

Apresentação

José Gabriel Porcile Meirelles

Departamento de Economia, Universidade Federal do Paraná

A importância da obra de Richard Nelson e Sidney Winter para o estudo da inovação, da concorrência a partir da geração e difusão de inovações, e da conseqüente transformação das firmas e estruturas de mercado, é difícil de exagerar. Surge a partir desses autores uma nova abordagem teórica, de inspiração schumpeteriana, que permite articular, de forma consistente, um conjunto de contribuições até então fragmentárias e dispersas, e abrir um horizonte de novas pesquisas teóricas e empíricas. Nesse processo, o artigo republicado neste número da **RBI** tem uma posição de destaque. Com efeito, Nelson e Winter desenvolvem nele as bases daquela que mais tarde será conhecida como teoria evolucionista da inovação, apresentando de forma já amadurecida alguns elementos centrais de seu livro clássico de 1982.¹

Primeiramente, o artigo aponta que a teoria sobre inovação disponível até aquele momento, baseada na distinção entre movimentos ao longo da função de produção e deslocamentos dessa função, era incompatível com as evidências empíricas obtidas a partir de um grande número de estudos sobre inovação e crescimento da produtividade, em nível da firma e dos distintos setores da economia. Tornava-se necessário articular essas evidências numa nova perspectiva, que permitisse definir um caminho consistente, rigoroso e cumulativo de pesquisa. Esse caminho estaria balizado por dois elementos-chave: 1) a necessidade de considerar o processo de geração e difusão de inovações como um processo estocástico, no qual agentes heterogêneos

¹ Nelson, R. R.; Winter, S., *An Evolutionary Theory of Economic Change*, The Belknap Press of Harvard University Press, 1982.

tomam decisões num ambiente de incerteza; *ii*) a necessidade de considerar o processo competitivo como caracterizado pelo desequilíbrio e por uma considerável complexidade e diversidade institucional.

Não é possível tratar aqui as muitas implicações desses elementos da nova teoria, que representam um corte claro com dois pilares da tradição neoclássica — a maximização como regra de decisão e o equilíbrio como ponto de chegada dos agentes e do conjunto das variáveis do sistema. Mas vale a pena mencionar duas implicações que são particularmente relevantes, a saber: *i*) a recuperação do conceito de racionalidade limitada e a plena incorporação desse conceito à dinâmica da concorrência; *ii*) a preocupação com a forma em que a diversidade institucional afeta os mecanismos de geração e difusão de inovações.

Racionalidade limitada e transformação industrial

No que tange à racionalidade dos agentes, não era novo entre os economistas o desconforto com relação ao suposto de maximização de alguma função-objetivo (lucro, utilidade) como representação adequada do processo de decisão dos agentes econômicos.² Adicionalmente, o suposto de existência de um agente representativo implicava na eliminação *ex hipotese* de qualquer vestígio de diversidade e heterogeneidade entre os agentes. Para Nelson e Winter, esses dois supostos eram especialmente indefensáveis no contexto da inovação, onde por definição não existe um conjunto perfeitamente conhecido e delimitado de opções e de resultados *ex ante*. Como alternativa, os autores recuperam os estudos sobre racionalidade limitada e racionalidade de procedimentos com informação imperfeita, dos quais Herbert Simon fora pioneiro. Nessa perspectiva, os agentes recorrem a heurísticas, “regras de bolso” ou padrões convencionais de comportamento, que orientam suas decisões em condições de incerteza. Essas regras são uma resposta racional, sistemática e inteligível frente a um ambiente complexo, mas que não requer a maximização de lucros nem o equilíbrio (a firma estaria apenas “satisfeita” com aquele conjunto de regras).

² O tema começou a ser amplamente debatido a partir do trabalho de Hall e Hitch, que mostrou que as firmas não utilizavam algoritmos de maximização para tomar decisões sobre preços (cf. Hall, R. L.; Hitch, C.J. “Price Theory and Business Behaviour”, in *Oxford Economic Papers*, 1939).

O conceito de heurísticas revela-se especialmente poderoso em termos de sua capacidade de organização da pesquisa por duas razões. A primeira e mais importante é que esse conceito é inserido num arcabouço dinâmico mais amplo, dado pela interação entre as heurísticas que orientam o comportamento das firmas e o ambiente de seleção. Essa interação é a que sustenta o processo recursivo de transformação das principais variáveis da estrutura industrial — número e tamanho das firmas, suas parcelas de mercado, a velocidade com que as inovações são adotadas por outras firmas, o nível e a taxa de aumento da produtividade. Mais ainda, essa perspectiva dinâmica, a partir da qual se geram, endogenamente, sucessivos “estados” da estrutura industrial, pode ser formalizada por meio de modelos de simulação (que os autores exploram em outros trabalhos).³ Isso torna possível o processo de fertilização mútua entre as teorias apreciativa e formal.⁴ A transformação industrial ocorre a partir da interação entre as heurísticas das firmas, a estrutura inicial da indústria (com destaque para as assimetrias tecnológicas entre as firmas) e as características da tecnologia.

A segunda dimensão tem a ver, precisamente, com a forma de analisar as características da tecnologia. Admite-se que o âmbito tecnológico obedece a regras e procedimentos que lhe são específicos, isto é, que tem heurísticas próprias, dotadas de certa autonomia com relação aos âmbitos científico e econômico. Esse tema já tinha sido sugerido por Nathan Rosenberg na sua crítica às teorias do *demand-pull*,⁵ mas Nelson e Winter avançam nessa direção através do conceito de “regime tecnológico”. Em particular, a autonomia do âmbito tecnológico permite explicar alguns paradoxos não resolvidos pelas teorias do *demand-pull* e do *technology-push*: a inovação responde aos in-

³ Cf. em particular Nelson, R. R.; Winter, S.; Schuette, H. L., “Technical Change in an Evolutionary Model”, in *Quarterly Journal of Economics*, v.90, 1976.

⁴ A teoria apreciativa sugere um conjunto de variáveis relevantes e uma estrutura de relações entre elas que representam um primeiro esforço de construção de um marco analítico consistente. Cabe à teoria formal exprimir essas relações em linguagem matemática, torná-las mais rigorosas e depurá-las de eventuais problemas de inconsistência lógica (cf. Nelson, R., “Economic Growth via the Co-Evolution of Technology and Institutions”, in Leydesdorff, L.; Basselaar, P. van der, *Evolutionary Economics and Chaos Theory*, Nova York: St Martin’s Press, 1994). Para Nelson e Winter, as teorias formal e apreciativa são fases distintas e complementares da construção teórica, e não caminhos alternativos.

⁵ Rosenberg critica a teoria do *demand pull* por ignorar que a tecnologia tem regras e uma temporalidade próprias. Cf. Rosenberg, N., *Perspectives on Technology*, Cambridge University Press, p.279, 1976. Rosenberg observa que a tecnologia em si mesma é um corpo de conhecimentos sobre certas classes de eventos e atividades, e não a mera aplicação de uma fonte externa de conhecimentos. Cf. Rosenberg, N., *Inside the Black Box: Technology and Economics*, Cambridge University Press, 1982.

centivos da demanda (isto é, ela é parcialmente endógena), mas o faz dentro dos limites ou restrições impostas pela tecnologia, que não são uniformes através dos setores e do tempo. Posteriormente, os conceitos de paradigma tecnológico e trajetória tecnológica oferecerão novos instrumentos para entender como tecnologia e economia se condicionam mutuamente no processo de inovação e concorrência.⁶

Instituições e políticas

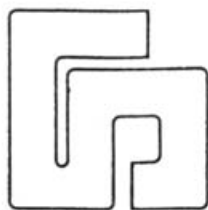
Outro elemento distintivo da abordagem de Nelson e Winter é a preocupação por incluir na análise a diversidade e a complexidade institucionais da inovação. Como resultado, não há um único padrão de geração e difusão de inovações, mas este pode variar grandemente em função das características da estrutura institucional. Essa estrutura depende de diversos fatores, entre eles: uma maior ou menor presença de instituições públicas de pesquisa e/ou de difusão de tecnologia; a existência ou não de modalidades formais ou informais de cooperação entre empresas privadas ou entre instituições públicas e privadas; o marco legal e de políticas que restringe as estratégias possíveis (p. ex., diferenças no sistema legal de patentes ou na legislação *anti-dumping*).

Como essa estrutura institucional difere entre países, regiões e setores, ela pode contribuir para explicar diferenças nas taxas de aumento da produtividade, assim como nas modalidades de difusão da inovação. Tipicamente, em alguns casos, a expansão da nova técnica se dará pelo aumento da parcela de mercado do inovador e pela paralela expulsão das firmas retardatárias; em outros casos, a inovação tenderá a ampliar sua participação no mercado pela via da imitação, à medida que os retardatários diminuem sua distância com relação à fronteira tecnológica e os lucros extraordinários do inovador se diluem. Naturalmente, o predomínio de um ou outro caminho de difusão terá conseqüências importantes em termos da concentração na indústria e do porte das firmas, e esse deve ser um aspecto a ser considerado na formulação das políticas.

⁶ Cf. Dosi, G., "Technological Paradigm and Technological Trajectories", in *Research Policy*, v.11., n.3, jun., 1982. Para Dosi, a "trajetória tecnológica é um *cluster* de possíveis direções tecnológicas cujas fronteiras externas são definidas pela natureza do próprio paradigma" (cf. Dosi, G., *Technical Change and Industrial Transformation*, Londres: The McMillan Press, p.17, 1984).

Em particular, a ênfase na diversidade da estrutura institucional sugere uma mudança com relação ao foco da política industrial. Para os autores, esse foco não se define em termos de uma firma, ou de um setor, ou em termos de uma abordagem “horizontal” que ignore (em prol da “neutralidade” dos incentivos) as especificidades do ambiente tecnológico e competitivo em que as firmas operam. Pelo contrário, a política deve ser desenhada para influenciar as interações entre um grupo diversificado de agentes (em nível nacional ou regional), que endogenamente geram a mudança tecnológica e os ganhos de produtividade. Esta nova perspectiva — que será posteriormente desenvolvida a partir do conceito de Sistemas Nacionais de Inovação —⁷ permanece como um traço distintivo do enfoque schumpeteriano sobre política industrial e é mais uma evidência da vitalidade da teoria da inovação de Nelson e Winter.

⁷ Sobre o conceito de Sistema Nacional de Inovação, cf. o artigo clássico de Christopher Freeman e a Apresentação de Eduardo da Motta e Albuquerque na *Revista Brasileira de Inovação*, vol.3, n.1, jan.-jun., 2004.



