_i$. Relatively large spillovers raise the returns to R&D because they raise the firm's R&D productivity (contemporaneous spillover effect); however, firms do not internalize the contribution of spillovers to reducing the future R&D costs of other firms (intertemporal spillover effect).

To keep the analysis tractable, we make the usual assumption of symmetry. All incumbents operate at the same productivity level Z_i . As a result, each firm has the same return on investment, given in (12).⁵

1.3. Entrants

By sorting out and adapting existing knowledge, entrepreneurs create new product lines. They consider entry as long as the sunk cost of entry is lower than the value of the firm,

$$V_i \leq \frac{1}{\beta} \left(\frac{Z_i}{S_i} \right)^\gamma \quad \wedge \quad L_N \geq 0, \quad (13)$$

where V_i is the (post-entry) value of a firm, the right-hand side of the first (in)equality is the entry cost in labour units, and L_N is labour engaged in entry activities. The entry cost increases with the productivity level Z_i at which a new firm is established and decreases with the spillovers S_i available to the entrepreneur. $\beta > 0$ and $\gamma > 0$ indicate, respectively, the level and the elasticity with respect to spillovers of the entry cost function. To preserve symmetry in the model, we

assume that new firms enter with a level of productivity that is the same as incumbents.

The value of a new firm is determined by the following no-arbitrage condition which states that return on investment in such a firm (profits and capital gains) equals the market interest rate r :

$$rV_i = \Pi_i + \dot{V}_i. \quad (14)$$

Profits follow from (5), (7), and (8),

$$\Pi_i = \frac{1}{\epsilon - 1} L_{xi} - L_{zi}, \quad (15)$$

since this is the profit flow that accrues to entrants once they become incumbents. If entrants are active, if $L_N > 0$, the value of a firm V_i equals the sunk cost of entry. Taking logs and time derivatives of (13) and substituting (15) into (14), we find

$$r = \left(\frac{L_{xi}}{\epsilon - 1} - L_{zi} \right) \beta (S_i/Z_i)^\gamma + \gamma (\hat{Z}_i - \hat{S}_i) \equiv r_N. \quad (16)$$

The right-hand side of this expression defines the rate of return to entry, r_N . Four determinants of this rate of return are worth noting. First, r_N increases with the size of the firm L_{xi} because the entry cost is sunk and fixed and a larger firm spreads it over a larger volume of production. Second, r_N decreases with L_{zi} because the cost of incumbency increases with the volume of R&D that the entrant expects to do in equilibrium once it becomes an incumbent. Third, r_N increases with the ratio of spillovers to firm-specific knowledge, S_i/Z_i , because entering a technologically advanced industry is costly but the cost is lower the larger the spillovers that entrants can exploit.

Fourth, r_N increases with the rate of change of firm-specific knowledge and decreases with the rate of change of effective spillovers, as captured by the term $\gamma (\hat{Z}_i - \hat{S}_i)$. Incumbents do not internalize the contribution of firm-specific knowledge accumulation to raising spillovers and therefore to indirectly reducing future entry costs (this is another component of the intertemporal spillover effect). Similarly, they do not take into account the contribution of firm-specific knowledge accumulation to directly raising future entry costs (this implies that incumbents do not try to pre-empt future entry by raising current R&D spending).

1.4. Spillovers⁶

The spillovers that benefit firm i , S_i , are a *weighted sum* of the stocks of firm-specific knowledge accumulated by all firms. To construct this weighted sum, we draw a distinction between the average level of knowledge, Z , and the variety of knowledge sources, N . Firms learn as they solve firm-specific problems that arise when they try to improve the productivity of their own production process. The problems that we have in mind here are of an applied nature, that is, they are well-defined and clearly structured in terms of the practical objective and the solution procedure the R&D team proposes to follow (for an excellent discussion of these issues, see Dosi, 1988). Applied knowledge of this type is structured in broad classes of problems. Within each class, problems have similar solution heuristics and require similar tools. Spillovers occur because when working on their own firm-specific problems, different firms may end up working on problems that belong to the same class. When this occurs, there is a *coincidence of interests* that gives rise to exchange of knowledge.

Imagine that firms work on one problem at the time. As they select which problem to work on, coincidence of interests gives rise to temporary networks of firms that provide each other with useful knowledge. The exchange of knowledge is limited to the firms that participate

in the network. The strength of the spillovers that benefit each firm depends on two variables: (a) the level of knowledge of the typical firm in the network and (b) the number of firms that participate in the network. The former is the firm-specific variable Z discussed above. The latter is our measure of the size of the network, to which we now turn.

When entrants introduce new products, they develop new production processes. To do this, they might formulate and solve problems that are completely new and, often, ill-defined and not clearly structured. When this occurs, they expand the technology space because they introduce new broad classes of problems that open up new possibilities for all firms. Not all entrants need to explore new classes of problems: some new products just involve new combinations of existing ideas. To capture this feature, we posit that the act of entry expands the technology space with some probability. When this occurs, firms have access to a larger menu of problems that they can select to work on.⁷

Let $J(t)$ be the number of classes of problems existing at time t . Whenever a new firm enters, it introduces a new class of problems with probability $0 < \delta < 1$. The number of problem classes then evolves according to $\dot{J} = \delta \dot{N}$, where $\dot{N}$ is the flow of firms entering. Integrating we find

$$J = J_0 + \delta(N - N_0), \tag{17}$$

where J_0 and N_0 are initial conditions. Because there is a constant probability of introducing a new class of problems whenever a new firm enters, the number of locations expands with the number of firms.

