Strategic Management Journal
Strat. Mgmt. J., **33**: 1384–1410 (2012)
Published online EarlyView in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.1985
Received 7 December 2010; *Final revision received 7 April 2012*

THE DYNAMICS OF WEALTH, PROFIT, AND SUSTAINABLE ADVANTAGE

MICHAEL G. JACOBIDES,[1]* SIDNEY G. WINTER,[2]
and STEFAN M. KASSBERGER[3]
[1] *London Business School, London, U.K.*
[2] *The Wharton School of the University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.*
[3] *Frankfurt School of Finance & Management, Frankfurt, Germany*

This paper shows how idiosyncratic resources can drive sustained profitability and persistent heterogeneity under competitive conditions. Generic inputs purchased in the market become idiosyncratic resources as the result of firms' investments in customization. Analytically, we show how heterogeneous firms coexist in equilibrium as a function of customization costs. Computationally, we show that sustainable profits can emerge without 'monopolistic' imperfections. We consider how capability heterogeneity, resource customization cost, and ease of expansion interact to drive short-run and sustainable profits. Our results illustrate that sustainable profits may represent a small part of the total wealth created over time by a firm or industry, and that changes in factors shaping a sector's evolutionary trajectory may be more important than changes in factors that determine profits' ultimate sustainability. Copyright © 2012 John Wiley & Sons, Ltd.

INTRODUCTION

The question of how profitability emerges and evolves has long been central in both strategic management and industrial organization research. In industrial organization and game theory, and strategy research drawing from it, attention has been focused on oligopolistic profits. Research in the widely accepted 'resource-based view' (RBV) has focused instead on the conditions under which, regardless of output-market imperfections, firms can have (persistent) profits. Our paper extends this approach. The RBV starts from the hypothesis that firms differ because they draw on different resources, which makes them differentially effective (Barney, 1991; Wernerfelt, 1984). Superior returns at the firm level have become more and more clearly identified with persistent scarcity rents on superior resources. The characteristics that resources must have if they are to yield such persistent returns have been extensively discussed; less attention has been devoted to the factors determining who gets the rents (but see Coff, 1999). RBV research has concluded that fairly stringent criteria must be met if firms are to become abnormally profitable in equilibrium (Barney, 1991; Peteraf and Barney, 2003). It has also shed light on some of the dynamic processes through which firms might establish the preconditions for capturing persistent returns in equilibrium (Dierickx and Cool, 1989).

Almost all the theoretical research in the strategy field has focused on how, *in equilibrium,* some firms are still more profitable than others after the hand of selection has eroded temporary profits. Guided by the concept of 'sustainable advantage,'

Keywords: competitive advantage; sustainability; heterogeneity; industry evolution; rent appropriation
*Correspondence to: Michael G. Jacobides, London Business School, Sussex Place, Regent's Park, London NW1 4SA, United Kingdom. E-mail: mjacobides@london.edu

Copyright © 2012 John Wiley & Sons, Ltd.

the goal has been to understand the conditions under which profits can be maintained in equilibrium, and to identify resource characteristics and firm strategies associated with this result.[1] While useful, this approach is too narrow.

In this paper, we extend and clarify the relationship between firm heterogeneity and profitability under competitive conditions. Our dynamic analysis emphasizes the importance of out-of-equilibrium profitability in long-run industry evolution. We show how persistent heterogeneity and sustainable profits emerge together from the historical trajectory of firm investments in idiosyncratic resources and capabilities. These investments establish an economically significant distinction between generic *inputs* and the idiosyncratic *resources* used by a firm in its own way. The persistence in equilibrium of the *returns on those investments* is, we claim, what 'sustainable advantage' and the resulting 'sustainable profits' are about.

Our focus on 'fully competitive conditions' is not motivated by skepticism about the existence of noncompetitive, or imperfectly competitive, conditions. Many situations depart from the competitive standard *to some degree*—but that degree matters. Where it is small, a competitive analysis may have direct empirical relevance. More generally, such analysis is valuable as a baseline or first approximation for assessing outcomes, particularly because the 'competitive case' is usually quite sharply defined relative to the innumerable alternatives.

After explaining our theoretical proposal, we formalize the analysis in a specific stylized model of industry evolution. The model includes a market in which firms compete for generic inputs. They also compete in the market for the final product, facing price-sensitive demand, and behave as price takers. Their distinctive competitive approaches are manifested in varying degrees of capability. Firms make idiosyncratic investments to convert generic inputs into the resources needed for production. These customizing investments, and their impact on profits and competition, are a key feature of our model. Capacity levels are fixed in the short term (a single period), but firms respond to profit signals by adjusting their capacity over time. The extension of capacity by more capable firms is the central mechanism driving (Darwinian) selection and shaping market structure over time. The industry displays an evolutionary trajectory, starting from an initial pool of heterogeneous firms with high dispersion of capability (and substantial efficiency profits for the best firms) to a more concentrated structure comprising only the most effective firms from the initial pool. Yet, as we demonstrate through our model, Darwinian mechanisms do not guarantee the full elimination of heterogeneity.