SUMMARY

This essay presents an overview of the prevailing theoretical literature on innovation, probes the adequacy of existing theory to guide policy regarding innovation, and sketches some directions for more fruitful theorizing. The focus is on the vast interindustry differences in rates of productivity growth, and other manifestations of differential rates of technological progress across industries. It is argued that the most important policy issues involve finding ways to make the currently lagging sectors more progressive, if in fact that can be done. Theory, to be useful, therefore must organize knowledge and guide research regarding what lies behind the uneven performance of the different economic sectors. In fact prevailing theory cannot do this, for two basic reasons. One is that theory is fragmented, and knowledge and research fall into a number of disjoint intellectual traditions. The second is that the strongest of the research traditions that bear on the differential innovation puzzle, research by economists organized around trying to 'fit' production functions and explain how production functions 'shift', neglects two central aspects of the problem; that innovation involves uncertainty in an essential way, and that the institutional structure supporting innovation varies greatly from sector to sector. The bulk of the paper is concerned with sketching a theoretical structure that appears to bridge a number of presently separate subfields of study of innovation, and which treats uncertainty and institutional diversity centrally.

in search of useful theory of innovation*

by

Richard R. NELSON

*Institution of Social and Policy Studies, Yale University, New Haven,
Connecticut 06520, USA*

and

Sidney G. WINTER

Economics Department, Yale University, New Haven, Connecticut 06520, USA

1. INTRODUCTION

This essay presents an overview of selected aspects of prevailing theoretical understanding of innovation, and attempts to sketch some directions that would seem fruitful to follow if we are to achieve a theoretical structure that can be helpful in guiding thinking about policy. We are using the term innovation as a portmanteau to cover the wide range of variegated processes by which man's technologies evolve over time. By a theory we mean a reasonably coherent intellectual framework which integrates existing knowledge, and enables predictions to go beyond the particulars of what actually has been observed. It seems apparent that if scholarly knowledge is to be helpful to deliberation about possible policy directions, theory must be wide enough to encompass and link the relevant variables and their effects, and strong enough to give guidance as to what would happen if some of these variables changed.

It is also apparent that, in trying to design policies aimed at so complex a web of social phenomena as innovation, intelligent policy makers are going to look to the scholarly community for advice. Indeed, research by economists and other social scientists on various aspects of innovation has had a major impact on recent policy thinking. In the United States, the Congressional hearings on science and technology policy, and the speeches of high officials on these matters, are full of 'facts' discovered by economists about the major role of

* This paper was supported partly under NSF Grant No. GS 35659 and partly under a grant to Yale from the Sloan Foundation. An earlier version of part of this paper was presented at a conference in Cambridge, Summer 1975.

technological advance in productivity growth, as a source of comparative advantage, etc. The Joint Economic Committee of the United States Congress recently commissioned a review of the literature. The economists' concept of externality is much bandied about in the policy dialogue in the UK as well as the US, and is a stock part of the rhetoric emanating from OECD*.

The current dialogue regarding policy toward innovation rests on two premises. The first is that technological advance has been a powerful instrument of human progress in the past. The second is that we have the knowledge to guide that instrument toward high priority objectives in the future. The first premise is unquestionable: the latter may be presumptuous. While all the attention recently given by politicians to scholars is flattering, we believe that the scholarly community has much less to say about appropriate policy toward innovation than many scholars like to believe. Prevailing theory of innovation has neither the breadth nor the strength to provide much guidance regarding the variables that are plausible to change, or to predict with much confidence the effect of significant changes.

Implicit in these critical remarks is a set of commitments on our part as to the nature of the major policy issues involving innovation. In a nutshell they are these.

First, at the present time the appropriate policy objectives are not well characterized as more effective general stimulation of technological advance, but involve more selective targeting. In the past technological advance has been extremely uneven across economics sectors and industries. Table 1, drawn from Kendrick's recent study, displays the vast intersectoral variations in growth of total factor productivity, and labor productivity, that were experienced over the 1948–1966 period [21]. Evidence of great intersectoral differences of productivity growth was provided earlier in Kendrick's 1961 study [20], and still earlier in work by Salter [48] and Schmookler [49]. The phenomenon seems to obtain in other countries as well as the US. Where other measures or indices of technological progress are available, they correlate reasonably well with the productivity measures, and show a similar cross industry dispersion.

The consequences of the imbalance of rates of productivity growth have been profound. We have experienced sharply rising relative costs and prices in the slow productivity growth sectors, and it is apparent that there is a widely held discontent with the performance of industries like housing, the complex of service industries, and with governmentally provided services like garbage

* The Joint Economic Committee paper is by Gilpin [11]. King [23] describes the science-policy thinking at OECD.

Table 1
Productivity growth in the private business economy by industry group and industry,
1948–1966

	Percentage yearly change in total factor productivity	Percentage yearly change in output per worker
<i>Private domestic business economy</i>	2.5	3.1
<i>Farming</i>	3.3	5.6
<i>Mining</i>	4.2	4.6
Metal	2.4	2.9
Coal	5.2	5.8
Crude petroleum and natural gas	3.2	2.3
Nonmetallic mining and quarrying	2.6	3.2
<i>Contract construction</i>	1.5	2.0
<i>Manufacturing</i>	2.5	2.9
Nondurables	2.6	3.2
Food, except beverages	3.0	3.4
Beverages	2.2	2.9
Tobacco	1.1	2.7
Textiles	4.0	4.3
Apparel	1.9	2.2
Paper and paper products	2.5	3.0
Printing	2.7	2.7
Chemicals	4.9	6.0
Petroleum refining	3.0	5.5
Rubber products	3.9	4.0
Leather products	1.7	1.7
Durables	2.4	2.8
Lumber	3.5	3.9
Furniture	2.9	2.9
Stone, clay, glass	2.4	3.2
Primary metal products	1.6	2.1
Fabricated metals	1.9	2.2
Machinery, except electric	2.6	2.7
Electric machinery	3.7	-4.1
Transportation equipment and ordinance	3.2	3.2
Instruments	2.9	3.7
Miscellaneous manufacturing	3.5	4.0
<i>Transportation</i>	3.4	3.7
Railroads	5.2	5.8
(Nonrail)	2.1	2.3
Local railroads and bus lines		-1.0
Intercity passenger		1.5
Intercity motor trucking		3.1
Water transportation	0.5	0.7
Air transportation	8.0	8.2
Pipelines		9.1

Table 1 (cont'd)
Productivity growth in the private business economy, by industry group and industry,
1948–1966

	Percentage yearly change in total factor productivity	Percentage yearly change in output per worker
<i>Communications and public utilities</i>	4.0	5.8
Communications	3.8	5.5
Electric, gas and sanitary service	3.9	6.1
Electric and gas	4.9	7.1
<i>Man-made</i>	2.5	2.9
Wholesale	2.5	3.1
Retail	2.4	2.7
<i>Finance, insurance and real estate</i>	NA	2.1
<i>Services</i>		
(except households; incl. Government enterprises)	NA	1.2

Source: Kendrick [21], from tables 5.1 and 5.5.

collection and street cleaning*. On the other hand, for many of the goods and services produced by the rapid productivity growth sectors, there is an apparent sense of surfeit. There is at once a reaction of *déjà vu* regarding the coming generation of supersonic commercial aircraft, and a strong felt need to improve the railroads.

Second, to the extent that the argument above is accepted, the hunt for appropriate policy instruments will not be an easy one. Macro measures will not do; thus proposals like a general R & D tax credit (which has been quite fashionable in the recent discussions in the US) are beside the point. Policies need to be designed to influence particular economic sectors and activities. Regarding these, the key policy problem will be to augment or redesign institutions rather than to achieve particular resource allocations *per se*. Improving the railroads does not look like an objective that can be met through funding a few well specified R & D projects. Rather, the policy search must be for a set of institutions that will allocate resources appropriately over a wide range of circumstances and time.

Third, the character of the appropriate institutional structure for the generation, screening, and effective exploitation of innovation depends on the

* Quantitative analyses of the relationship between relative productivity growth, and relative price changes, are contained in Kendrick [20, 21] and Salter [48]. Baumol [1] articulates the malaise about the slow productivity growth industries.

underlying technologies, the nature of the demands for the goods and services, and the characteristics of the organizations supplying them. These critical variables differ from sector to sector. General analytic arguments, for example about externalities that are inherent in certain kinds of R & D, have little bite since they ignore sectoral differences. Useful analysis must focus on and illuminate these differences.

While these arguments will be developed in more detail shortly, we believe to some extent the plausibility of these basic premises is apparent on their face. If they are accepted then, to be useful to thinking about policy, theory must be wide enough to relate the technological progressivity of the sector to the institutional structure involved. And the theory must be strong enough to guide plausible thinking about the effects of the various changes in institutional structure. It is apparent that any useful and coherent theory of innovation must recognize explicitly the factors that differ across industries. Our present bag of theory of innovation does not do this, and thus is not very useful in the current policy dialogue.

Sect. 2 will review selected aspects of the literature bearing on the inter-industry productivity puzzle, attempting to pinpoint where the real weaknesses are, and to identify the essential building blocks for a mere useful theory of innovation. Sects. 3 and 4 present a way of theorizing about innovation that we believe has promise of broadening and strengthening understanding of the innovation process and the variables impinging on it, so as to improve our ability to design efficacious policy. We certainly do not profess to have yet achieved a complete intellectual edifice. However, we can show you the floor plans, and point to certain parts of the structure that are taking shape here and there. In sect. 5 we shall sketch a few questions relating to institutional structure.

2. THE STATE OF CURRENT UNDERSTANDING

The weakness of present understanding of the reasons behind the differential productivity growth puzzle is in part due to lack of facts. But it is due at least as much to lack of theory that will enable us to knit together and give structure to what we know and extend our knowledge beyond particular facts. While there has been a considerable volume of research by economists, other social scientists, and historians of science and technology, that ought to bear on the differential productivity puzzle, that research is not well connected. This makes review, much less integration, of what is known quite difficult. More important, it means that knowledge is in the form of congeries of semi-isolated facts, rather than a connected intellectual structure.

2.1. The economists' model of differential productivity growth

There has been, however, one cluster of research that has been aimed directly at the differential productivity growth puzzle, and which has had a sufficiently strong theoretical structure so that knowledge is relatively integrated and has been cumulative. This structure derives of the text book economic theory of production. Studies within this tradition generally take the form of regression analyses attempting to explain differential productivity growth by differences in research and development and other factors which are seen as 'shifting' the production function.

Kendrick's early study, particularly the work of Terleckyj contained in that study, set the style for much of the subsequent work based on production theory [20]. The analysis was concerned with explaining cross sectoral differences in growth of total factor productivity, rather than output per worker. While the two measures are highly correlated, it is significant that the analysis simply took as given productivity growth accounted for by increases in capital per worker, and aimed to explain the 'residual.' For our purposes, the primary conclusion was that research and development intensity of an industry (measured in several different ways) was a significant factor explaining differences in total factor productivity growth across manufacturing industries (the only industries for which research and development data existed) both in simple and multiple regressions. Several other industry characteristics, for example growth of total output in the industry and sensitivity of output to cyclical fluctuations, also were strongly correlated with growth of total factor productivity. For these variables the possibility of two-way causation was recognized explicitly. For the most part, subsequent studies attempting to explain across industry differences in rates of productivity growth have followed roughly in this pattern set by Kendrick–Terleckyj, but with a wider and finer scan of variables and industries, and (in some cases) more concern with theoretical specification of the regression equations.

Mansfield's study, as Kendrick–Terleckyj, was limited to a sample of manufacturing industries and was focused in the relationship between growth of total factor productivity and R & D spending [26]. Unlike Kendrick–Terleckyj, Mansfield worked with an explicit production function formulation in which accumulated and depreciated research and development expenditures were treated as a form of capital. A number of different specifications were explored; for example, both embodied and disembodied technical progress models were regressed. Like Kendrick–Terleckyj, Mansfield found a significant effect of an industry's research and development spending upon its

measured rate of technical progress. This conclusion was robust under different specifications, although the magnitude was sensitive to the exact specification.