Each period, firms choose which type of problems to work on -- and hence the spillovers-network to participate in -- at random from the menu of available classes. In particular, the

probability that any given firm works in network j is $1/J$. It follows that the expected number of firms that meet in location j is $N_j = N/J$. This ratio is our measure of network size. We assume that if firm i operates in the same temporary network as firm k it draws a fraction σ of firm k 's knowledge. Hence, firm i expects to receive

$$E(S_i) = \int_0^{N_j} \sigma Z_k dk = \sigma Z \frac{N}{J}, \quad (18)$$

where we have taken into account the symmetry of the model by writing, $Z_k = Z$. The number of firms that operate within network j is a random variable but, given our assumption that firms visit locations with equal probability, this source of heterogeneity averages out. According to this equation, therefore, all incumbents expect to receive the same expected amount of spillovers regardless of the location they visit. Furthermore, expected spillovers equal realized spillovers because N is a continuum. This feature simplifies the analysis considerably because it preserves the symmetry of the model.⁸

We can now express spillovers as a function of Z and N by substituting (17) into (18) and dropping the expectation operator:

$$S_i = Z_i \cdot \left(\frac{\sigma N}{J_0 + \delta(N - N_0)} \right) \equiv Z_i \cdot s(N), \quad (19)$$

where $s(N)$ denotes spillovers per unit of firm-specific knowledge. The key properties of this spillovers function are that (a) it is linear in average knowledge, Z , (b) it is bounded from above in the number of firms, N , and (c) it may rise or fall with entry. The latter two properties are emphasized by the elasticity of spillovers per unit of knowledge with respect to the number of firms,

$$n(N) \equiv \frac{\partial \log s(N)}{\partial \log N} = \frac{J_0 - \delta N_0}{J_0 + \delta (N - N_0)}. \quad (20)$$

This elasticity is positive (negative) if $\delta < (>) J_0/N_0$, and goes to zero as the number of firms goes to infinity.

Entry affects spillovers through two channels. First, a new firm raises average spillovers in a network, since with more firms there are more sources of knowledge to learn from. This *knowledge expansion effect* is proportional to the number of firms, N . Second, the partition of technology in classes of problems becomes finer and it is more difficult to absorb the knowledge of other firms because the probability that any two firms work on similar problems decreases. This *knowledge differentiation effect* is captured by an increase in J . If $\delta < J_0/N_0$, entrants mostly join existing networks and network size increases with entry, thus stimulating spillovers. If $\delta > J_0/N_0$, in contrast, entrants mostly create new networks and network size falls with entry.

If the number of firms is very large, a large number of networks is operating. Then, the probability that a given inframarginal firm meets the marginal firm while working on technologically related problems converges to zero as the number of firms grows large (that is, n in (20) approaches zero). More importantly, the marginal contribution of the number of firms to spillovers falls asymptotically to zero because the knowledge expansion effect exactly offsets the knowledge differentiation effect.⁹

It is insightful to interpret the number of networks, J , as a measure of the average *technological distance* between an arbitrary pair of firms. An increase in J makes firms more distant on average because it reduces the probability that they participate in the same network. Loosely speaking, we can then say that average technological distance increases with the number of firms. Moreover, because the probability that the marginal firm is technologically infinitely distant from the typical inframarginal firm (i.e., that it operates in a different network) converges

to one asymptotically, the number of firms in a given network is bounded from above. Hence, the size of the network within which a firm operates grows less than proportionally to the number of firms and eventually becomes independent of it.

In the appendix we extend the present set up. Firms may choose to work on more than one class of problems and thus benefit from knowledge spillovers from several spillover networks at the same time. By making the natural assumption that the effectiveness of knowledge exchange on one particular problem falls with the number of problems a firm works on (holding fixed total research effort), the same conclusions regarding spillovers and entry emerge. The number of networks a firm optimally chooses to participate in grows at a slower pace than the number of firms and spillovers become bounded.

2. General Equilibrium and Steady State

2.1. General Equilibrium

Goods market equilibrium implies that total spending equals the value of total supply. Taking into account (5) and (8) we find

$$LE = P_i N X_i = \frac{\varepsilon}{\varepsilon - 1} N L_{Xi}. \quad (21)$$

Labour devoted to entry is

$$L_N = \frac{1}{\beta} \left(\frac{Z}{S} \right)^\gamma \dot{N} = \frac{\dot{N}}{\beta s(N)^\gamma}. \quad (22)$$

Total supply of labour is fixed at L . Labour market equilibrium requires

$$NL_{Xi} + NL_{Zi} + \frac{1}{\beta \cdot s(N)^\gamma} \dot{N} = L. \quad (23)$$

Capital market equilibrium is determined by the interaction of households, incumbents and entrants. Households are willing to postpone consumption as long as the rate of return is large enough. The required rate of return to savings follows from (4) and (21),

$$r = \rho + \hat{L}_{Xi} + \hat{N}. \quad (24)$$

Arbitrage between the returns to productivity improvement and to creation of new product lines ensures that either both incumbents and entrants obtain resources against the rate of return that households require, or that only the group that offers the higher rate of return obtains them. The rate of return that incumbents and entrants offer are given by (12) and (16). Using (19), we can express them as:

$$r_Z = \theta s(N)^\psi L_{Xi} - \psi \hat{Z}_i - \psi n(N) \hat{N}, \quad (25)$$

$$r_N = \frac{\beta}{\varepsilon - 1} s(N)^\gamma L_{Xi} - \beta s(N)^{\gamma - \psi} \hat{Z}_i - \gamma n(N) \hat{N}, \quad (26)$$

where $n(N)$ is the elasticity of spillovers, s , with respect to the number of firms, N , that we derived in the previous section.