We derive some limited conclusions from our model by analytical methods linking the heterogeneity of firm capability (recently underscored by Bloom and Van Reenen, 2010, and Syverson, 2011) in equilibrium to the extent of customization. However, much richer implications flow from our computational experiments, in which we vary the degree of heterogeneity and the cost of the 'customizing' investments that give resources their firm-specific character. We examine how these parameter changes affect wealth creation as well as firm profits and their distribution. This analysis is supplemented by a robustness check that explores the influence of the pace of dynamic adjustment and also of oligopolistic rather than strictly competitive behavior.

Our results provide a clear view of what might be called the 'strong RBV' case. We show how firms can appropriate the rents generated by idiosyncratic resources that they do not (and cannot) own. Our model demonstrates how this can happen as a result of the creation of idiosyncratic resources through customization, even when input and output markets are fully competitive. Thus, our analysis clarifies RBV logic at a basic level, and fits with its intuitions on competition, firm heterogeneity, and financial performance.[2]

We also propose, however, an important qualification to the RBV emphasis on sustainability. Our model calculations indicate that a firm's sustainable profits may be relatively unimportant compared to the total wealth created in the entire course of its industry's evolution. This result illustrates

[1] In this paper we regard 'sustainable advantage' as necessary but not sufficient for 'sustainable profitability,' with the question of value appropriation intervening between the two. Our analysis treats both issues of advantage, in the sense of net value creation, and profitability, or value appropriation.

[2] We take the term 'heterogeneity' to *imply* differences in financial performance, as graphically discussed in an early paper by Walker (1887). That is, we set aside the subset of heterogeneity cases where 'equifinality' obtains in the sense that observably different firms wind up at the same observed performance level.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

the joint effect of two straightforward considerations. First, the early years of an industry's development are likely to be marked by relatively high price-cost margins, at least for the more competent firms. Later, aggregate industry capacity is aligned with industry demand, resulting in lower prices and margins. Second, sooner beats later simply because of the time value of money. In fact, a dollar accrued 10 years from now is worth less than half of its present-day equivalent, based on the historically plausible interest rate we employ. Thus, we need to assess outcomes from an investment standpoint to understand how wealth creation relates to the dynamic factors that drive industry adjustment paths (and the resulting 'out-of-equilibrium' profits)—especially in the early years of an industry's evolution.

The following section covers the theoretical background of our investigation, and outlines our general proposals. We then describe the motivation for our specific model, its detailed workings, and its partial analytical solutions, before turning to the situations considered in the illustrative calculations. The final three sections present the results, their interpretations, and the broader implications. In the latter discussion, we relate our notion of customization cost to previous work by Lippman and Rumelt (1982).

PROFITS, WEALTH, RESOURCES, AND SUSTAINABILITY: AN RBV FRAMEWORK

Profit and the 'semi-permanent' attachment of resources

The field of strategic management revolves around the analysis of differential profitability and its sustainability over time. But what is 'profit,' and how does it emerge? The question receives multiple answers even within economics, and additional perspectives are offered in the fields of accounting, finance, and strategic management. For the RBV, a cautionary point is that apparently high profits (accounting net income) can derive from sunk investments that are actually failing to yield a normal rate of return. To address this issue, we view financial performance in terms of the potential gains of *investors in the firm* over longer periods, focusing on net present value (NPV) of cash flows, returning to the definition of profit in our discussion.

A related conceptual issue concerns how firms benefit from resources that are 'attached' to them. Ownership of resources is one important form of attachment. Yet some types of resources, and especially human capital, cannot be *owned* by the firm. We thus have to consider how exactly resources are 'semi-permanently attached' to the firm (Wernerfelt, 1984) and yield profits. To tackle this challenge, we distinguish sharply between *generic inputs* and *resources*. As Dierickx and Cool note, 'firms do not employ "generic labor," but people endowed with firm-specific skills and values. "Generic labor" is rented in the market; firm-specific skills, knowledge and values are accumulated through on the job learning and training.' (1989: 1505). We argue that it is this process of accumulation and customization that explains the 'semi-permanent attachment.'

Focusing on this process, we can explain how firms can benefit from resources they cannot own, as well as those whose services they command through ownership. It is the *customization* itself that offers the potential for the generation of rent and profit. Explaining these dynamics is particularly important given that human resources are typically critical in organizational capabilities, and that associated issues of tacit knowledge and social complexity are identified as important sources of 'inimitability' (Hatch and Dyer, 2004).

In our framework, we focus on a key mechanism of attachment of (human) resources that has strong causal links to the RBV themes of firm heterogeneity and idiosyncrasy. We assume that the attachment of employees to a specific firm is determined primarily by the degree to which their skills are more valuable in that firm than they are elsewhere. This happens not only through 'matching' but also through deliberate and costly actions by firms to impart firm-specific skills to the employee, through processes that are generally known as 'specific training.' The key insights on this topic have long been available in the analysis of human capital pioneered by Becker (1964, 3rd edn. 1993). In our model below, we follow Becker's classic analysis closely. A firm makes a costly investment to customize the human resource, that is, to convert the generic resource to its own idiosyncratic type. As Becker explained, a key point is that the benefits of this investment are realizable only within that firm. Thus the conversion does not create outside opportunities and attendant bargaining power that could be used to extract a significant fraction of

Copyright © 2012 John Wiley & Sons, Ltd.

the benefit that the investment produces. On the other hand, by conceding a *trivial* fraction of the rent (modeled as zero), the firm can make its offer superior to the market alternative indefinitely, thus assuring the attachment and avoiding the need to reinvest in a new resource.