Leonard's study also dealt only with manufacturing [24]. His most significant contribution to the analytic dialogue was a separation of research and development spending financed by the industry itself, and research and development spending done in an industry but financed by government. His empirical conclusion was that self-financed research and development spending had a much larger impact than governmentally financed research and development upon both growth of industry output, and growth of output per worker. Leonard did not work with a total factor productivity measure, although it is apparent that he would have got the same qualitative results using it as he got with his output per worker measure. It is an open question as to whether his results are identifying a real difference between the effectiveness of funds from different sources that would apply across the industry spectrum, or whether his procedure simply has separated out a few major industries, like aircraft and missiles, and electronics, where the government is both a significant buyer, and a significant research and development financier. If the latter, government research and development spending may be a proxy for several special factors that obtain in these industries. One might conjecture that in the defense industries R & D has been pushed into a regime of very low marginal returns, and that this is what the low weight on government R & D is capturing. Also, there are difficulties (which may exert a downward bias to measured technical advance) with the price deflator used to measure growth of real output in industries (like the aerospace complex) which sell specially designed new products to a monopsonistic buyer.

Brown and Conrad developed a quite complex theoretical specification of the factors influencing the rate of growth of industry output, which, however, was simplified significantly as the authors moved from the theoretical equations to the regression equations actually fitted [4]. Their sample of industries, as those in the studies listed above, included only those in the manufacturing sector. Their major contribution to the analytical dialogue was to include in their regressions a measure of research and development done by other industries and embodied in the intermediate goods purchased by the industry in question, as well as research and development funding of the particular industry. Their regression results show a significant impact of such indirect research and development. In Brown and Conrad, own and indirect R & D is simply added together. In a later study, Raines entered these separately in his regressions [45]. His results indicated, strangely enough, that indirect research and development spending seems to have a more powerful effect than

own research and development spending.

Perhaps the most interesting recent study is that by Terleckyj [56]. He considers nonmanufacturing industries as well as manufacturing industries. This is an important and provocative extension. The average rate of total factor productivity growth in the nonmanufacturing sector has been at least as fast as that in the manufacturing sector, and yet their own research and development intensity is drastically lower in nonmanufacturing than in manufacturing. In his analysis of the effect of own research and development spending, Terleckyj found that applied research and development funding provided by companies in an industry was a more powerful explanation of differences in productivity growth across manufacturing industries than total research and development expenditures done by the industry. Since the principle difference between the two measures is government financed research and development, his findings are consistent with those of Leonard.

Terleckyj's most significant contribution is an enriched and sophisticated extension of the earlier work of Brown and Conrad, and Raines, assessing the contribution of research and development done by other industries and embodied in inputs. He distinguishes between research and development embodied in capital equipment, and in intermediate inputs. In simple regressions, both are significant factors in explaining differences in productivity growth across the spectrum of industries, nonmanufacturing as well as manufacturing. However, when own R & D and R & D embodied in inputs are both placed in multiple regression equations, own research and development spending does not carry much explanatory power although it is slightly more important for manufacturing than for nonmanufacturing. This conclusion, anticipated by Raines, could result from a variety of factors. One possibility is that research and development spending financed by an industry is largely focused on new product design that, because of weaknesses in price indices, does not show up adequately in measures of growth of industry output; in contrast process improvements, which show up more reliably in increased productivity, stem largely from improved inputs and capital equipment. Another possibility is that own R & D spending and purchased inputs from R & D intensive industries are strong complements. This would show up if the regression were correctly specified. But under slight misspecification the regression weight can shift to one variable or the other.

The studies have been useful and provocative, but have not cut very deep. Severe measurement problems make interpretation of some of the statistical results, difficult. Some of these measurement problems were sketched above. Griliches' recent work provides a wider catalogue [14].

But the fundamental problem is that of specification. The regression equa-

tions involve complex interactions in which factors influencing the demand side and the supply side of the technical change process are intertwined, and confounded with other influences. Kendrick's recent analysis shows this sharply [21]. In simple correlational analysis, he finds that productivity growth in an industry is related to research and development spending. However, productivity growth also is strongly correlated with growth of the output of the industry over time. When Kendrick employs a step-wise multiple regression technique, the two explanatory variables that carry the most weight in explaining productivity growth are growth of output, and extent of unionization (which has a negative effect). With these variables included, other variables like research and development spending add little to the goodness of fit. But interpretation is hazardous. Obviously, there are a tangle of causations, from R & D to productivity growth, from productivity growth and lowered prices to growth of output, from growth of output in the presence of scale economies to productivity growth, from expansion of the industry to greater incentives for R & D, and so on. One also can wonder whether unionization is a cause, or an effect, of sluggish technical change in an industry. While Kendrick's particular analytic strategy makes this problem explicit, it is implicit in the studies above which simply related productivity growth to research and development spending.

And even were it granted that the causation runs, at least in part, from research and development spending to industry progressivity, what explains the uneven allocation of research and development spending? Among the range of possible explanations two stand in stark contrast regarding their policy implications. One is that research and development activity is more powerful when directed toward the technologies of certain industries than toward the technologies of others; therefore, the disparities in rates of technical progress reflect some kind of innate differences on ability to advance efficiently the different kinds of technologies. The second possible explanation (not mutually exclusive) focuses not on possible innate differences in what R & D can do in different sectors, but on differences in institutional structure that influence the extent to which R & D spending is optimal and the results of R & D effectively employed. The proposition is that industries differ significantly in the extent to which the results of research and development spending are internalized by the sponsoring firms, that in some industries but not in others there is significant government subsidization of research and development where externalities are important, and that industries also differ significantly in the speed and reliability of the mechanisms by which new technology is screened, and the use of efficacious innovation spread throughout the sector. Needless to say, the differences in

these explanations matter profoundly in terms of their policy implications. Some headway can be made on the key questions by augmenting and enriching the basic production function framework above. For example, R & D spending might be treated *not* as an independent variable but as a variable to be explained by other factors, some institutional, some proxies for possible innate differences across industries. While, to our knowledge, this kind of integrated analysis is rare, the 'productivity growth as a result of R & D' studies are quite conformable with other studies relating R & D spending in an industry to firm size and concentration variables. Similarly, it ought to be possible to try to explain the magnitude and perhaps the productivity of applied research and development spending in an industry by certain proxy indices of the strength of the scientific base.

We would argue, however, that the breadth and strength of the production function framework is inherently limited. To obtain a more solid understanding of innovation and what can be done to influence innovation, it is necessary to study in some considerable detail the processes involved and the way in which institutions support and mold these processes. Since the 'production function framework' contains at best a rudimentary characterization of process and relevant institutional structure, a considerably more fine grained theoretical structure is needed for these more microscopic studies.

2.2. Building blocks for a broader theoretical structure

A considerable body of research has attempted to take a more microscopic look.* There is a rapidly increasing literature on the nature of the research and development process, the links between science and invention, the sources of invention (large firms, small firms, private inventors), the kinds of organizational and other factors associated with successful choice and carrying out of a project, etc. Other studies have probed at learning phenomena and, more generally, the way technologies (or a particular technology) evolve over time. A significant literature exists on organizational factors influencing the decision to adopt an innovation. Diffusion of innovation has been a fertile research field in several disciplines. While, for the most part these studies have not been concerned explicitly with inter-industry differences, innovations in many different industries have been studied, and it ought to be possible to make some cross-industry comparisons.

* There have been several recent relatively comprehensive reviews of pieces of this literature. Among the best are Mansfield [28], Pavitt [42], Freeman [9] and Kennedy and Thirlwall [22].

Unfortunately, these studies add much less than we would hope to our understanding of the differential productivity growth puzzle. The basic problem was stated briefly earlier — by and large, these studies have proceeded within disjoint theoretical frameworks. These are virtually no conceptual bridges between project SAPPHO which probes at conditions for successful innovation [55], the Jewkes, Sawers, and Stillerman study of the sources of invention [18], and studies by economists such as Mansfield [26] and Griliches [13] on diffusion. Most important, there is no way to link these studies and the body of research concerned directly with differential productivity growth. Our knowledge is Balkanized. We cannot, in general, bring together several different bodies of analysis to focus on any one question, or tie together the various pieces to achieve an integrated broader perspective. Thus, while a considerable amount of research had probed at the details of process, and at the nature of prevailing institutional structure and how this has influenced process, there is no way in which we can confidentially link this knowledge to our understanding of the factors behind the differential productivity growth rates we have experienced. However, the micro studies have documented a number of ‘facts’ about innovation with which any widely embracing theory must be consistent. Two of these facts indicate that it is not promising to use the theoretical structure behind the productivity growth studies as a starting point.

The first of these facts is that innovation involves uncertainty in an essential way. The implicit process characterization of the ‘production function’ models would appear to be not only rudimentary, but fundamentally misleading. The problem cannot be patched up by reposing the theory in terms of expectations, introducing considerations of risk aversion, etc. Rather, a theoretical structure must encompass an essential diversity and disequilibrium of choices. Because of the uncertainty involved, different people, and different organizations, will disagree as to where to place their R & D chips, and on when to make their bets. Some will be proved right and some wrong. Explicit recognition of uncertainty is important in thinking about policy. One fundamental implication is that it is desirable for the institutional structure to generate a variety of innovations. Another implication is that a major function of an effective institutional structure is that it screen innovations effectively, accepting and spreading the good, winnowing out the bad.

A second fact that the microcosmic studies have illuminated is that the institutional structure for innovation often is quite complex within an economic sector, and varies significantly between economic sectors. Thus, in agriculture, there is considerable public subsidization of research done by predominantly non-profit institutions (largely universities) and a subsidized

federal-state extension service for the dissemination of information regarding new technological developments to farmers, interacting with the network of private farms, and industries that produce and sell farm equipment, fertilizers, etc. The commercial aircraft industry is equally complex, but must be described in quite different terms. Innovation in medicine involves a set of institutions different from either of these. This institutional complexity and diversity would seem to be where the focus of a policy attention should be; however, it does not seem possible to extend the rudimentary institutional assumptions built into the production function model sufficiently to really grip these dimensions.

If there is to be any hope of integrating the disparate pieces of knowledge about the innovation process, a theory of innovation must incorporate explicitly the stochastic evolutionary nature of innovation, and must have considerable room for organizational complexity and diversity. Merely having these attributes is, of course, no guarantee that a theoretical structure will be able to integrate what is known, or have the power to predict the effect of reasonable changes in institutional structure. But, these attributes can give some guidance to the search for a useful theory.

Over the past several years, we have been developing the outlines of such a theoretical structure*. Our objective is to develop a class of models based on the following premises. First, in contrast with the production function oriented studies discussed earlier, we posit that almost any nontrivial change in product or process, if there has been no prior experience, is an innovation. That is, we abandon the sharp distinction between moving along a production function and shift to a new one that characterizes the studies surveyed earlier. Second, we treat any innovation as involving considerable uncertainty both before it is ready for introduction to the economy, and even after it is introduced, and thus we view the innovation process as involving a continuing disequilibrium. At any time there is coexistence of ideas that will evolve into successful innovations and those that will not, and actual use of misjudged or obsolete technologies along with profitable ones. Over time selection operates on the existing set of technologies, but new ones continually are introduced to upset the movement toward equilibrium.

More formally, in an accounting sense we view productivity growth as explained within our proposed theoretical structure in terms of first, the generation of new technologies, and second, changes in the weights associated with the use of existing technologies. This accounting distinction reflects an

* See in particular refs. [35, 36, 38, 39].

analytic break. We are attempting to build conformable sub-theories of the processes that lead up to a new technology ready for trial use, and of what we call the selection environment that takes the flow of innovations as given. (Of course, there are important feedbacks.)

The structure we propose is capable of quite formal articulation and modeling. We have built several formal models and reported on these in earlier papers. In this paper, the emphasis will be on the theory as providing a useful language for talking about dynamic process, for integrating presently disparate knowledge, and for guiding future research.