2.2. Steady State

We focus on steady states where R&D drives growth while entry peters out (the appendix deals with transitional dynamics). With a constant allocation of labour across activities, constant productivity growth obtains because accumulation of firm-specific knowledge is linear in Z ; combine (6) and (18) to see this. However, the returns to entry fall as entry takes place. Combining (22) and (18), we find

$$\hat{N} = \beta L_N s(N)^\gamma / N. \quad (27)$$

As long as $n\gamma < 1$, the rate of entry is decreasing in N . This condition is satisfied for a large number of firms because n approaches zero as N grows large; see (20). The steady state is therefore characterized by constant values of production per firm, L_{xi} , R&D per firm, L_{zi} , and the number of firms, N . Economic growth is captured by the rate of productivity growth,

$$g = \theta \dot{Z}/Z = \theta (S/Z)^\psi L_{zi} = \theta s(N)^\psi L_{zi}. \quad (28)$$

Moreover, capital market equilibrium requires $r = \rho$; see (24).

Whenever growth is positive, the rate of return to R&D equals the required rate of return, ρ . After substitution of $r_z = \rho$, $\hat{L}_{xi} = 0$ and $\hat{N} = 0$ into (23), (24), (25) and (28), this condition can be written

$$g = \frac{\theta}{\theta + \psi} \left[\frac{\theta s(N)^\psi L}{N} - \rho \right] \equiv g^{LL}(N). \quad (29)$$

This equation describes the economy's resources constraint. We label it LL. No-arbitrage in the capital market requires that entry and R&D yield the same rate of return, which in steady state

equals the required rate of return, $r_Z = r_N = \rho$. Substituting $\hat{N} = 0$ and (28) into (25) and (26), solving (25) for $\theta s^\psi L/N$, and substituting the resulting expression into (26), we find

$$g = \frac{\theta \rho}{\beta[\psi - \theta(\epsilon - 1)]} \left[\theta(\epsilon - 1) \cdot s(N)^{\psi - \gamma} - \beta \right] \equiv g^{NN}(N). \quad (30)$$

This equation describes the capital market no-arbitrage condition for incumbents and entrants and summarizes the industry's Nash equilibrium when the number of firms is determined endogenously. We label it the NN curve. Note that it slopes upward if $n(N) \cdot (\psi - \gamma) > 0$ and downward if $n(N) \cdot (\psi - \gamma) < 0$.

Since we want to study the effect of entry on the growth rate, we focus on the situation in which both entry and innovation opportunities determine the steady state. In other words, we rule out that one type of investment (entry or cost reduction) dominates the other for all states of the economy. In the Appendix we show that such a steady state exists if both expressions in square brackets in (30) are positive.

Fig. 1 characterizes growth in steady state. Equations (29) and (30) are depicted as the NN and LL curves in the (N, g) plane. The intersection of the two curves at (N^*, g^*) represents a steady state in which the number of firms is determined endogenously. The economy converges gradually to this steady state if it starts out with an initial number of firms smaller than N^* . The bold segments represent continua of steady states where the returns to entry fall short of the returns to R&D and entrants cannot obtain resources in the capital market. As a result, the number of firms is fixed and equilibrium is determined only by R&D opportunities and the economy's resources constraint. These points are described by the LL curve. When the economy starts out with an initial number of firms larger than N^* , there is no transitional dynamics and the model exhibits hysteresis; see the Appendix. Moreover, if the initial number of firms is larger

than $\bar{N}$, growth is zero because firm size is too small and the returns to R&D fall below the required rate of return, ρ .

INSERT Fig. 1. *Steady-state equilibria and the scale effect*

3. The Scale Effect

The scale effect can be negative or positive but always vanishes as the economy becomes large. This is illustrated in Fig. 1. An increase in the scale of the economy, L , affects only the LL curve that shifts out and traces the NN curve, which is upward or downward sloping depending on the sign of $n(N) \cdot (\psi - \gamma)$. More importantly, the NN curve approaches a horizontal asymptote so that growth is bounded and the scale effect vanishes.

An increase in the size of the economy increases the returns to R&D because knowledge is non-rival. This means that the cost of producing an idea is independent of the scale of production to which it is applied (R&D is a fixed cost). This basic cost-spreading mechanism is captured by the upward shift of the LL curve: holding constant N , growth increases with the scale of the economy. Since this effect is driven by firm's perception of a larger market, we call it the *market-size effect*. However, a larger market attracts entry because the value of the firm increases. Entry of new firms generates (a) *dispersion* of resources over a larger number of firms, (b) *expansion* of the aggregate public knowledge stock, and (c) *differentiation* of knowledge which raises average technological distance. The balance of these forces ultimately determines whether the scale effect is positive or negative. First, a larger number of firms implies a smaller size of the firm. Labour is dispersed over more business units so that, while the scale of the total

economy increases, individual firms may become smaller. Since knowledge is firm-specific and the size of the firm determines the returns to R&D, entry mitigates the market-size effect by limiting firm's ability to exploit the cost-spreading mechanism driving R&D decisions. As a result, R&D effort per firm falls. Second, entry expands the aggregate stock of public knowledge because entrants create new product lines that bring new sources of knowledge to the industry. This raises the returns to R&D and reinforces the market-size effect. However, and third, the arrival of new classes of problems increases the specialization of research and limits the extent to which the larger aggregate stock of public knowledge generates spillovers that benefit the individual firm.