A 'holdup problem' is sometimes alleged here: once the investment is sunk, why can't the worker bargain for a significant piece of the return? This possibility is realistic in some cases, but we are assuming a situation where a large number of similar workers are potentially available. In such a case, building a 'reputation for toughness'—that is, 'just say no' to the employee who wants a raise—is a viable strategy for the employer, as game-theoretic analysis of similar situations has long confirmed.[3] If the situation involved a single employee with valuable idiosyncratic traits, or a group engaged in collective bargaining, a different result would likely ensue (see Coff, 1997, 1999, 2010, and Lippman and Rumelt, 2003a, 2003b, Brandenburger and Stuart, 1996 for relevant discussion).[4]

Financial performance in the short and long run

The second conceptual point we wish to address also draws on Dierickx and Cool (1989), who argue that resources underpinning sustainable advantage are typically the result of firms' sustained accumulation of idiosyncratic assets. We extend their analysis to consider explicitly the dynamic calculus inherent in the choices of firms that wish to expand. We take the perspective of a hypothetical long-term *investor* in one of these 'idiosyncratic' firms soon after it is founded.

This approach is consistent with the implicit but logically inescapable assumption in the RBV that it is the process of accumulation over longer periods that underpins 'abnormal profitability.' This assumption alone implies that it is inadequate to discuss financial performance without reference to interest rates and returns on investment, an economic logic that is obviously of central concern to investors. In particular, high (accounting) profits in the present may not represent good performance if they are a consequence of low rates of return on 'sunk' investments that no longer show up in the financial statements, and the returns could actually fall short of the normal level.[5] High current profits are explained by prior investments (Dierickx and Cool, 1989)—but the picture of 'financial performance' is seriously incomplete if rates of return are ignored.

In our framework, the intertemporal perspective is provided by considering the benefits that accrue to investors as reflected by the NPV of the cash flows that the firm produces. If the discount rate used corresponds to returns on alternative investments, NPV is an appropriate *ex post* indicator of whether investing in the firm was a 'good idea.' In an actual, real-time context, share prices fluctuate and differences in transaction timing mean that some investors in a given firm do well, others poorly. Nevertheless, NPV—calculated *ex post* but from a 'time zero' early in the firm's history—remains a good conceptualization of the *realized financial performance of a particular firm*. It is certainly much better than 'profit' measured over a time frame that does not include key investments—like children reporting a 'profitable' afternoon at their lemonade stand by neglecting yesterday's costs of lemons, sugar, pitchers, glasses, lumber, chairs, paper, and crayons.

Heterogeneity and selection under full competition

In the final part of our framework, we consider how heterogeneous firms can compete and coexist even in the absence of product-market imperfections or spatial competition. Of course, a commonsense appraisal endorses the idea that quite different firms coexist and 'compete' in more or less the same market for long periods; consider, for example, Toyota and GM. However, the meaning of 'compete' here is somewhat vague, and is probably closest to what Porter (1980) termed 'rivalry.' Under 'competition' in the narrow economic sense,

[3] The analysis referred to arose primarily from efforts to resolve Selten's 'chain store paradox' and the related problem of explaining cooperation in finitely iterated prisoner's dilemma games. For an overview, see Fudenberg and Levine 1989.

[4] For *owned* resources, customization can also be the source of profits, in addition to the benefit a firm can derive from the resources' appreciation in value (see Winter, 1995; Jacobides and Winter, 2007).

[5] Accounting conventions afford many opportunities for the investments underpinning current income to disappear entirely from view; perhaps the most blatant of these is the expensing of research and development.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

it is hard to explain how heterogeneity can persist in the long run.[6] Lippman and Rumelt (1982) answered that challenge persuasively and rigorously—but also with artfully contrived inattention to the dynamic issues addressed by our paper. We build on their contribution by modeling firms with differing levels of efficiency, but we consider the time-consuming nature of the competitive process that they modeled as occurring instantly. We thus avoid the assumption of decreasing returns to scale as the mechanism that keeps the most efficient firm(s) from dominating the market.

This approach allows us to revisit the ways in which 'competition' acts as an antagonist of 'abnormal return'—as economists have argued at least since Adam Smith. In counterpoint to that tradition, a number of considerations have been identified over the years that might explain persistent profitability in spite of 'competition.' We omit discussion of these here, because they are substantially irrelevant to a core understanding of the RBV.[7] For our framework, we adopt assumptions that are largely consistent with, yet subtly different from, standard arguments that competition eliminates profits. We assume that firms produce under conditions of long-term constant returns to scale, implying that a profitable input-output combination can always be replicated at arbitrary scale.[8]

Consistent with the understanding of 'competition' in the economics textbooks, we further assume that firms calculate on a 'price-taking' basis—that is, they assume that their actions alone do not change market prices. Together, these assumptions are ordinarily taken to imply that no long-run equilibrium can involve positive profits, since such profits would imply an incentive to expand production. In a finite world, such expansion would ultimately drive input prices up, or output prices down, to an extent that would eliminate the incentive represented by the positive profits. In our framework, however, we depart from standard assumptions by assuming that replication is costly, and this underpins our account of persistent profitability.