3. THE GENERATION OF INNOVATION

Fragmentation of empirical work certainly characterizes the state of knowledge regarding the genesis of innovation. However, almost all students of innovation agree that purposive acts of investment are an important part of the process. Many scholars have recognized two classes of factors influencing the allocation of effort: factors that influence the demand for or pay-off from innovation, and factors that influence the difficulty or cost of innovation.

Demand side factors have been studied by a variety of research traditions. In economics, Schmookler, for example, has articulated and provided strong supporting evidence for a simple model in which changes in the composition of demand for goods and services across industries chain back to influence investment patterns, which in turn influence the relative return to inventors working on improvements in different kinds of machines. The theoretical work by economists on factor bias of innovation, and the empirical work which started with Habakkuk [15], also is focused on the effect of demand side factors. Recently Hayami, Ruttan and followers, have provided strong empirical support for the proposition that relative factor prices influence the nature of innovation, at least in agriculture*. A wide variety of more micro studies, by economists and other social scientists, examining particular inventions, or the influences that bear on allocation decisions within a particular R & D organization, also have identified the importance of perception of demand or pay off.

Other studies have focused on the cost or supply side. In contrast with

* The key reference is Hayami and Ruttan [16]. An extension and updating will be published soon by Binswanger and Ruttan [3].

research on demand side factors, research aimed at exploring differences in the difficulty or cost of different kinds of innovation has had but limited conceptual and empirical pay-off. A list of variables and conjectures has been proposed, for example, that the scientific base for innovation differs from industry to industry, or that innovation is intrinsically easier in industries that produce physical products than in those that produce services. It also has been proposed that mechanization, where relevant, is a particularly easy route for innovation and similarly, that latent scale economies often provide a route that is easy to follow*. Empirical support for these propositions, however, has been weak. Nonetheless, there would appear to be only a few dissenters to the proposition that factors on the cost side as well as factors on the demand side differ across industries and technologies, and that these differences are important in explaining the pattern of innovation that has occurred.

3.1. The profit maximization hypothesis and its limitations

The theoretical problem is how to organize what we know so that the whole adds up to more than the sum of the parts and knowledge extends beyond the particulars. At first thought it is tempting to adopt the economists' model of the profit maximizing firm as the basis for a theory of innovation genesis. The semi-empirical fact of purposeful and intelligent behavior regarding the allocation of resources to innovation, which almost all scholars believe characterizes the true situation, would be theoretically translated as profit maximizing behavior. Factors on the demand and supply side then would be brought in as they are in the traditional theory of the firm.

The models aiming to explain differential productivity growth, discussed in the previous section, in general have taken R & D expenditures as given, rather than as something to be explained. However, it seems apparent that the authors had in mind something like a profit maximizing model behind the scenes. And several profit maximizing models aimed at exploring R & D spending have been constructed. For example, it is possible to deduce the Schmookler conclusions simply by treating accumulated R & D spending as a form of capital, and extending the traditional result in the theory of the firm regarding the affects of an increase in demand for a product upon the amount of investment it is profitable to undertake. The literature by economists on factor bias, mentioned above, is self consciously within the 'profit maxi-

* For the general discussion see Nelson, Peck and Kalachek [34]. Baumol [1] has stated the hypothesis about the service sectors. See Rosenberg [46], David [7], Hughes [17] and Levin [25] regarding mechanization and economies of scale.

mizing' frame. The key concept is that of an innovation possibility set associated with a given level of expenditure, or with different elements associated with costs of different amounts. Firms are assumed to choose the profit maximizing element*. The effects of changes in factor prices on the direction of inventing can be deduced within this model.

A profit maximizing interpretation of purposive and intelligent behavior clearly has considerable organizing power. However, there are certain important difficulties with this formulation that need to be recognized.

The basic limitations relate to our remarks in the preceding section. In many cases, the organizations doing R & D are not motivated by profits at all, but are governmental, or private not-for-profit institutions. The difficulty here can be resolved on the surface by treating the term 'profit' very broadly to stand for whatever objectives the organizations happen to have. We shall adopt such a tactic in the following section. But, at the least, this calls for rather detailed empirical inquiry as to the nature of the organizations doing the R & D before the 'profit' concept can be made operational in a model. And, actually, the problem is even more complicated than that. In many sectors there are a complex of R & D organizations, some profit oriented, some governmental, some academic, doing different things, but interacting in a synergistic way. In particular, in medicine, agriculture, and several other sectors, private for-profit organizations do the bulk of R & D that leads to marketable products, but academic institutions play a major role in creating basic knowledge and data used in the more applied work.

Relatedly, everyone agrees that R & D is an uncertain business. Uncertainty resides at the level of the individual project, where the 'best' way to proceed seldom is apparent and the individuals involved instead have to be satisfied with finding a promising way. Uncertainty also resides at the level of R & D project selection. The enormity of the set of possible projects, the inability to make quick cheap reliable estimates of benefits and costs, and the lack of convenient topological properties to permit sequential search to home in rapidly on good projects independently of where that search starts, means that project choice, as well as outcome given choice, must be treated as stochastic. The problem with the maximization metaphor is not that it connotes purpose and intelligence, but that it also connotes sharp and objective definition of the range of alternatives confronted and knowledge about their properties. Hence it suggests an unrealistic degree of inevitability and correctness in the choices made, represses the fact that interpersonal and

* For good recent reviews and criticism see Binswanger [2] and Nordhaus [40]

interorganizational differences in judgment and perception matter a lot, and that it is not all clear ex-ante, except perhaps to God, what is the right thing to do.

Largely because of these limitations, the profit maximization formulation has been unable to cope with certain salient aspects of the innovation generation process. While it has had qualitative success in dealing with certain variables on the demand side, it has tended to ignore externalities, implicitly presuming that 'consumer' valuations are reflected in profit opportunities for firms doing R & D. It also has neglected the range of questions relating to how consumers find out about and value new products and processes. On both of these matters, the particular structure of institutions, and the way uncertainties are resolved, are key. Our discussion of the 'selection environment' for innovations is aimed to deal with these kinds of problems more coherently.

The profit maximization formulation has had very little grip on factors on the cost or feasibility of innovation side. Here a more process oriented characterization of the innovation generation process seems essential. We shall sketch such a characterization in the remainder of this section.

3.2. R & D strategies and probabilistic outcomes

In this section we outline what we believe may be a fruitful way to deal with some of these difficulties. The components of our proposal are these. An R & D project, and the procedures used by an R & D organization to identify and screen R & D projects, can be viewed as interacting heuristic search processes. A quasi stable commitment to a particular set of heuristics regarding R & D project selection can be regarded as an R & D strategy*. Often it is possible to identify a few R & D strategies that are prevalent in a particular sector in a particular era. An R & D strategy might be modeled extensively in terms of the heuristics employed in the search processes and their consequences. Or a strategy can be represented in terms of a conditional probability distribution of innovations (or innovation characteristics) given certain conditions facing the organization.

By a heuristic search process we mean an activity that has a goal, and a set of procedures for identifying, screening, and homing in on promising ways to get

* Note that the concept of a strategy employed here is *not* the same as the strategy concept used in statistical decision theory or game theory. In general a strategy, as we are employing the term, will not involve an explicit ex-ante solution to what will be done under all possible contingencies. Nor is our kind of strategy derived from any explicit maximization calculations.

to that objective or close to it*. The procedures may be characterized in terms of the employment of proximate targets, special attention to certain cues and clues, and various rules of thumb. While they may be fruitful in the sense of yielding relatively satisfactory outcomes a good percentage of the time, they do not guarantee a good outcome or even a unique one. That is, they are heuristics, rather than an algorithm for calculating an optimum. While a costless maximizing algorithm would be preferred by decision makers to good heuristics, in complex decision problems a maximization algorithm is likely to be highly expensive to employ, even if one is known. In most cases one simply doesn't exist. Thus good heuristics is the best one can hope for.

The notion that an R & D project can be viewed as a search is quite widespread in the literature. Recently, Evanson and Kislev have developed a specific model of the search process involved in trying to find a new seed variety with improved performance in certain dimensions [8]. They place their model in a maximizing framework, in the sense that they calculate the number of elements (seed varieties) that 'should' be sampled. However, their model glides over exactly those aspects of an R & D project where heuristics are important – the judgments as to what elements to sample first, assessment of what has been learned from the early draws that provides guidance as to the next steps, etc. We shall not elaborate here the notion of intra project heuristics save for one point. We propose that an important aspect of the question of the role of science in invention can fruitfully be posed in terms of the ways in which scientific knowledge enables powerful search heuristics.

While this has been less extensively discussed, it is apparent that the project selection process also must be heuristic. For all of the reasons discussed above, the selection by a firm of a set of projects to pursue cannot be maximizing in any serious sense. Rather, the process must employ various proximate targets, attend to particular cues and clues, use various rules of thumb.

One of the characteristics of heuristics is that often they factor a complex interrelated decision problem into parts, which then are treated as more or less independent. For example, obviously decisions regarding the size of the overall R & D budget, and decisions regarding the projects that will be undertaken, are strongly interdependent. However, there is considerable evidence that many business firms employ R & D decision heuristics that separates these dimensions, at least at first cut. Thus setting the R & D budget as a certain fraction of sales seems to be a widely used heuristic for the first stage, at least, of the budgetary process [26].

* For a rich and sophisticated discussion see Simon and Newell [54].

There similarly seem to be certain widely used heuristics for hunting for promising projects to undertake, which proceed independently of the research budget, save for an 'adding up' constraint.

It is convenient to employ the term 'strategy' to refer to a stable widely used set of heuristics for project selection. A model of a strategy ideally would be able to pick out, probabilistically, the nature of the R & D projects that would be selected by an organization employing it, given certain conditions under which project selection proceeds, for example, the pattern of consumer demand for different products and product improvements, and the state of scientific knowledge (which affect project ranking) and total firm sales (which affect the firm's total R & D budget). The outcomes, in terms of the degree and kind of success achieved from the projects actually undertaken, also will be probabilistic, given project heuristics and external conditions. Since project selection implies project heuristics, an R & D strategy can be viewed as defining a probability distribution of number and kinds of innovations, given certain variables that influence project selection and project outcome.

A considerable portion of research on the generation of innovation can be viewed as attempts to find, describe, and analyze the consequences of the employment of certain widely studied strategies. One would expect that the nature of strategies would differ greatly depending on the nature of the organization being studied. For example, government agencies would have different strategies than business firms; however, for the most part research has focussed on business firms, and this will be our focus here.

A good R & D strategy must attend to factors on the demand side and factors on the supply side. It is no good to pick out projects that are technologically exciting and doable, but which have no demand, or to undertake projects which if successful would have a high payoff, but where there is no chance of success. However, one can imagine a strategy that focusses on the 'demand' side and picks out a list of inventions that if made would have a good market, for later screening according to feasibility and cost. Or one can imagine a strategy that initially picks out projects where major technological breakthroughs seem possible, for later checks for market ability. A considerable literature has evolved that implicitly posits that R & D-strategies can be dichotomized into these two. The first strategy has been named 'demand-pull'; the second 'capabilities-push'*. Both, presumably, pick out projects within some kind of a prior budget constraint (say R & D as a certain percent of sales).

* For a discussion see Pavitt [42] and Freeman [9].

Of course, if one were committed to the proposition that project choice was truly maximizing, then it should make no difference whether choice proceeded by first listing projects with good demand prospects and then doing a cost or feasibility check on this list, or by prescreening of technical feasibility and then checking for demand. The order of calculation would be irrelevant. However, if one views the first cut as defining a region for consideration, with subsequent more detailed calculation over the selected subset, demand-pull and capabilities-push are quite different strategies. In a given external regime of stimulus, they will select different projects, and presumably they will have different consequences in terms of the payoffs achieved.