An important property of the model is that the dispersion of resources over a larger number of firms is never fully offset by stronger spillovers. The reason is that a one percent increase in the number of firms reduces firm size by one percent but raises spillovers by less than one percent. In other words, if the economy disperses resources over more product lines, it also dilutes the aggregate public knowledge stock. As a result, entry of specialized firms dampens or even reverses the market-size effect.¹⁰ If the probability that a new firm brings knowledge that is useful to a specific incumbent is very small, because new firms are very likely to join a different network (J is very large), then the knowledge differentiation effect offsets the knowledge expansion effect. In the limit, spillovers are not affected by entry because the entrant is not likely to exchange knowledge with the typical incumbent. In this case, there is no scale effect because the dispersion effect exactly offsets the market-size effect. The reason is that the returns to R&D and entry now depend only on firm-level variables, see (25) and (26), and no longer on the number of firms. An increase in L , therefore, raises firm size, L_{xi} , which triggers higher R&D and entry of new firms. Entry in turn causes rates of return to fall gradually via the dispersion effect. The steady-state rate of return, $r_N = r_Z = \rho$, obtains when L_{xi} and L_{zi} are back

to their pre-shock levels. Hence, entry stops when firms are of the same size as before the expansion of the market.

Dilution of the aggregate stock of public knowledge is related to the degree of specialization in the economy. The extent of the market determines the degree of specialization, as Adam Smith argued a long time ago. In our model this applies not only to specialization in production but also to specialization in research. If L increases, new firms enter, each following its own firm-specific technological trajectory and offering its own differentiated product. Consumers have love-of-variety preferences and, all else equal, benefit from higher product market specialization.¹¹ Specialization in research, on the other hand, means that technological distance increases on average, which implies that on average firms learn less from each other. As a result, although there is a larger aggregate stock of public knowledge that firms might exploit, their ability to do so is actually reduced and spillovers do not rise proportionally. If the increase in the size of the market triggers a very large increase in knowledge differentiation, spillovers may even fall.

The sign of the scale effect depends on the sign of the term $n(N) \cdot (\Psi - \gamma)$, which determines the slope of the NN curve. If $n(N) > 0$, the knowledge expansion effect dominates the knowledge differentiation effect and spillovers increase with entry. In this case, entrants and incumbents benefit from stronger spillovers. Entrants benefit relatively more if $\gamma > \Psi$, in which case investors shift their financial support from incumbents to entrants and the additional amount of labour is mostly allocated to entry. In the new steady state, firms are smaller and growth is slower (see the right-hand side panel of Fig. 1). In contrast, if creation of new firms does not rely much on spillovers, while cost reduction by established firms does, i.e. if $\Psi > \gamma$, then the additional amount of labour is mainly allocated to existing firms. In steady state, firm size is larger and supports higher R&D per firm and faster growth (see the left-hand side panel of Fig.

1). These results are reversed if $n(N) < 0$ because in this case spillovers *fall* with entry. Thus, if $\psi > \gamma$ incumbents are hurt by entry more than entrants and growth falls with the size of the economy (see again the right-hand side panel of Fig. 1). The reverse happens if $\psi < \gamma$.

It is interesting to look at the determinants of the growth rate for a very large economy. If L approaches infinity, s approaches σ/δ (see 19). Substituting this result into (30), we obtain a closed-form solution for the growth rate. Then (29) solves for N . The number of firms increases one-for-one with L .¹² Most comparative statics results with respect to the growth rate are similar to those of other endogenous growth models. An increase in β makes entry more profitable, reduces average firm size and hurts long-run growth. The effects of knowledge spillovers and knowledge differentiation, parameterized by σ and δ respectively, are specific to this model and are worth mentioning. Increases in σ and δ have opposite effects, the direction of which depends on the sign of $\psi - \gamma$. For reasons mentioned above, an increase in spillovers σ only increases growth if spillovers stimulate R&D relatively more than entry, i.e. if $\psi > \gamma$.

4. Conclusion

We presented a growth model where dilution of public knowledge causes the scale effect to vanish asymptotically. The mechanism behind this result is increasing technological distance. As the economy becomes larger, because the workforce increases, entrants introduce new, differentiated technologies. As a result, the aggregate stock of public knowledge rises but, because firms become more specialized, the spillovers that each individual firm exploits in its own R&D activity rise less than proportionally or may even fall. This property highlights the importance of the interaction between innovation and market structure dynamics in a class of

growth models that have been proposed recently as solutions to the problem of scale effects. In our model, favorable conditions for entry, which are determined by the larger scale of the economy, imply a significant change in the structure of the market and technology. If entry raises spillovers and if entrants benefit from them while incumbents rely mainly on firm-specific knowledge, then the scale effect is negative. The scale effect is positive, in contrast, if incumbents benefit from spillovers more than entrants. Regardless of its sign, moreover, the scale effect always vanishes when the economy becomes very large.