The plausibility of the price-taking assumption depends on whether the firm has appreciable market power. In a dynamic setting with heterogeneous firms, market shares change and concentration typically increases over time. This may mean that price-taking becomes increasingly implausible over time, in which case it should be abandoned in favor of some account of pricing under monopoly or oligopoly. Yet as our primary concern here is to explicate RBV logic under competitive conditions, we feature results based on price-taking. We also perform a simple robustness check to explore the quantitative impact of emergent oligopolistic behavior. (See Makadok, 2010, for an analysis of market power vs. efficiency differences.).

Our proposal begins with a simple story that is at the heart of a number of evolutionary economics models. In this story, individual firms create production methods through path-dependent learning. Although an underlying fund of public knowledge may provide a common starting point, the contingencies of path-dependent learning in complex situations make different firms' operational methods significantly diverse. Once they have achieved their idiosyncratic solutions to the production problem, firms settle on their methods quasi-permanently and seek to scale them up when this is profitable. A production method is replicable by the firm that created it (constant returns), but is not fully imitable. (The reason for inimitability, whether natural or artificial [e.g., patents], is not consequential for this analysis as long as there is nothing that can be done to change the situation (Koopmans, 1957). In the simple models (e.g., those in Nelson and Winter, 1982), the replication required for expansion is not only feasible but also perfect and costless.

In our key departure from this path, we assume, instead, a positive value for the cost of customizing resources. A successful firm that cannot create additional amounts of customized resources might be able to grow, but not in a way that reproduces the specific advantages that account for its prior success. Its long-run fate might therefore diverge

[6] As Syverson (2011: 326, 333) says in his review, '[Studies] have documented, virtually without exception, enormous and persistent measured productivity differences across producers, even within narrowly defined industries... Given the large differences in productivity within an industry that I discussed above, a natural question is to ask how they could be sustained in equilibrium. The ubiquity of this dispersion suggests there must be some real economic force at work.'

[7] For example, in the context of competitive general equilibrium theorizing in the Arrow-Debreu tradition, positive equilibrium returns arise when a finite number of firms produce under conditions of diminishing returns to scale, throughout. Whether those equilibrium returns should be called 'profits' is a terminological issue.

[8] We follow the terminology and respect the substance of existing understanding of the economics of production, which began with Ricardo's analysis of rent. For a modern treatment see an advanced microeconomics textbook, for example, Mas-Colell, Whinston, and Green (1995: 127–160).

Copyright © 2012 John Wiley & Sons, Ltd.

sharply from early indications. We postulate, further, that these customization costs are in large part *one-time* costs associated with the *initial creation* of new capacity, which often involves setting up new factories, retail outlets, branch offices, and the like. In the course of getting 'up and running,' a wide range of decisions are made about structures, both physical and organizational, that will influence coordination and communication patterns for decades to come. Within the contexts set by such structures, the initial occupants of the various roles seek out stable patterns of productive interaction. This process involves a good deal of costly trial-and-error learning, even when the intention is to replicate an existing establishment (Winter, 2010). Like the larger-scale aspects of structure, these more detailed patterns will tend to persist long after the 'founding generation' has left the scene.

When the costs associated with added capacity are strongly front-loaded by investments made at the time of expansion, an important asymmetry is created between the economics of expansion and the economics of continuing operation or contraction. The asymmetry arises not only from investments in capability creation and early learning as just described but also from investments in durable equipment—which often requires costly customization itself. The more durable it is and the more its resale value is diminished by the customization,[9] the more it contributes to the asymmetry and hence to path dependence and persistent heterogeneity.

Illustrating the model's mechanisms through an example

Consider a hypothetical world of perfect foresight, and two differentially capable firms, taking the perspective of an investor. All investors might agree that the potential path of Firm A will yield better returns than that of Firm B, when both paths are assessed overall at the start. *In the world of real historical time and imperfect foresight, though, this comparison quickly becomes irrelevant (to investors) once some substantial B-specific investments have been sunk.* When B has achieved substantial scale via many such investments, its resilience in adverse market conditions is essentially subsidized by its past. B might not just coexist with A; it might even manage to drive A out. Investors who could see the whole picture would know that B overpaid for its dominance, and that therefore *approaching a new situation by mimicking B's behavior would be a bad idea*. But this insight cannot change the history that produced it.

The asymmetry produced by scale-related sunk costs is a key feature of our formal analysis. Our mechanism fuses two familiar points—the fact that sunk costs are not relevant *per se* to forward-looking calculation, and the fact that increases in production capacity require investment.