Several different studies have concluded that, if strategies can be so dichotomized, demand-pull is by far the more common of the two [42]. Further, when applied, demand-pull is more likely to result in a commercially successful project than a strategy of capabilities-push. However, capabilities-push selected projects, when they do pay off, pay off handsomely. None of these results is surprising. One might well have expected that screening first for innovations which, if they could be achieved, would yield high payoffs, would focus attention on a more fruitful part of the choice set than screening first for things that technological wizardry can accomplish. One also might have expected that, when the R & D cost or feasibility criteria were applied to a project list preselected on demand considerations, the outcome would be a project the objectives of which can be achieved with modest cost and high confidence. However, the 'look first at technological possibilities' strategy should occasionally come up with a big winner. Thus categorizing R & D strategies in this way would appear to be a fruitful intellectual endeavor.

However, if one thinks about it for a moment, both a pure demand-pull and a pure capabilities-push strategy would appear to be naive. One might posit that a strategy that involves more backing and forthing between demand and supply side considerations would be more powerful. Further, it is patently implausible that project generators look first at 'all' R & D projects with high demands, or at 'all' major achievable breakthroughs. These sets are too vast. Also, the very staffing of an R & D organization inevitably limits the range of projects it has the capabilities to undertake.

We propose that most R & D organizations de facto are limited to strategies which involve a precommitment to one or a small number of classes of R & D projects each of which has a certain similarity of broadly defined targets, procedures for reaching these targets, and R & D resources required. Binswanger [2] has used the term 'R & D activity' in a way that captures what we mean here. Following the work of Hayami and Ruttan, he notes that in agriculture one can identify various classes of R & D project. One can

distinguish between projects aimed to improve mechanical equipment, projects that aim to improve seed varieties, etc. Within these classes both the targets, and the procedures followed by those doing the R & D, have a certain similarity. Relatedly, an R & D activity defines, or is defined by, certain kinds of skills, equipment, and organization. He proposes that most R & D projects, at least in agriculture, can be viewed as belonging to one of a relatively small group of R & D activities, in the sense above. An R & D organization can staff itself to work with one or a few of these R & D activities. This involves a strong limitation on its R & D strategy. Given this precommitment, demand-pull or technology-push may guide the period by period details of choice. Or other procedures may be employed.

In the Binswanger formulation the set of projects associated with a particular R & D activity is closely circumscribed. Thus there is no way to try to adjust for changes in demand conditions, or changing costs of R & D, operating within that activity. Rather, changes in the kind of R & D done, in response to changes in external conditions, must come about through changes in the mix of R & D activities employed. We see no inevitable reason why the examples of R & D activity, given by Binswanger, and Hayami and Ruttan, must possess this internal inflexibility. For example, mechanical devices comprise a large class of items, and the various members of the class each possess a number of relevant dimensions. Certainly, R & D aimed at improving mechanical design can aim in a variety of directions. Similarly, there are a number of possible attributes one might aim for in trying to develop a new seed variety. In any case, the economists' bias is that a viable strategy ought to be able to adjust to changing demand and cost conditions, and thus must involve a mix of activities, or a very flexible activity.

However, it may be that there are certain powerful intra project heuristics that apply when a technology is advanced in a certain direction, and payoffs from advancing in that direction that exist under a wide range of demand conditions. We call these directions 'natural trajectories'. If natural trajectories exist, following these may be a good strategy.

3.3. *Natural trajectories*

In some of the writing on technological advance, there is a sense that innovation has a certain inner logic of its own. In contrast with the central focus of the economists' models – a fine tuned responsiveness to demand conditions and their changes, and a scan of a wide range of projects to assess profitability – particularly in industries where technological advance is very rapid, advances seem to follow advances in a way that appears somewhat

'inevitable' and certainly not fine tuned to the changing demand and cost conditions. Rosenberg talks of 'technological imperatives' as guiding the evolution of certain technologies; bottlenecks in connected processes, obvious weak spots in products, clear targets for improvement, etc. [46, 47]. These provide sharp signals that certain R & D projects are doable and are worth doing under a wide range of particular demand and cost conditions. Marginal changes in external conditions influence at most the ranking in terms of profitability of the set of profitable projects associated with pushing technology in a particular direction. We suggest that such natural trajectories are important, and can be studied.

In many cases natural trajectories are specific to a particular technology or broadly defined 'technological regime'. We use the 'technological regime' language to refer to much the same kind of thing as Hayami and Ruttan mean by a meta production function. Their concept refers to a frontier of achievable capabilities, defined in the relevant economic dimensions, limited by physical, biological, and other constraints, given a broadly defined way of doing things. Our concept is more cognitive, relating to technicians' beliefs about what is feasible or at least worth attempting. For example, the advent of the DC3 aircraft in the 1930's defined a particular technological regime; metal skin, low wing, piston powered planes. Engineers had some strong notions regarding the potential of this regime. For more than two decades innovation in aircraft design essentially involved better exploitation of this potential; improving the engines, enlargening the planes, making them more efficient*.

Binswangers R & D activities, our R & D strategies, often are intimately connected with a given technological regime, in a sense not well developed by Binswanger. The sense of potential, of constraints, and of not yet exploited opportunities, implicit in a regime focuses the attention of engineers on certain directions in which progress is possible, and provides strong guidance as to the tactics likely to be fruitful for probing in that direction. In other words, a regime not only defines boundaries, but also trajectories to those boundaries. Indeed these concepts are integral, the boundaries being defined as the limits of following various design trajectories.

In many cases the promising trajectories and strategies for technological advance, within a given regime, are associated with improvements of major components or aspects thereof. In aviation, engineers can work on improving the thrust-weight ratio of engines, or on increasing the lift-drag ratio of airframes. General theoretical understanding provides clues as to how to

* See Miller and Sawers [30] and Phillips [44].

proceed. In jet engine technology, thermodynamic understanding relates the performance of the engine to such variables as temperature and pressure at combustion. This naturally leads designers to look for engine designs that will enable higher inlet temperatures, and higher pressures. In airframe design, theoretical understanding (at a relatively mundane level) always has indicated that there are advantages of getting a plane to fly higher where air resistance is lower. This leads designers to think of pressurizing the cabin, demanding aircraft engines that will operate effectively at higher altitudes, etc.

Often there are complementarities among the various trajectories. Advances in engine power, and streamlining of aircraft, are complementary. Developing of seeds that germinate at the same time and grow at the same rate facilitates mechanical harvesting.

While natural trajectories almost invariably have special elements associated with the particular technology in question, in any era there appear to be certain natural trajectories that are common to a wide range of technologies. Two of these have been relatively well identified in the literature: progressive exploitation of latent scale economies, and increasing mechanization of operations that have been done by hand.

There are a wide variety of industries and technologies where the advance of equipment involves the exploitation of latent scale economies. In chemical process industries, in power generation, and in other sectors where designing of equipment of larger capacity will permit output expansion without a proportional increase in capital or other costs, the objectives of cost reduction apparently lead designers to focus on making equipment larger. Hughes has documented the way in which designers of electric power equipment have aimed progressively to push forward the scale frontier [17]. Levin has provided a general theoretical discussion of the phenomenon and provided case studies of the process in operation in the manufacture of sulfuric acid, ethylene, ammonia, and petroleum refining [25]. In the development of aircraft technology, designers long have understood that larger planes in principle could operate with lower costs per seat mile. Of course in aviation, as in electric power, the possibilities for exploiting latent scale economies are limited by the market as well as by engineering. In aviation, high volumes and long hauls provide the market targets of opportunity. And, historically, these have tended to grow in importance over time. This has permitted engineers to follow their design instincts. As a rule, each generation of commercial aircraft has tended to involve larger vehicles than those in the predecessor generation. Another quite common natural trajectory is toward mechanization of processes that have been done by hand. Mechanization seems to be viewed by designers of equipment as a natural way to reduce costs, increase reliability

and precision of production, gain more reliable control over operations, etc. This point has been stressed by Rosenberg in his study of 19th century innovation in American industry [46]. That this tendency to mechanize still exists has been suggested by Piore [43], and documented in considerable detail by Setzer [53] in her work on the evolution of production processes at Western Electric. Inventors and research and development engineers, operating under a higher order objective to look for inventions and design chances that will reduce costs, look for opportunities to mechanize. Engineers, through training and experience, apparently acquire heuristics that assist the design of machinery. For this reason, hunting for opportunities for mechanization, like trying to exploit latent scale economies, can serve as a useful focuser for inventive activity.

David, in a fascinating and important recent essay, proposes a different but complementary hypothesis [7]. While the studies above point to 'easy invention' in directions that increase the capital labor ratio, David suggests that in the late 19th century technologies that already were capital intensive were easier to improve in a 'neutral' direction, than were technologies that involved a lower degree of capital intensity. Here the proposition was that during that period of time there was a 'lot of room' for improving mechanized operations, and engineer-designers had some clever ways of moving in that direction.

Exploitation of latent scale economies, and opportunities for further mechanization, are important avenues for technological advance in many sectors at the present time, as well as in the 19th century. Many of the studies cited above are of relatively contemporary examples. However, there is no reason to believe, and many reasons to doubt, that the powerful general trajectories of one era are the powerful ones of the next. For example, it seems apparent that in the 20th century two widely used natural trajectories opened up (and later variegated) that were not available earlier: the exploitation of understanding of electricity and the resulting creation and improvement of electrical and later electronic components, and similar developments regarding chemical technologies. As with the case of mechanization during the 19th century, these developments had several different effects. For example, improvements in ability to understand electrical phenomena and growing experience with electrical and electronic equipment led to a substitution of these kinds of components for others. And technologies that had many and important electronic components were better able to benefit from the improvements in these components than other technologies.

It is apparent that industries differ significantly in the extent to which they can exploit the prevailing general natural trajectories, and that these dif-

ferences influence the rise and fall of different industries and technologies. During the 19th century, cotton gained ascendancy over wool in large part because its production processes were easier to mechanize. Quite possibly both Rosenberg trajectories and David trajectories were involved. In the 20th century, Texas cotton drove out southeastern cotton in good part because the area was amenable to mechanized picking. In the current era, where considerable power has been developed to design and improve synthetic products, synthetic fibers have risen in importance relative to natural ones.

One aspect of natural trajectories, whether specific to a particular technology or more general, whether 19th century or contemporary, is that underlying the movement along them is a certain knowledge on the part of the technicians, engineers, scientists, involved in the relevant inventive activity. The knowledge may be quite specific, as understanding of the tactics for hybrid development of seeds, or the operating characteristics of jet engines. The knowledge may involve more art and feel than science; this certainly was so of the knowledge behind the mechanization and scale economies of trajectories during the 19th century. But in the middle to late 20th century, many scholars have been strongly tempted by the hypothesis that underlying the technologies that have experienced the most rapid advance, or built into a key component of these, is a relatively well articulated scientific knowledge. This does not mean that the 'inventors' are active scientists, nor that 'inventing' exploits knowledge produced by recent science. But the fact that college educated scientists and engineers now comprise the dominant group doing applied research and development indicates that, at the least, scientific literacy is in the background. One then is led to the proposition that a key factor that differs across industries and that partly resolves the differential productivity puzzle is the strength of scientific understanding relevant to seeking improvements. We shall not review that tangled literature here. However, in the concluding section we will pose the question as to whether such differences, if important, are inherent or institutional.