A crucial assumption in our model is that new entrants tend to introduce new knowledge that makes the total knowledge stock more diverse and more specialized. Hence, we model the incremental process of increasing complexity in an expanding economy. In reality, entry can sometimes imply the introduction of knowledge that brings firms together. If new knowledge has the character of a general purpose technology (GPT), like the principles of a new energy system or a new way to organize R&D, many firms may change the direction of their research in order to absorb this knowledge and to take advantage of the GPT. General purpose technologies might thus decrease technological distance and temporarily disrupt the process of knowledge differentiation. Models that deal explicitly with GPTs stress that they lead to a temporary period of adjustment during which the economy adapts to the new GPT (Helpman and Trajtenberg, 1998). We have focussed on how incremental innovation and trend growth react to the expansion of the market, rather than on the effect of discrete arrivals of radically different technologies. Nevertheless, our model may provide a framework for studying GPTs. At least two changes are required. First, the GPT must be modeled as an unexpected shock to the structure of the knowledge stock: it may suddenly bring together existing fields of research and thus effectively reduce the number of classes of problems (J) on which firms do research. This raises the rate of return to entry and R&D. Second, as is common in the GPT literature, an adjustment cost

associated to swiching to the new research direction must be added. Either research itself is temporary less productive (research have to spend time redirecting research) or the first fruits from the new direction in research underperform relative to old products and processes. Together, the two effects produce the usual “wave” in innovation and growth, but the model may yield new insight in how entry of new firms is affected by the GPT and how this mitigates its effect on growth.

Another property of the model is that we enrich the structure of similarities and differences among technologies, but keep fixed the degree of substitution of goods on the user side. In principle, the technology to produce a good and the services it yields to users are separate concepts. Yet, product characteristics are likely to change with the process of innovation. It is, however, not clear whether more differentiated technologies make goods more or less substitutable. To address this question, one needs to further model the microeconomics of production, research and knowledge spillovers, and take a stand on how they relate to users’ demands. This requires a structural model in which producers endogenously choose product characteristics and trade off competition from substitute goods against technology spillovers (as in Weitzman, 1994). This would take us beyond the Dixit-Stiglitz model that has proven so powerful and tractable in the analysis of how market structure affects technological innovation and economic growth.

Duke University

Tilburg University

Date of receipt of first submission:

Date of receipt of final typescript:

Appendix A: Dynamics

In this section we study the model's general equilibrium dynamics. We first study firm behaviour and determine the conditions under which entrants and/or incumbents obtain resources to invest. Next, we consider household behaviour and determine the supply of resources. Finally, we combine firm and household behaviour and draw the phase diagram for the economy in the (N, NL_{xi}) plane. To simplify things, we introduce the following notation: $M \equiv NL_{xi}$ (total Manufacturing labour) and $b \equiv \beta s(N)^{\gamma-\psi}$. We simply write s and n for $s(N)$ and $n(N)$; so it should be kept in mind that b , s , and n depend on N . We show that a unique transition path towards a steady state with R&D exists under the following assumptions:

Assumption 1: $b < \theta(\epsilon - 1) < \psi$,

Assumption 2: $b < \gamma$.

These assumptions require comment. In the case $n \cdot (\gamma - \psi) > 0$, they reduce to a simple restriction on the parameters because the function $b \equiv \beta s(N)^{\gamma-\psi}$ is increasing and bounded from above; $s(N)$ converges to σ/δ as N becomes large. In the case $\gamma > \psi$, they require that N be larger than a critical value determined by $\beta s(N)^{\gamma-\psi} = \min \{ \gamma, \theta(\epsilon - 1) \}$.

We begin with a characterization of the instantaneous Nash equilibrium in R&D and entry strategies. If both R&D and entry take place, no-arbitrage requires $r_Z = r_N = r$, where r is the market interest rate. From (25) and (26) we can write:

$$\hat{Z} = \frac{1}{\psi} \left(\theta s^\psi \frac{M}{N} - r \right) - n\hat{N}, \quad (\text{A1})$$

$$\hat{Z} = \frac{1}{b} \left(\frac{bs^\psi}{\varepsilon-1} \frac{M}{N} - r \right) - \frac{\gamma}{b} n\hat{N}. \quad (\text{A2})$$

These two curves can be interpreted as the reaction functions of, respectively, incumbents and entrants. Since they are downward sloping, R&D and entry are strategic substitutes. Intuitively, entry fragments the market and depresses the incentives to spend on R&D; on the other hand, R&D increases the cost of incumbency and depresses the incentives to enter. The Nash equilibrium exists and is stable if (A2) intersects (A1) from above; see Fig. 2. This requires

$$[\psi - \theta(\varepsilon-1)]b \frac{s^\psi M}{N} > (\psi-b)(\varepsilon-1)r. \quad (\text{A3})$$

which implies that the intercept of (A2) is higher than that of (A1), and

$$[\gamma\theta(\varepsilon-1) - \psi b] \frac{s^\psi M}{N} > (\gamma-\psi)(\varepsilon-1)r, \quad (\text{A4})$$

which implies that (A2) intersects the horizontal axis at a point to the left of the one where (A1) intersects it. Assumption 2 ensures that (A2) is steeper than (A1).