Suppose for simplicity that the investment never has to be fully renewed (for reasons suggested above). In that case, when Firm B has achieved a capacity level Q_B, it can remain in the market as long as its forward-looking unit cost c_B is less than the prevailing output price P. Meanwhile, Firm A may have a lower unit cost than B if customization cost is excluded. But to actually expand its capacity—so as to increase output, thus pushing price down and driving B out of the market—it must invest up front in further customization, and this investment obviously need not be attractive, considering that its attractiveness depends on A's customization costs. In such a case, B may survive indefinitely. Whether B is 'making money' in an overall present-value sense depends on the historical circumstances in which it sunk the investment that created capacity Q_B—perhaps price was higher then. Given that $P > c_B$, it is certainly making 'profits' in an ordinary sense. Further, these returns correspond precisely, in the short run, to the concept of economic rent: they exist because the B-customized resource is in strictly inelastic supply (as it is for the corresponding resources of Firm A and other viable competitors).

Customization is what keeps Firm A from expanding; and both the cost, and its discount rate of the future (reflecting its opportunity cost of funds) will determine the extent to which it is deterred from growing and potentially squeezing Firm B out of the market. Thus, the dynamic calculus of customization and assessment of the returns to capacity expansion will determine the

[9] 'Customization' may arise even from the simple fact of transportation and installation costs. These can create a significant economic distinction between a machine shown in a manufacturer's catalog and the purchased machine on-site, in the door, set up, bolted down, connected to complementary devices, and so forth. That happens despite the fact that the machine *per se* remains precisely as advertised in the catalog. Beyond that, the strength of the 'lock in' may be increased because the 'lemons problem' depresses the market value of used complex equipment (Akerlof, 1970).

Copyright © 2012 John Wiley & Sons, Ltd.

profit levels that can be sustained in equilibrium under a competitive structure.

THE MODEL

Overview and antecedents

Our industry model draws on the tradition of evolutionary economics (Dosi *et al*., 1995; Malerba *et al*., 1999; Nelson and Winter, 1982; Winter, 1964, 1971), and extends the analytical framework used by Jacobides (2000, 2008). It is a 'temporary equilibrium' model in which an individual period represents a Marshallian 'short run.' Short-term prices determine profits, and profits in turn drive firm investment behavior. Firms do not have the structural knowledge to anticipate (even in probabilistic terms) the long-run evolution of their industry.[10] The iteration of the single-period equilibrium and change processes generates the dynamic path of the industry. Within this evolutionary setup, our model portrays a process of industry development in which selection plays a key role. The more efficient firms expand and put competitive pressure on the less efficient ones, causing them to shrink. Since firms behave as price-takers, both in the short run and in their investment decisions, the expansion of the more successful firms is not checked by voluntary restraint derived from perception of the market power associated with a large market share. (The role of market power is considered as a modeling extension, described below, and we also explore the robustness of the results to a change in the number of model firms.)

Modeling (idiosyncratic) resources

The most distinctive feature of this model is its treatment of idiosyncratic resources. We assume that there is a single resource, and the level of that resource is the single state variable characterizing a firm in a time period. Seeking to rebalance the picture presented by the many familiar models in which such a central state variable would be an owned asset called 'capital' or 'capacity,' we make the contrasting assumption that the resource is not owned. Rather, it is hired, period by period, at a price that is endogenously determined in the market. The obvious interpretation is a trained employee.[11] There is a second input, which merely 'supports' or 'enables' the focal one. This input represents generic facilities and equipment or services that are not owned by the firm, but are available during every period at a market price. For example, the focal resource might be talented professionals, while the supporting input could be the office space where they work. (Because the supporting input lacks any attachment to the individual firm beyond a single period, it would not be consistent with RBV thinking to call it a 'resource.')

The nature of 'attachment' and its consequences is modeled after the notion of firm-specific training, as noted earlier. A firm makes a costly investment to customize the generic resource; it concedes a trivial fraction of the rent (modeled as zero), making its offer superior to the market alternative indefinitely, and thus assuring the attachment and avoiding the need to repeat the conversion investment with a resource newly acquired from the marketplace. We abstract from the problems of turnover and finite lifetimes, and make the attachment potentially permanent. We simply assume that the customization investment is a one-time cost that remains effective as long as the resource is available to the firm; the substantive point is that some of the outlays attending initial expansion to a given level do not have to be *fully* repeated later. A key implication is that the costs of customization investments are borne only when the firm is growing.

Firm-level attributes: capabilities, firm choices, and competitive interaction

Heterogeneity in production methods is formally represented by a firm-specific multiplicative constant affecting the firm's production function. Thus, firms are (allowed to be) differentially capable, capability being reflected in the efficiency with which a firm turns a set of (customized) resources into outputs. We conceive of these capability differences as arising from independent exploratory search efforts at earlier stages of industry history.

In addition to its resource, a firm needs an amount of the supporting input that is appropriate

[10] This is the key point that distinguishes this type of model from full-rationality (rational expectations) models of industry evolution such as those put forward by Jovanovic (1982) and others.

[11] The logic that follows is equally applicable to customized equipment as to trained employees, except that the aspects relating to the incentives faced by the resource are not relevant for equipment.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

to the historically determined scale of its resource stock. Although the supporting input is hired for a single period, we assume that the firm must commit to a certain level of it before prices and actual production levels are known. Thus, there is a maximum amount of output that a firm can offer the market during a given period, determined by the available stock of the focal resource and the level of the supporting input. What is open to influence by the market is the quantity of the resource that is actually hired, and the resulting output level. We assume that, in setting the amount of the supporting input, the firm anticipates full utilization of the available resource and chooses the support level that would be required for long-run cost minimization given the resource level. This assures that full utilization corresponds to operating at a point on the long-run average cost curve. In terms of the example given, the amount of office space hired for the period is the amount that would be appropriate in this cost-minimization sense given the present size of the firm's available workforce. The utilization of that workforce, but not the office rent, depends on the contracts that come in that period. When the stock of the resource is underutilized, some portion of the excess is no longer available in the following period—as suggested by the fact that experienced workers who are unemployed tend to drift away from the production locales of their declining employers.