Whether a 'strategy' involves a natural trajectory or not, whether or not there are certain identifiable R & D activities among which firms must choose some small subset, whatever the ways used to assess chances of benefits and feasibility, an R & D strategy determines R & D outcomes, probabilistically. There will be some winners and some losers. The next analytic question is, 'what next?'

4. THE SELECTION ENVIRONMENT*

4.1. Elements of the selection model

The heart of the conceptualization problem discussed in the preceding section was to characterize the generation of innovation as purposive, but inherently stochastic. Despite a tendency of some authors to try to slice neatly between invention, and adoption, with all of the uncertainty piled on the former, one cannot make sense of the micro studies of innovation unless one recognizes explicitly that many uncertainties cannot be resolved until an innovation actually has been tried in practice. While the organizations watching the flow of new innovations may be trying to behave rationally, as with the analysis of the generation of innovation, rational behavior should not be translated as maximizing behavior unless explicit recognition is made of differences in perception or luck. Relatedly, analysis of the ways that innovations are screened, some tried and rejected, others accepted and spread, must be explicitly dynamic. We propose the concept of a 'selection environment' as a useful theoretical organizer. Given a flow of new innovations, the selection environment (as we are employing the terms) determines how relative use of different technologies changes over time. The selection environment influences the path of productivity growth generated by any given innovation, and also it feeds back the influence strongly of the kinds of R & D that firms and industry will find profitable to undertake.

There is an important conceptual issue that needs to be clarified at the outset. In much of the literature on technological change a sharp distinction has been drawn between inventing and innovating (where the latter term is used, more narrowly than we are using it, to refer to a decision to try out technology in practice). The distinction harks back to Schumpeter of the *Theory of Economic Development*. While technological invention was not center piece in his analysis, regarding invention he described a world in which independent inventors had to link up with firms in being, or entrepreneurs seeking to establish new firms, to implement their inventions. We would argue that in the current institutional environment with much of innovation coming from internal R & D, the old Schumpeterian distinction is much less useful than it used to be. While there are examples of inventions that were economically viable without further R & D that simply lay around waiting for some to try them out, this seems a rare occurrence. Further, the earlier experimental use of a new technology often is integrated with the last stages of the research

* An earlier version of this section was published in Nelson and Winter [37].

and development process.

There is, however, a distinction that has some of the flavor of the old Schumpeterian distinction that ought to be recognized. Often an innovation is produced by a firm for sale to customers who will use it. Thus there are two acts of innovation (in the narrow sense of the term) that are involved. In the case of the advent of jet passenger aircraft, DeLavaland, the company that produced the first commercial jet, was an innovator. But so was the airline that bought the plane. More generally, if the focus is on any economic sector, it is useful to distinguish between two kinds of innovation. Some of these may bubble out of the research and development activities of the firms in the sector. Others may be largely in the form of materials, components or equipment offered by supplying firms. However, for the moment let us repress that distinction and focus on an economic sector which is experiencing a flow of new innovations, some of which may be viable, and others not. While the range of possible innovations, and the characteristics of the sectors, obviously are extremely diverse, the analytic task is to develop a conceptual framework which at once identifies commonalities and enables the differences to stand out.

Consider, then, the following diverse set of innovations and industries: the first model 707 aircraft produced by the Boeing Aircraft Company, the first use of the oxygen process on a commercial basis by a steel company in Austria, a new seed variety tried by a farmer, a pioneering doctor trying a anti-cancer drug, a district court trying the system of release on own recognizance without bail for select group of people accused of crime, a school trying an open classroom. The range of possible innovations, and the characteristics of the organizations that introduce them, is enormous.

A necessary condition for survival of an innovation is that, after a trial, it be perceived as worthwhile by the organizations that directly determine whether it is used or not. If the innovation is to persist and expand in use, the firm must find a new product or process profitable to produce or employ, the doctor must view the treatment as efficacious, the school system must be persuaded that the new classroom technique is good educational practice and worth the cost. We shall call all of these primary organizations 'firms' and use the term profitable to indicate value in the eyes of the firms, without implying that the objectives are money profit rather than something else, or that the organization is private, not public. Neither do we imply any social merit in firms' objectives. Firms may be motivated by little more than the prestige of being first. Sectors obviously differ in terms of the objectives of the firms.

The question of whether or not the firms find the innovations profitable

depends not only on the objectives of the firms. In almost all economic sectors the firms – for-profit private organizations, public agencies, individual professionals – are subject to monitoring mechanisms which influence the innovations that score well or poorly according to the objectives of the firms and which may more directly impose constraints on firm behavior. A key part of this monitoring mechanism involves the individuals or organizations who are the demanders or beneficiaries of the goods or services produced by the firms in the sectors. Thus the profitability to Boeing of producing 707 type aircraft depends on how the airlines react to these planes. Consumers must be willing to buy the corn that the new seed produced at a price that covers cost. Patients must agree to the new treatment. School systems and legal systems are constrained by funds proposed by high order executives, and voted by legislatures. In some sectors there are additional constraints imposed on firms by agencies that are assigned a legal responsibility to monitor or regulate their activity. Thus the Boeing 707, before it could be put into commercial use, had to pass FAA tests. New pharmaceuticals are regulated, etc. Selection environments differ greatly in the structure of demanders and monitors and the manner and strength in which these mold and constrain the behavior of firms.

There are, roughly speaking, two roughly distinct kinds of mechanisms for the spread of a profitable innovation. One of these is greater use of an innovation by the firm that first introduces it. If the firm produces a variety of products or undertakes a variety of activities, this may occur through substitution of the new activity for older ones. Or, the firm may grow, absolutely, and (if there are competitors) relatively by attracting new resources. In sectors which involve a number of administratively distinct organizational units on the supply side, there is a second innovation spreading mechanism that needs to be considered – imitation. Imitation of certain innovations may be deliberately spurred by the institutional machinery. Thus the agricultural extension service encourages widespread adoption by farmers of new seed varieties. If the innovation is produced by a supplying firm, its sales agents will try to encourage rapid adoption. Or the institutional machinery may deter or block imitation, as the patent system blocks the adoption by one firm of patented innovations created by a competitor.

The relative importance of these mechanisms differs from sector to sector. Dieselization of a nationalized railroad system must proceed largely through substitution of diesels for other kinds of railroad engines, although improvement in the service may enable a nationalized railroad system to gain additional funds. If, on the other hand, there are a number of organizationally separate railroad systems, when one railroad is a successful innovator,

spread of diesels may to a considerable extent require imitation by other railroads. The success of the 707 encouraged and enabled Boeing to expand its production facilities. And other aircraft producers were spurred, at their peril, to design and produce comparable aircraft. Bail reform has spread in part by greater use within particular districts, but since one jurisdiction is not permitted to expand relative to another, and since there are many thousands of jurisdictional districts, the ultimate spread of innovations in the Criminal Justice System depends upon imitation.

We propose that a rigorous general model of the selection environment can be built from specification of these three elements: the definition of 'worth' or profit that is operative for the firms in the sector, the manner in which consumer and regulatory preferences and rules influence what is profitable, and the investment and imitation processes that are involved. In the remainder of this section we shall discuss some important qualitative differences in sectoral selection environments that become the focus of attention once one poses the theoretical problem in the way we have. Market sectors differ significantly among themselves. And many sectors involve important nonmarket components that have special characteristics.

4.2. The market as a selection environment

The perception that market competition in a sector operates like a selection environment was explicit in the writings of many of the great 19th and early 20th century economic theorists. Schumpeter was well within the classical tradition. In a stylized Schumpeterian evolutionary system, there is both a carrot and a stick to motivate firms to introduce 'better' production methods or products. Better here has an unambiguous meaning: lower cost of production, or a product that consumers are willing to buy at a price above cost. In either case the criterion boils down to higher monetary profit. Successful innovation leads to both higher profit for the innovator and to profitable investment opportunities. Thus profitable firms grow. In so doing they cut away the market for the noninnovators and reduce their profitability which, in turn, will force these firms to contract. Both the visible profits of the innovators and the losses experienced by the laggards stimulate the latter to try to imitate.

It would seem likely that the Schumpeterian dynamics would differ somewhat depending on whether the innovation were of a new product or a new process. For product innovation, the profitability to the firm is bound tightly to the reactions of potential consumers. For process innovation, which does not change the nature of the product, consumer constraints are far more

blunt. The firm can make an assessment of profitability by considering the effects on costs, with far less concern for consumer reaction. Further, and reinforcing these differences, product innovation usually comes from a firm's own R & D; process innovations likely come from the R & D done by suppliers and are embodied in their products. To the extent this is so, imitation by a competitor of a process innovation is likely to occur relatively rapidly, and to be encouraged by a marketing supplier, rather than being retarded by a patent.

Both expansion of the innovator, and imitation by competitors are essential to the viability of Schumpeterian process. In the standard descriptions of dynamic competition, expansion of the innovator is likely to be stressed. It is surprising, therefore, that the relationship between innovation and investment has been studied hardly at all empirically. The principal studies of firm investment have been based on neoclassical theory modified by Keynesian considerations, and tend to ignore the relationship between innovation and expansion of a firm. The Meyer-Kuh retained earnings-capacity pressure theory would imply that successful innovators tend to expand [29]. Presumably a successful innovation both yields profits, and attracts demand which may, initially at least, exceed capacity. A more straight laced neoclassical theory also would predict that firms that come up with better processes and products ought to want to expand their capacity to produce. But the major studies of firm investment have, virtually without exception, ignored the influence of innovation on investment.

The exceptions are studies where the author's basic hypothesis is oriented around the Schumpeterian interactions. Mueller does find that lagged R & D expenditure by a firm has a positive influence on its investment in new plant and equipment [31]. In a later study, Grabowski and Mueller used lagged patents as a measure of R & D output, but find that the influence on plant and equipment investment is weak statistically [12]. Mansfield's studies give stronger support for a 'Schumpeterian' view. In examining investment at an industry level, he finds that the number of recent innovations is a significant explanatory variable, augmenting more traditional variables [26]. But perhaps his most interesting results involve comparisons of firm growth rates, where he finds that innovating firms in fact tend to grow more rapidly than the laggards. However, while the advantage of the innovators tends to persist for several periods, the advantage tends to damp out with time, apparently because other firms have been able to imitate, or to come up with comparable or superior innovations.

In contrast to the sparseness of studies of the relationship of investment to successful innovation, a large number of studies have focused on the spread of

innovation by diffusion (imitation) in profit-oriented sectors*. There have ranged across a variety of sectors from agriculture (study of the diffusion of hybrid corn among farmers) to railroads (diesel engines) to brewing, to steel. Many have documented the role of profitability of an innovation in influencing the speed at which that innovation spreads. However, other studies have concluded that the calculations made by firms tend to be haphazard, and that even ex-post the firms had little idea, quantitatively, how profitable the innovation turned out to be [33]. Several have found that, for innovations that are costly to put into operation, large firms (with greater financial resources) tend to adopt a new technology earlier than do smaller firms, although there are exceptions. Most of the studies show an S shaped pattern of use of the new innovation over time. In many cases this has been attributed to the fact that the later users are observing the behavior (and perhaps performance) of the earlier adopters before making their own decisions. In some instances the innovations were inputs provided by a supplier, and the early adopters of the innovation were not in a position to block subsequent use of their competitors. In other instances this was not the case. For example, a glass producing company, Pilkington, holds the basic patents on the float glass process and presumably had an interest in limiting diffusion to other firms except where Pilkington was blocked from the market. It is interesting that the analysts of diffusion have not in general been cognizant of these differences.