INSERT Fig. 2. *Nash equilibria (a) with entry and R&D, (b) no entry, (c) no R&D.*

An equilibrium with no entry results if (A3) is violated; see Fig. 2b. We then have $\hat{N} = 0$ and $r = r_Z$. Substituting these results in (A3) with reversed inequality sign, and using (25) and the resources constraint $L = M + \hat{Z}N/d^\psi$, we find that whenever

$$M [(\psi - b)(\epsilon - 1) + \theta(\epsilon - 1) - b] > (\psi - b)(\epsilon - 1)L \quad (\text{A5})$$

entry is zero and only R&D can be positive. An equilibrium with no R&D results if (A4) is violated; see Fig. 2c. We then have $\hat{Z} = 0$ and $r = r_N$. Substituting these results in (A4) with reversed inequality sign, and using (25) and the resource constraint $L = M + \hat{N}N/d^\psi b$, we find that whenever

$$M [(\psi - \gamma)(\epsilon - 1)nb + \theta(\epsilon - 1) - b] < (\psi - \gamma)(\epsilon - 1)nbL \quad (\text{A6})$$

R&D is zero and only entry can be positive.

Conditions (A5) and (A6) define, respectively, the no-entry region and the no-R&D region in the (N, M) plane. $M \in [0, L]$ defines the feasible region.

It is now clear why we need Assumption 1. Condition (A6) shows that we need $b < \theta(\epsilon - 1)$ for R&D to be positive in steady state. Condition (A5) then shows that we need $b < \psi$ to have entry. Condition (A3), finally, shows that we need $\theta(\epsilon - 1) < \psi$ to have both R&D and entry. Fig. 3 depicts the regions that violate these restrictions as shaded areas (note that the no-R&D region disappears if $\gamma > \psi$).

INSERT Fig. 3. *Phase diagrams*

We now turn to household behavior, which is characterized by $\hat{M} = r - \rho$; see (24). To find the $\hat{M} = 0$ locus in the (N, M) plane, we eliminate r . If both entry and R&D occur, we set $r = r_Z = r_N$ and use (25), (26) and (23) to derive

$$\hat{M} = \left[\frac{\psi - \theta(\varepsilon - 1)}{(\psi - b)(\varepsilon - 1)} + \psi n \left(\frac{\gamma - b}{\psi - b + (\gamma - \psi)nb} \right) \left(\frac{(\psi - b)(\varepsilon - 1) + \theta(\varepsilon - 1) - b}{(\psi - b)(\varepsilon - 1)} \right) \right] \frac{bs^\psi}{N} M$$

$$- \left[\psi n \left(\frac{\gamma - b}{\gamma - b + (\gamma - \psi)nb} \right) \frac{bs^\psi}{N} L + \rho \right].$$

From this equation, we obtain the $\hat{M} = 0$ locus outside the no-entry and no-R&D regions,

$$M = (\varepsilon - 1) \left(\frac{\psi n L (\gamma - b) + \rho \frac{N}{bs^\psi} [\psi - b + (\gamma - \psi)nb]}{\psi - \theta(\varepsilon - 1) + n[(\theta + \psi)\gamma(\varepsilon - 1) - \psi \varepsilon b]} \right). \quad (A7)$$

For N sufficiently large, the locus is upward sloping (since n becomes small if N becomes large).

Above (below) the locus, M rises (falls).

The $\hat{M} = 0$ locus in the no-entry region results from $r = r_Z$, $\hat{N} = 0$, (25) and (23),

$$M = \frac{\psi L + \rho N s^{-\psi}}{\theta + \psi}. \quad (A8)$$

This curve is upward sloping and intersects (A7) on the border of the no-entry region.

The $\hat{M} = 0$ locus in the no-R&D regime results from $r = r_N$, $\hat{Z} = 0$, (26) and (23),

$$M = \frac{\gamma n L + \rho N / bs^\psi}{1/(\varepsilon - 1) + \gamma n}. \quad (A9)$$

This curve has a similar shape as (A7). If (A9) is inside the no-R&D region and if it intersects the border of the no-R&D region, then (A9) and (A7) intersect at a point on this border; this requires $\psi b > \gamma \theta(\varepsilon - 1)$.

Finally, we have to consider the situation with neither entry nor R&D. The resources constraint in this case is $M = L$, which yields $\hat{M} = \hat{L} = 0$. This is consistent with (24) only if $r = \rho$ and $\max \{r_Z, r_N\} < \rho$ for $\hat{N} = \hat{Z} = 0$. From (25) and (26), we see that this requires $N > \bar{N}$, where $\bar{N}$ is defined by $\theta_s(\bar{N})^\psi L/\bar{N} = \rho$. Hence, the $\hat{M} = 0$ locus for $N > \bar{N}$ reduces to $M = L$. Note that (A8) and $M = L$ intersect at $N = \bar{N}$.

The phase diagrams in Fig. 3 depict the borders between the regions defined by (A5) and (A6). The $\hat{M} = 0$ locus connects (A9) to (A7) to (A8) and to the $M = L$ line in the relevant region. The arrows indicate the motion of N and M . All unstable trajectories are ruled out because they eventually violate the transversality conditions for firms and/or households or the resources constraint. There is thus a unique perfect-foresight, general-equilibrium path. If the initial number of firms is to the left of N^* , then the economy jumps on the saddle path and converges to the steady state (N^*, M^*) ; if the initial number of firms is to the right of N^* , then the economy jumps on the $\hat{M} = 0$ locus and enters immediately a steady state with positive growth if $N < \bar{N}$, or zero growth if $N \geq \bar{N}$.