Formal analytics: firm decisions and market equilibrium in the short run

The level of the resource is represented by R. Output, Q, is produced according to a production function of the Cobb-Douglas form,

$$Q = a \cdot R^b \cdot K^{1-b} \tag{1}$$

Here, R is the amount of the focal resource and K is the quantity of the supporting input. Firms have different values of the efficiency parameter, a, but the same value of b. For the time being, we analyze the 'representative firm' and suppress the firm indexes, as well as the time index.

At the start of a period, a firm has level R of the resource available, reflecting its past decisions to customize units of the generic input. To support this level of the resource, the firm commits to an appropriate level K of the supporting input. Denoting by W the price of one unit of R and by v the price of one unit of K, the amount of the supporting input that corresponds to cost minimization with that R satisfies

$$W \cdot R \cdot \frac{1}{K \cdot v} = \frac{b}{1-b} \tag{2}$$

This is the familiar condition for cost minimization when the production function is Cobb-Douglas. We assume that the price v is given exogenously, and without loss of generality that its value is 1—the unit of K is 'a dollar's worth.' This yields

$$K = W \cdot R \cdot \frac{1-b}{b} \tag{3}$$

With up to R units of the resource available, and expecting to pay W for units actually employed, the firm will purchase/rent this amount of the supporting input.[12]

With K and R determined, the firm faces, during period t, the problem of choosing an optimal level X of actual utilization of the resource:

$$\begin{gathered} \mathit{Maximize}\ \Pi(t) \\ \mathit{subject\ to}\ 0 \le X \le R(t) \end{gathered} \tag{4}$$

where

$$\begin{aligned} \Pi(t) = {} & P(t) \cdot a \cdot X^b \cdot K(t)^{1-b} \\ & - W(t) \cdot X - 1 \cdot K(t) \end{aligned} \tag{5}$$

The final constant does not, of course, affect the maximizing choice, but it is by definition part of the maximand. We call the optimal X in this problem $R^*(t)$, and the resulting optimized level of operating profit is $\Pi^*(t)$. The optimization yields the short-run input demand curve, relating the firm's resource use to prevailing prices and the extent of input usage within that period:

$$R^*(t) = min \left\{ \left(a \cdot b \cdot \frac{P(t)}{W(t)} \right)^{\frac{1}{1-b}} \cdot K(t), R(t) \right\} \tag{6}$$

[12] Since the K decision is made before W is known, the dynamic specification logically requires an expectation rule for W. We skip over this detail because our actual computations in this paper take W as a given constant.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

On the output side, the firms face the market demand curve

$$Q = D \cdot P^{-\varepsilon_D} \tag{7}$$

Market clearing in the short run involves equating supply and demand in the two markets, one for the resource and one for the final product. At this point, clarity is served by making firm-specificity explicit:

$$S \cdot (W - W_0)^{\varepsilon_S} = \sum_j R_j^*(t) \tag{8a}$$

$$\sum_j a_j \cdot R_j^*(t)^b \cdot K_j(t)^{1-b} = D \cdot P^{-\varepsilon_D} \tag{8b}$$

After substituting in from Equation 6, Equations 8a–b become two equations in the two prices P and W; the solutions are $P(t)$ and $W(t)$.

In sum, the logic of each period starts from the firms' available amounts of their specific resources R, which are a consequence of their past decisions. They plan to complement their resources with appropriate, long-run optimal levels of the supporting input. The short-run market process produces a joint determination of actual levels R^* and the market prices P and W. The two equations above then determine the input and output prices for the period and also individual firm resource levels, hence outputs and profits.

Model dynamics

The linkage between periods is provided by assumptions regarding expectations and decision rules affecting the choice of $R(t+1)$. For every additional unit of the resource that the firm wishes to have available in $t+1$, beyond the amount available in t, the firm incurs a cost in the amount C, a parameter we call 'customization cost.' Such costs are incurred only in the anticipation that the investment will prove profitable, when judged by their required rate of return i. All firms face the same discount rate, the required rate of return; and they do not differ in terms of the efficiency with which they raise capital or engage in customization.

First, firms need to assess whether the expansion will be profitable. Although numerous historical examples testify to the point that volatile expectations can be a source of instability in early-stage industry evolution, we put this set of issues aside. Our assumptions about expectations are chosen to yield orderly, stable adjustment paths, are consistent with our long-run equilibrium analysis, and have the virtue of simplicity.