It also is quite surprising that in no study of which we are aware has there been an attempt to study the dual roles of expansion of the innovator, and imitation of the imitator, together. It would seem apparent that in order for a market selection environment to work effectively, a rather fine balance is required between the two mechanisms. We will return to this issue in the concluding section.

4.3. Nonmarket selection environments

While economists have concentrated their attention on market sectors, research on the selection environment of nonmarket sectors has been undertaken principally by anthropologists, sociologists, and political scientists. This in itself would have led to some significant differences in focus and analysis. But to a considerable extent the differences in analysis appear to reflect real differences in the selection environments.

* For a review see Mansfield [27].

An essential element in most theorizing about market selection environments is a relatively clear separation of the 'firms' on the one hand, and consumers and regulators on the other. Consumers' evaluation of products – versus each other, and versus their price – is presumed to be the criterion that ought to dictate resource allocation. Firms can be viewed as bidding, and competing, for consumer purchases, and markets can be judged as working well or poorly depending on the extent to which the profitability of a firm hinges on its ability to meet consumer demands as well as or better than its rivals. The viability of an innovation should depend on consumers' evaluation of it.

A hallmark of nonmarket sectors is that the separation of interests between firms and customers is not as sharply defined as in market sectors. The relationship between a public agency, like a school system, and its clientele (students and parents) and sources of finance (mayor, council, and voters), simply does not have an arm's length distance quality that marks the relationship between seller and potential buyer of the new car. Relatedly, the question of how legitimate values are to be determined is much more complex than in market sectors. The public agency is expected to play a key role in the articulation of values, and to internalize these and work in the public interest of its own volition. This is so in many nominally 'private sector' activities, like the provision of medical services by doctors. The doctor is not supposed to make his decisions regarding the use of a new drug on the basis of whether this will profit him, but rather on his expectation of how this will benefit his patients. Further, he is supposed to know more about that than do his patients. This is not to say that in fact interests of firms and consumers always are consonant. In most nonmarket sectors (as in market sectors where competition is lax) the firm has a good deal of discretionary power regarding what it is to provide, and the customer may have little direct power to reward or to punish performance. But in general the appropriate 'control' mechanism over a provider of goods and services in a nonmarket sector is not viewed as competition among providers for the consumer dollar. For these reasons, the motivations of the firms in a nonmarket sector cannot simply be presumed to be monetary profit. This makes analysis of the operative values relating to acceptance or rejection of an innovation difficult. As in the theory of consumer behavior, as contrasted with the text book theory of the firm, tastes matter, these may be hard to analyze, and may not be stable. Even in situations where there is a relatively clear cut goal, and the decision to employ an innovation or not hinges on assessment of efficacy relative to that goal, it has proved hard to identify relevant criteria. Thus, in the Coleman, Katz and Menzel study of the diffusion among physicians of a new pharmaceutical, the authors did not even attempt to specify quantita-

tively the ways in which the new product was superior medically to pre-existing alternatives [5]. In Warner's study of the decision by doctors to use new chemotherapeutic techniques for the treatment of cancer, in several of the cancer varieties where a significant fraction of patients were so treated, there was no quantitative evidence that the therapy had any effect [58]. The physicians made their decisions on hope but on no objective evidence. Friedman, in his study of the acceptance and spread of a certain program of bail reform, was able to identify a few rather specific reasons why the key agencies might find the reform attractive [10]. But the reasons were largely qualitative and it is interesting that, after adopting the reform, there was no real monitoring to check that the programs were performing as hoped. In fact, the performance of the program eroded over time in at least one key dimension, and no one noticed.

Political and regulatory control over firms cannot provide the pervasive, if not always coercive set of value signals and incentives that is provided by consumer sovereignty in market sectors. Thus there is greater room left for autonomous and discretionary behavior on the part of suppliers. However, the employment of regulatory and political mechanisms of governance, as contrasted with consumer sovereignty, means that in many cases several different parties may have to go along before an innovation can be operative. In Friedman's study of bail reform, the police and the courts both had to agree to the proposal, and legislative agreement was necessary where budget was involved. Government agencies often have to gain specific agreement from both political chief executives and legislature before they can proceed with a new program.

Nonseparation of suppliers and demanders leaves little room for firms to compete among each other for consumer dollars. Where there is a single supplying entity — like the United States Postal Office, or the Department of Defense — spread of use of an innovation is a matter of internal decision making constrained and pressured to some degree by the higher order political processes. Where there is a range of suppliers — as in medicine, or in state and local governmental agencies — innovations must spread largely through imitation across the spectrum of noncompeting firms. At the same time there is no incentive for the innovating firm to deter imitation. Organizations that cannot expand into the terrain of others and know that others cannot encroach on their territory have little to gain from preventing others from adopting their successful innovations. Indeed, in most of the sectors under consideration here, there are formal arrangements for cooperation and flow of information across firms. In many, professional organizations set values and judge the merit of new innovations. Professional stamp of approval, and

the adoption pattern it stimulates, often are the only criteria for merit available to a non-professional.

Consider the quasi market for the provision of physician services. Without strong constraints afforded by consumers or outside regulators, consumer welfare is guarded (perhaps not so securely) largely by professional standards of efficacy of treatment. To assess the efficacy of new treatments doctors consult with each other and apparently aim for professional consensus guided by the judgment of certain key experts. Mohr's study of the spread of new practices and policies across local public health services reveals a similar professionalism at work [32]. Walker's study of the lead and lag pattern among state governments in the adoption of new programs indicates the presence of regional groups with intraregional leaders (generally populous, urban, and wealthy states) to which officials in departments in other state governments look for references and models [57].

Professional judgments are moderated by political constraints on spending limits, and other governmental regulatory processes which impinge on decision making in a more detailed way. Thus in Mohr's study, the speed with which a local public health service adopts new practices was found to be positively related to the extent to which public health professionals were in control of the key office. However, the professional bias toward adoption of new techniques was moderated by political and budgetary constraints. These, which had to do with the composition and presumably the attitudes of the local 'consuming' populations, did limit, if in a loose way, the innovations that local public health services could afford to adopt. Similarly, Walker's study showed that budgetary constraints imposed by state political systems significantly moderated the proclivity of state officials to adopt progressive programs (read, programs adopted by other states whose judgments they admired).

Crain's study of the spread of fluoridation across American cities is perhaps the most revealing example of a sector in which the 'firms' have a bias toward adopting an innovation based on notions of professional or technical appropriateness, but consumers tend to resist it [6]. He notes that the spread of fluoridation first occurred quite rapidly, in a context where local health professionals were in charge of the decision. As time elapsed, fluoridation became a more openly political issue, and mayors began to take the decision-making authority out of the hands of the professionals. The spread of fluoridation slowed significantly. Still later, it became common for citizen referendum to become the vehicle for decision. This development brought the spread of fluoridation to a virtual halt.

The pattern in all of these cases is quite different from that in the market

sectors studied by economists. It is, however, easy enough to see the same broad elements of modeling that need to be stressed: motivations of the firms in the sector (in general, not characterizable in terms of monetary profit), the ways (if any) in which consumers (often voters) and financiers (often legislatures) constrain firm behavior, and the mechanisms of information and value sharing among firms in the imitation process (which is the dominant mechanism by which an innovation spreads).

5. THOUGHTS ON THE EFFECTS OF INSTITUTIONAL STRUCTURE

The preceding two sections sketched some pieces that we hope can be filled in, extended, added to, so that a useful theory of innovation may evolve. The conceptualization has taken very seriously the two critical requirements for a useful theory of innovation that were identified earlier: that innovation be treated as inherently stochastic, and that the formulation be capable of encompassing considerable institutional complexity and variety.

Simply keeping in mind the uncertainty and institutional diversity surrounding innovation can help make thinking about policy issues more sophisticated than has been the norm. Consider, for example, the literature that has evolved on the role of firms with substantial market power in the innovation process*. To a distressing extent, that literature has placed the problem in a static frame. Yet Schumpeter, certainly the primal intellectual sources of the current discussion, never viewed the innovation problem statically. Always he had at the center of his analysis that innovation was an uncertain business, that it was important to get new things tried out, to sort out the good from the bad, and that doing this effectively was the principal virtue of capitalistic competition. In his *Theory of Economic Development*, his argument about the inherent monopoly power of successful innovators was concerned explicitly with transient monopoly as a consequence of, and lure for, innovation — not structural monopoly as a base for innovation. There is something more of a structural argument in his *Capitalism, Socialism and Democracy*. But the desirable structure that he proposed is not that of sheltered monopoly. The firms of the *Theory of Economic Development* have grown larger, but 'The Process of Creative Destruction' scarcely is a proposal for a quiet life for monopolists.

Further, most of the analysis has proceeded as if the presence, or lack of it, of

* For a fine review of that literature see Kamien and Schwartz [19].

large firms with considerable market power were the key institutional difference between economic sectors. In fact, intersectoral differences in the presence of 'large firms' is not a powerful explicand of inter-sectoral differences of productivity growth. If this were the key institutional variation possible, there would be little hope of designing powerful policy. But, of course, there is far more to institutional variation than differences in the average size or market power of firms. In some of the sectors the critical institutions are not firms at all in the ordinary sense (e.g., medical care, garbage collection, etc.). And, even where it can be persuasively demonstrated that large firms can internalize benefits from R & D much more completely than can small firms, an alternative policy would be to establish governmental or not-for-profit organizations to carry on these activities, in lieu of reliance on large firms.

The reason we have stressed uncertainty and institutional complexity and diversity is that these salient attributes of innovation have been ignored in much of the literature concerned with policy towards innovation. This is particularly so in the literature focussed on the inter-industry productivity growth puzzle, which we have proposed is perhaps the most important phenomenon that needs to be comprehended if analysis is seriously to contribute to the policy dialogue. However, merely recognizing these aspects cannot carry us very far.

We believe that the two major theoretical proposals we have introduced above – modeling innovation generation as the conditional probabilistic outcome of various R & D strategies, and modeling the fate of an innovation in terms of the workings of a selection environment – have promise of taking us further. For example, the 'selection environment' language seems useful for describing some of the important institutional differences among sectors, and for beginning to explore some of the consequences of these for the inter-industry productivity growth puzzle. Clearly there is considerable variation among sectors in their 'selection' environments, and these differences can affect both the speed and extent of spread of any innovation. It might be conjectured, at first thought, that these differences would influence the level of productivity at any time but not its rate of growth. We think this is wrong for two reasons. First, even if one assumes that the rate of advance of 'best practice' is not influenced by the selection environment, it is not apparent that sectors need be characterized by a constant ratio of average to best practice. In some non-market sectors it is hard to identify strong forces that will prevent that ratio from falling. This suggests that a particularly fruitful arena for investigation is the nature of innovation evaluating and information dissemination systems in non-market sectors (since imitation carries such a large part of the

load in these sectors for spreading of the innovation). If we understood these better, it is likely that we would see how to improve upon them.

Second, and even more important, the selection environment feeds back to influence the incentives for R & D by the firms in a sector. We propose that our 'selection environment' concept is much better suited than conventional economic concepts for probing at 'externalities' from R & D. Regarding non-market sectors it is hard to make many generalizations, even tentatively. The most important one is that there is likely to be only a casual connection between the incentives for R & D on the part of public or not-for-profit suppliers, and consumer benefits.