Appendix B Endogenizing the Number of Networks a Firm Participates in

We now provide an extension of the basic setup that allows firms to choose the number of problems they work on. Let $A_i \leq J$ denote the number of networks the firm decides to participate in. Let

$$e\left(\frac{L_{Zai}}{L_{Zi}}\right), \quad e'(\cdot) > 0,$$

denote the effectiveness of knowledge exchange as a function of the share of R&D resources that the firm allocates to location a , where $a \in [0, A_i]$. The firm then receives spillovers

$$S = \int_0^{A_i} \left(\int_0^{N_{ai}} e \left(\frac{L_{Zai}}{L_{Zi}} \right) Z_k dk \right) da.$$

We model the choice of the number of networks as

$$\max_{A_i} S_i \text{ subject to } \int_0^{A_i} L_{Zai} da = L_{Zi},$$

where L_{Zi} is the firm's R&D budget determined in the main text. If the firm allocates R&D employment in equal shares across networks, we can write

$$S = \int_0^{A_i} \left(\int_0^{N_a} e \left(\frac{1}{A_i} \right) Z_k dk \right) da = A_i e \left(\frac{1}{A_i} \right) Z \frac{N}{J},$$

and the optimal choice of A_i requires

$$\frac{e'(1/A_i) \cdot (1/A_i)}{e(1/A_i)} = 1.$$

Suppose this expression has a solution $A_i^* < J$, where A_i^* is a number that depends only on the properties of the function $e(\cdot)$, and therefore is the same for all firms. Then the firm does not participate in all existing networks, the effectiveness of knowledge exchange reduces to $e(1/A_i^*)$, and we can write $S = [A_i^* \cdot e(1/A_i^*)] \cdot ZN/J$. Hence, spillovers are proportional to the network size N/J as in the main text with exogenous research portfolio size (note that the term in brackets replaces σ). It might happen that the economy starts out with $J_0 < A^*$. Then firms choose $A_i = J$, that is they work on all problems. However, as N grows, J becomes large and eventually exceeds A^* which makes firms choose A_i^* again. Thus, our asymptotic result that in a very large economy effective spillovers are bounded and become proportional to average firm-specific knowledge

simply requires that there exist some finite A_i^* that solves the above first order condition. If the finite value $A^* < J$ does not exist, then the firm sets $A = J$ and visits all locations. This yields $S = e(1/J)ZN$. In this case, the property that effective spillovers are bounded in N is preserved if the term $e(\cdot)N$ converges to a constant. This requires that the function $e(\cdot)$ be asymptotically linear.

The difference between the basic setup and this extension sheds further light on the mechanism that drives the model. In the basic setup, bounded spillovers follow from the fact that the firm interacts with a subset of all existing firms because we “forced” it to work on only one class of problems. In this extension, in contrast, it follows from the fact that the function $e(\cdot)$ implies that as the firm disperses resources over too many problems, its research becomes unfocused and less productive. More generally, it is optimal for the firm to narrow the scope of its R&D effort, work on a subset of all possible problems, and therefore interact with only a subset of all existing firms, whenever knowledge becomes so differentiated that the menu of possible problems is too large.

References

- Adams, J.D., and Jaffe, A.B. (1996). ‘Bounding the effects of R&D: an investigation using matched establishment-firm data.’ *Rand Journal of Economics*, vol. 27, pp. 700-721.
- Aghion, Ph. and Howitt, P. (1990). ‘A model of growth through creative destruction.’ *Econometrica*, vol. 60, pp. 323-352.
- Aghion, Ph., and Howitt, P. (1998). *Endogenous Growth Theory*. Cambridge: MIT Press.

- Backus, D., Kehoe, P., and Kehoe, T. (1992). 'In search of scale effects in trade and growth.' *Journal of Economic Theory*, vol. 57, pp. 377-409.
- Cohen, W., and Klepper, S. (1996). 'A reprise of size and R&D.' *ECONOMIC JOURNAL*, vol. 106, pp. 925-951.
- Dinopoulos, E. and Thompson, P. (1998). 'Schumpeterian growth without scale effects.' *Journal of Economic Growth*, vol. 3, pp. 313-335.
- Dosi, G. (1988). 'Sources, procedures, and microeconomic effects of innovation.' *Journal of Economic Literature*, vol. 36, pp. 1120-1171.
- Geroski, P. (1991). 'Entry and the rate of innovation.' *Economics of Innovation and New Technology*, vol 1, pp. 203-214.
- Griliches, Z. (1990). 'Patent statistics as economic indicators: a survey.' *Journal of Economic Literature*, vol. 28, pp. 1661-1707.
- Griliches, Z. (1992). 'The search for R&D spillovers.' *Scandinavian Journal of Economics*, vol. 94 (supplement), pp. S29-S47.
- Grossman, G. and Helpman, E. (1991). *Innovation and Growth in the Global Economy*. Cambridge: MIT Press.

Helpman, E. (1998). *General Purpose Technologies and Economic Growth*. Cambridge: MIT press.

Helpman, E. and Trajtenberg, M. (1998). 'A time to sow and a time to reap: growth based on general purpose technologies.' In Helpman (1998), pp. 55-84.

Howitt P. (1999). 'Steady endogenous growth with population and R&D inputs growing.' *Journal of Political Economy*, vol. 107, pp. 715-730.

Jaffe, A.B. (1986). 'Technological opportunity and spillovers of R&D: evidence from firms' patents, profits, and market value.' *American Economic Review*, vol. 76, pp. 984-1001.

Jones, C.I. (1995). 'Time series test of endogenous growth models.' *Quarterly Journal of Economics*, vol. 110, pp. 495-525.