We assume that firms assess the trend in variable profit per unit of the resource, extrapolate that trend for a single period, and assume constancy after that. This assessment tends to be optimistic with respect to the long run, since profitability is declining. By contrast, the one-period extrapolation tends to be pessimistic because the rate of the decline is itself diminishing. As time passes, the pace of change drops, and both of these biases decline in magnitude until, at the threshold of equilibrium, the estimates are essentially accurate. The threshold of equilibrium is precisely where it matters that they be accurate, from the viewpoint of industry-level outcomes, since those expectations shape the last investment decisions as the system closes in on equilibrium.

The one-period extrapolation is derived from an 'adaptive expectations' mechanism that assesses the trend in 'variable profit per unit of the resource,' denoted by $V_j(t)$—that is, operating profit with the fixed cost of the supporting input added back in, divided by $R^*(t)$.

$$\begin{aligned} V_j(t) &= [\Pi_j^*(t) + 1 \cdot K_j(t)]/R_j^*(t) \\ &= P(t) \cdot a_j \cdot [R_j^*(t)/K_j(t)]^{b-1} - W(t) \end{aligned} \tag{9}$$

This quantity is always nonnegative by virtue of the simultaneity of profit maximization and market clearing. Our experiments show that it provides a good basis for the expectations governing the investment decisions. In effect, the firm calculates the expected change of variable profit per unit as a weighted average of recent percentage changes, with geometrically declining weights. The details of this are more cumbersome than the concept, and are consigned to an appendix.

To establish the profitability test for an incremental investment in R_j, which we denote by $T_j(t)$, we take into account variable profit, but also the fixed costs per unit, the customization cost, and time discounting, as well as the expectation of percentage changes in unit profit, which we term z_j.[13] Firms extrapolate the recent changes one period

[13] Although firms do not borrow in our model, the interest rate matters because it describes the returns available on alternative investment opportunities available to the firm or its investors.

Copyright © 2012 John Wiley & Sons, Ltd.

forward to estimate changes in future profitability. Specifically, the investment profitability test given the data of time t is based on the sign of the expression

$$T_j(t) = \frac{(1+z_j)}{i} \cdot V_j(t) - \frac{(1-b)\cdot W(t)}{i \cdot b} - C \tag{10}$$

That is, $T_j(t)$ is the excess of the present value of the future rents per unit earned per unit of the resource over the cost of the customizing investment. Note that $W(t)$ enters this expression explicitly by virtue of its role in determining the level of the supporting input (Equation 3), and we assume that the realized period t value is extrapolated for this purpose.[14] A positive value of $T_j(t)$ indicates an attractive investment in present-value terms, given our specific assumption as to how firms evaluate the returns to be expected in a future that is unknown at the time the investment is made.

When the test $T_j(t)$ is passed, the next question is how many new units of the resource should be customized. We assume that this question is answered by reference to the amount of cash available to cover the customization investment; the level of investment is determined by the ratio of available cash to customization cost per unit of resource. As the customization cost is considered to be incurred at the start of a period, contemporaneously with the commitment to K for that period, the net cash flow of period t is

$$\Pi_j^n(t) = \Pi_j^*(t) - C \cdot max(0, R_j(t) - R_j(t-1)) \tag{11}$$

The investment in customization is irreversible, so no cash is released when the R level declines.

For reasons discussed subsequently, we assume that only a fraction of this total is applied to customization. To determine the actual number of resource units customized, we divide the dollar investment by the customization cost per unit, C.

$$R_j(t+1) = R_j(t) + \begin{cases} max\left(0, \theta \cdot \frac{\Pi_j^n(t)}{C}\right), & T_j(t) > 0 \\ 0, & T_j(t) \le 0, \\ & R_j^*(t) = R_j(t) \end{cases} \tag{12}$$

[14] Here again, an extrapolation scheme for W is in principle needed. However, in our computations we take W to be a constant, and choose here to avoid the additional complexity introduced by such an extrapolation.

$0 \le \theta \le 1$ is a parameter controlling the fraction of the cash flow available for investment.

On the other hand, if the investment test is not passed, and if also a firm did not use up all of its customized resource, that is, $R_j^*(t) < R_j(t)$, it is assumed that a fraction λ of the excess resource availability is permanently lost to the firm in the next period. (This might reflect, for example, changing expectations of resource owners about the prospects of further work with the particular firm, or the dwindling of commitments to its particular locale.) In that case,

$$R_j(t+1) = R_j(t) - \lambda \cdot (R_j(t) - R_j^*(t)) \tag{13}$$

Table 1 provides a glossary of the variables used, and an online Appendix provides a Figure.

ANALYTIC RESULTS

Before turning to computational experiments with the model, we examine special cases in which the character of the long-run equilibrium can be explored analytically. Generally speaking, we have exact analytical results only for the cases in which all firms are identical. However, when customization cost C is zero, we know that the only firms that survive in the long run are all those that have the highest efficiency level, regardless of the initial heterogeneity, for selection always entails homogeneity in the long run. By contrast, when C is positive, firms with differing efficiencies can coexist in the long run, as we will demonstrate. The mix of firms that coexist in the long run is, however, a path-dependent, historical phenomenon. Our analysis can illuminate the extent of the potential indeterminacy in the mix of survivors, but no general conclusions beyond that are possible.