Regarding market sectors, the situation clearly is more complex than economists writing about 'externality' have tended to recognize, and the externalities are inextricably connected with dynamics. For example, consider the comparison we have made elsewhere of selection environments in agriculture and in aircraft production [35]. The fact that producers of aircraft can expand their plant rapidly, and that imitation is hard, provides great stimulus to the firms to do R & D, perhaps over stimulus. On the other hand, in agriculture, the fact that expansion is slow compared with imitation means that there is little incentive to firms to do R & D. Support for R & D is dependent upon suppliers, and public agencies. The division between 'own' and 'supplier' R & D, and between private and public finance, analyzed in the differential productivity growth literature, clearly reflects underlying structural differences of this sort. The selection environment concept enables these to be analyzed. One can interpret the naive form of the 'Schumpeterian' hypothesis as positing that the returns to R & D are internalized to a greater extent when there are a few large firms than when there are many small ones. But clearly there is a lot more going on than merely differences in size of firms, and further, the sizes of firms are largely endogenous to the system, not a given.

Indeed the explicit dynamic treatment of the selection environment enables one to see potential anti-trust problems, which although they have been discussed verbally, defy formal treatment within traditional models*. Where, as in aviation, a significant innovation can give one firm a major advantage over others, where firms can grow rapidly and there are a few limits on firm size, and where imitation is difficult, monopoly structure may develop out of the competitive process itself, and for reasons that textbook discussions focusing on economies of scale may badly obscure. The question of what to

* For a discussion see Phillips [44].

do about monopoly structure in the production of civil aircraft should that develop, or what to do about IBM, is extremely complex. But at least the theoretical structure proposed here enables these issues to be seen.

In contrast with discussion of selection environments where the focus was explicitly institutional, there was little explicit institutional discussion in our analysis of R & D strategies. In particular, the discussion of natural trajectories, which we conjectured (interpreting the conjectures of others) were associated with sectors where technological advance has been most rapid, may have given the flavor of 'innate' differences. However, we suggest that it is an open question whether it is inevitable that natural trajectories exist for certain technologies but not for others. We have put forth the proposition that underlying natural trajectories there is a certain body of knowledge that makes the traverse relatively easy, and that in the recent half century formal science has been an important part of that knowledge. The key question then becomes: to what extent are the directions in which science advances inevitable, and to what extent can these be molded by conscious policy.

It is apparent that the evolution of basic scientific understanding has a certain logic, and momentum, of its own. To try to guide that evolution with any precision or to believe that improvements of fundamental understanding can be won simply because the payoffs are high, is foolishness. However, the fields of aerodynamics and applied thermodynamics did not evolve as rapidly and fruitfully as they did merely because they were 'ripe' and groups of academics were interested in them. Rather, they were carefully nurtured and funds and institutions provided for their development. Similarly, both in agriculture and in medicine public institutions and public funds have been established to further the advance of knowledge that feeds into the technologies.

We propose the following. While externalities pervade the innovation process they are greatest in the activities that generate understanding and data. In all of the sectors that have been blessed with strong scientific underpinnings to their technologies, institutions other than the 'firms' in the sector have played a major role in developing that science. In many cases these institutions are 'universities' and the fields defined by academic interests. But in some cases it has been possible to establish institutions that have over the years successfully built up a scientific knowledge relevant to a technology, and which continue to expand that base of knowledge. Study of these cases should be high on the priority list of anyone interested in the problem posed by the great imbalance of technological advance.

But we do not want to press these particular arguments too far. Rather, these kinds of questions, and tentative propositions, are put forth as illustrative of

those that arise rather naturally from our proposed theoretical formulation. We hope we have persuaded you that, at the least, there is some promise here for a view of innovation, and a language for talking about innovation, that can integrate a non-trivial share of the present congeries of disjoint research traditions. That would be an important step forward.

REFERENCES

- [1] W. Baumol, The Macro Economics of Unbalanced Growth, *American Economic Review* 57 (June 1967) 415.
- [2] H. Binswanger, A Micro Economic Approach to Induced Innovation, *Economic Journal* 84 (December 1974) 940.
- [3] H. Binswanger and V. Ruttan, *The Theory of Induced Innovation and Agricultural Development* (John Hopkins, Baltimore, 1976).
- [4] M. Brown and A. Conrad, The Influence of Research and Education on CES Production Relations; M. Brown (ed.) *The Theory and Empirical Analysis of Production* (Columbia University Press for NBER, 1967).
- [5] J. Coleman, E. Katz and H. Menzel, The Diffusion of an Innovation Among Physicians, *Sociometry* 20 (December 1957) 253.
- [6] R. Crain, Fluoridation: The Diffusion of an Innovation Among American Cities, *Social Forces* (June 1966) 467.
- [7] P. David, *Technical Change, Innovation and Economic Growth* (Cambridge University Press, Cambridge, 1974).
- [8] R. Evanson and Y. Kislev, *Agricultural Research and Productivity* (Yale University Press, 1975).
- [9] C. Freeman, *The Economics of Industrial Innovation* (Penguin, Harmondsworth, 1974).
- [10] L. Friedman, *Innovation and Diffusion in Nonmarkets: Case Studies in Criminal Justice*, Yale University, Ph.D. Dissertation (1973).
- [11] R. Gilpin, *Technology, Economic Growth and International Competitiveness: A Report Prepared for the Joint Economic Committee of the U.S. Congress* (July 19, 1975, USGPO, Washington).
- [12] H. Grabowski and D. Mueller, Managerial and Stockholder Welfare Models of Firms Expenditures, *Review of Economics and Statistics* 54 (Feb. 1972) 9.
- [13] Z. Griliches, Hybrid Corn: An Exploration in the Economics of Technological Change, *Econometrica* 25 (October 1957) 501.
- [14] Z. Griliches, Research Expenditures and Growth Accounting, in B. R. Williams (ed.) *Science and Technology in Economic Growth* (Halsted Press, New York, 1973).
- [15] H. J. Habakkuk, *American and British Technology in the 19th Century* (Cambridge University Press, Cambridge, 1962).
- [16] Y. Hayami and D. Ruttan, *Agricultural Development: An International Perspective* (John Hopkins, Baltimore, 1971).
- [17] W. Hughes, Scale Economies and Electric Power in W. Capron (ed.) *Technical Change in Regulated Industries* (Brookings, Washington, 1971).
- [18] J. Jewkes, D. Sawers and R. Stillerman, *The Sources of Innovation* (Norton, New York, 1961).
- [19] M. Kamien and N. Schwartz, Market Structure and Innovation: A Survey, *Journal of Economic Literature* 13 (March 1975) 1.
- [20] J. Kendrick, *Productivity Trends in the United States* (Princeton University Press for the NBER, 1961).

- [21] J. Kendrick, *Postwar Productivity Trends in the United States, 1948-1969* (Princeton U. Press for the NBER, 1973).
- [22] C. Kennedy and R. Thirlwall, Surveys in Applied Economics: Technical Progress: A Survey, *Economic Journal* 82 (March 1972) 11.
- [23] A. King, *Science and Policy: The International Stimulus* (Oxford University Press, 1974).
- [24] W. Leonard, Research and Development in Industrial Growth, *Journal for Political Economy* 79 (March and April, 1971) 232.
- [25] R. Levin, *Technical Change, Economies of Scale, and Market Structure* (Yale University Ph.D Dissertation, 1974).
- [26] E. Mansfield, *Industrial Research and Technological Innovation* (Norton, New York, 1968).
- [27] E. Mansfield, Determinants of the Speed of Application of New Technology, in: B. R. Williams (ed.), *Science and Technology in Economic Growth* (Halsted Press, New York, 1973).
- [28] E. Mansfield, The Contribution of R & D to Economic Growth in the U.S., *Science* 175 (February 1972) 477.
- [29] J. Meyer and E. Kuh, *The Investment Decision: An Empirical Study* (Harvard University Press, 1957).
- [30] R. Miller and D. Sawers, *The Technical Development of Modern Aviation* (Routledge and Keegan Paul, 1968).
- [31] D. Mueller, The Firms Decision Process: An Econometric Investigation, *Quarterly Journal of Economics* LXXXI (Feb. 1967) 58.
- [32] L. Mohr, The Determinants of Innovation in Organization, *The American Political Science Review* 63 (1969) 111.
- [33] L. Nasbeth and G. Ray, *The Diffusion of New Industrial Processes* (Cambridge University Press, 1974).
- [34] R. Nelson, M. J. Peck and E. D. Kalachek, *Technology, Economic Growth and Public Policy* (Brookings, Washington, 1967).
- [35] R. Nelson and S. Winter, Neoclassical Vs. Evolutionary Theories of Economic Growth, *Economic Journal* 84 (December 1974) 886.
- [36] R. Nelson and S. Winter, Factor Price Changes, and Factor Substitution in an Evolutionary Model, *Bell Journal of Economics* 6 (Autumn 1975) 466.
- [37] R. Nelson and S. Winter, Growth Theory from an Evolutionary Perspective: The Differential Productivity Growth Puzzle, *American Economic Review* 65 (May 1975) 338.
- [38] R. Nelson, S. Winter and H. Schuette, Technical Change in an Evolutionary Model, *Quarterly Journal of Economics* 90 (February 1976) 90.
- [39] R. Nelson and S. Winter, Dynamic Competition and Technical Progress, in: B. Balassa and R. Nelson (eds.) *Private Incentives, Social Values and Public Policy: Essays in Honor of William Fellner* (North-Holland, Amsterdam, 1976).
- [40] W. Nordhaus, Some Skeptical Faults on the Theory of Induced Innovation, *Quarterly Journal of Economics* 87 (May 1973) 208.
- [41] S. Nyman and A. Silberston, *An Approach to the Study of the Growth of Firms*, paper presented at Second Conference of Economics of Industrial Structure, Holland (1971).
- [42] K. Pavitt, *Conditions of Success in Technological Innovation* (OECD, Paris, 1971).
- [43] M. Piori, The Impact of the Labor Market upon the Design and Selection of Productive Techniques within the Manufacturing Plant, *Quarterly Journal of Economics* 82 (November 1968) 602.
- [44] A. Phillips, *Technology and Market Structure* (Heath Lexington, 1973).

- [45] F. Raines, The Impact of Applied Research and Development on Productivity, Washington University Working paper No. 6814.
- [46] N. Rosenberg, *Technology and American Economic Growth* (Harper Torch Books, New York, 1972).
- [47] N. Rosenberg, The Direction of Technological Change: Inducement Mechanisms and Focusing Devices, *Economic Development and Cultural Change* 18 (October 1969) 6.
- [48] W. E. H. Salter, *Productivity and Technical Change* (Cambridge University, 1966).
- [49] J. Schmookler, The Changing Deficiency of the American Economy, 1869–1938, *Review of Economics and Statistics* 34 (August 1952) 214.
- [50] J. Schmookler, *Innovation and Economic Growth* (Harvard University Press, 1966).
- [51] J. Schumpeter, *The Theory of Economic Development* (Harvard University Press, 1934).
- [52] J. Schumpeter, *Capitalism, Socialism and Democracy* (Harper and Row, New York, 1950).
- [53] F. Setzer, *Technical Change over the Life of a Product: Changes in Skilled Inputs and Production Processes*, Yale University Ph.D Dissertation (1974).
- [54] H. Simon and A. Newell, *Human Problem Solving* (Prentice Hall, New York, 1972).
- [55] Science Policy Research Unit, *Success and Failure in Industrial Innovation* (Center For Study of Industrial Innovation, London, 1972).
- [56] N. Terleckyj, *The Effect of R & D on Productivity Growth in Industries* (National Planning Association, November 1974).
- [57] J. Walker, The Diffusion of Innovation Among American States, *American Political Science Review* 63 (September 1969) 880.
- [58] K. Warner, *Diffusion of Leukemia Chemotherapy*, Yale University Ph.D Dissertation (1974).