Malerba, F. and Orsenigo, L. (1995). 'Schumpeterian patterns of innovation.' *Cambridge Journal of Economics*, vol.19, pp. 47-65.

Malerba, F., Orsenigo, L. and Peretto, P. (1997). 'Persistence of innovative activities, sectoral patterns of innovation, and international technological specialization.' *International Journal of Industrial Organization*, vol. 15, pp. 801-826.

Peretto, P. (1998a). 'Technological change, market rivalry, and the evolution of the capitalist engine of growth.' *Journal of Economic Growth*, vol. 3, pp. 53-80.

Peretto, P. (1998b). 'Technological change and population growth.' *Journal of Economic Growth*, vol. 3, pp. 283-311.

Peretto, P. (1999). 'Cost reduction, entry, and the interdependence of market structure and economic growth.' *Journal of Monetary Economics*, vol. 43, pp. 173-195.

Romer, P. (1990). 'Endogenous technological change.' *Journal of Political Economy*, vol. 98, pp. S71-S102.

Smulders, S., and van de Klundert, T. (1995). 'Imperfect competition, concentration and growth with firm-specific R&D.' *European Economic Review*, vol. 39, pp. 139-160.

Weitzman, M.L. (1994). 'Monopolistic competition with endogenous specialization.' *Review of Economic Studies*, vol. 61, pp. 45-56.

Young, A., 1998, "Growth without scale effects", *Journal of Political Economy*, 106, 41-63.

Footnotes

* We thank Jan Boone, Michelle Connolly, Greg Crawford, Wolfgang Keller, Kent Kimbrough, Theo van de Klundert, Georg Goetz, Richard Nahuis, Etsuro Shioji, Jonathan Temple, Jeroen van de Ven, and two anonymous referees for useful comments on an earlier version. Smulders' research was financed by the Dutch Science Foundation (NWO) and the Royal Netherlands

Academy of Arts and Sciences.

1. Malerba and Orsenigo (1995) and Malerba *et al.* (1997) show that industries in which persistent innovation by existing firms is the dominant mode of technological advance are much more numerous than industries characterized by creative destruction whereby innovations are typically introduced by new firms.

2. Geroski (1991) documents the relation between innovation and entry. After controlling for fixed industry effects, he finds that in the time-series dimension entry has a negative effect on innovation while innovation has no effect on entry. Hence, innovation does not open opportunities for entry while entry crowds the market and depresses the incentives to innovate. This supports our view that explicit consideration of market structure requires to abandon the view that there is a one-to-one correspondence between the formation of new firms and innovation. Geroski also finds that across industries innovation is correlated with entry. Our model predicts the same: since entrants build on the same knowledge stock as incumbents, industries with large spillovers provide rich innovation opportunities for incumbents and low barriers to entry for entrants.

3. Alternatively, (2) can be interpreted as the production function of the single final good with C_i as intermediates.

4. Since cost-reducing and quality-improving innovations are formally similar, we focus on the former with the understanding that our results apply also to the latter.

5. Peretto (1998a, 1998b) discusses in detail conditions under which a symmetric equilibrium obtains in models of this class.

6. In working on this section, we benefited substantially from discussions with Greg Crawford and John Seater.

7. There is a fundamental reason why we posit that only entrants introduce new classes of

problems. Incumbents do research that is focussed on *improving existing* production processes. Entrants, in contrast, do research that is focussed on *creating new* production processes. This implies that entrants are much more likely to challenge the established research approaches and heuristics and try something entirely new. Unlike incumbents, they are not constrained in their innovation efforts by an already existing technology that generates problems that are well-defined, focussed and highly structured.

8. This symmetry does not mean that firms are identical: each firm has a qualitatively different knowledge stock because it has a unique and idiosyncratic history of learning and interaction with specific groups of other firms.

9. The scale effect does not vanish only in the knife-edge case $\delta = 0$, which implies that network size is proportional to the number of firms. Another special case is $\delta = J_0/N_0$, which yields constant network size and no effect of entry on spillovers. This is the special case of zero scale effect discussed in the papers cited in the introduction.

10. This mechanism cannot apply in models where the rate of entry *is* the rate of innovation. In the variety-expansion models of Romer (1990) and Grossman and Helpman (1991, Ch. 3), for example, positive growth requires positive entry rates and a balanced growth path obtains only if the number of firms does not affect the returns to innovation-entry. This requires that the dispersion effect exactly offsets the knowledge expansion effect (the knowledge differentiation effect is zero by construction). As a result, there is nothing left to sterilize the market-size effect.

11. We could reinterpret the consumption index C in (1) as consumption of a homogenous final good produced by means of intermediate inputs according to the technology (2). In this case, an increase in specialization implies higher TFP.

12. As discussed in the Introduction, this is the condition that has to be imposed in Jones's (1999) model to remove the scale effect. It arises endogenously from the microeconomic structure of our

model when L is large.

Figure titles

Fig. 1. *Steady-state equilibria and the scale effect*

Fig. 2. *Nash equilibria (a) with entry and R&D, (b) no entry, (c) no R&D. Nash equilibria*

Fig. 3. *Phase diagrams*

(a) $n(N)(\psi - \gamma) > 0$

(b) $n(N)(\psi - \gamma) < 0$

Fig. 1. *Steady state equilibria and the scale effect*

Fig. 2. Nash equilibria (a) with entry and R&D, (b) no entry, (c) no R&D.

Fig. 3. *Phase diagram*