Equilibrium with identical firms

The analytic results we can derive relate to firm behavior as the sector converges to long-run equilibrium. We consider the case of homogeneous firms, and begin the analytical exercise by supposing hypothetically (contrary to the dynamic model) that the total availability of the resource (or rather, generic input) to the industry is fixed, and that the level of the supporting input is appropriately matched to it. On that assumption, the return to the resource is a Ricardian rent in the strictest

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

Table 1. Model glossary

Parameters and single-period variables		
$Q_j(t)$	Output	Endogenously derived
a_j	Production efficiency coefficient (capability)	Firm-specific, varied within and between scenarios
R_j	Level of resource useable by firm j in period t	State variable (short-run); choice variable (dynamically)
R_j^*	Level of resource used by firm j in period t	Choice variable
b	Cobb-Douglas production elasticity for R	Parameter
K_j	Supporting resource	Choice Variable
$W(t)$	Price of resource R	Endogenously derived (constant in calculations here)
v	Price of resource K	Parameter
$\Pi_j(t)$	Operating profit	Endogenously derived
$\Pi_j^*(t)$	Optimized level of operating profit	Endogenously derived
D	Demand constant	Parameter
$P(t)$	Price of final good	Endogenously derived
ε_D	Elasticity of demand	Parameter
W_0	Supply constant for resource	Parameter
ε_s	Elasticity of supply	Parameter
Dynamic update parameters and variables		
C	Customization cost	Parameter, varied between "setups" (sets of scenarios)
$V_j(t)$	Variable profit per unit of resource	Calculated (by firms) derived value
i	Interest/required rate of return	Parameter
$H(t)$	Herfindahl index at t	Parameter derived from period t output levels
$z_j(t)$	Expectations for changes in per unit profit	Calculated (by firms) derived value - see Appendix
d	Weight on recent observations to derive $z_j(t)$	Parameter
$T_j(t)$	Investment test, net present value	Calculated (by firms) derived value (>0 means invest)
$\Pi_j^n(t)$	Net cash flow	Derived value - realized profit minus investment costs
θ	Fraction of net (free) cash flow investable in C	Parameter
λ	Fraction of unused resource lost to the firm	Parameter

sense. What would be the equilibrium level of that rent, per unit of the resource, and how would that rent vary with the assumed total availability? As may be obvious, the smaller the postulated level of resource availability, the smaller the level of production and the higher the final good price—and the higher the rent per unit assigned to the resource. Now compare that rent with what the actual supply curve of the resource would yield with the supply price at the assumed resource level. The difference between the two is the true rent per unit that accrues to the firms when resource use is at the assumed level.

Following this intuitive logic, the rent schedule construction is straightforward. Given a value of R, we consult the supply curve of the resource to identify the corresponding W. That value, in conjunction with the exogenously given price of the supporting input, determines optimal input proportions between R and K. With those proportions known, the production function tells us the output quantity Q that corresponds to the initial R. Consulting the demand function, or rather its inverse, we find the output price P. The value of output and its cost (excluding rent) are now determined, the difference between the two is the total rent, and that quantity divided by R is the unit rent value for that R.

Following this path analytically, we first consider the supply-price curve $s(R)$

$$W = s(R) \tag{14}$$

Invoking Equations 3 and 1 above, we find that long-run quantity is related to R by

$$Q = A(R) \cdot R \tag{15a}$$

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

where

$$A(R) = a \cdot \left(s(R) \cdot \frac{1-b}{b} \right)^{1-b} \tag{15b}$$

Note that a relationship of the form of Equation 15a would prevail for any production function displaying constant returns, but the particular Cobb-Douglas specification is obviously involved in Equation 15b.

Let the demand-price function be $h(Q)$:

$$P = h(Q) \tag{16}$$

Then the desired result, the rent per unit (W_V), is obtained as

$$\begin{aligned} W_V &= \left(Q \cdot h(Q) - W \cdot R - 1 \cdot W \cdot R \cdot \frac{1-b}{b} \right) / R \\ &= \frac{Q \cdot h(Q)}{R} - W - W \cdot \frac{1-b}{b} \end{aligned} \tag{17a}$$

Here, Equation 3 has been invoked again, while 15a and 15b are implicit. This gives us the rent value corresponding to a given R.

If there were *nothing* to prevent firms from expanding, this expression would be zero in the ultimate equilibrium. In this case, we could substitute zero at the left-hand side, and thus define the equilibrium relationship between Q and R. In general, however, there is a cost to expanding; and this cost is a function of the customization cost C and of the interest rate i. The prospect of a perpetual rent stream in excess of $i \cdot C$ is enough to motivate the investment C in creating another unit of the resource. So, barring any other influences on firm growth (such as deliberate output restraint), we can match the benefit from the use of resource to the cost of creating it via customization. Invoking Equation 15a, R divides out, and Equation 17a can be rewritten as

$$W_V = A(R) \cdot h(A(R) \cdot R) - s(R)/b \tag{17b}$$

For given values of interest rate i and customization cost C, the equation

$$A(R) \cdot h(A(R) \cdot R) - s(R)/b = i \cdot C \tag{18}$$

thus determines the equilibrium value of R (and hence all other variables) in the qualified sense described above, provided all firms are identical—specifically, this expresses the relevant marginal condition for capacity *expansion*. Note the particularly simple form that appears if $s(R)$ is a constant, with the result that $A(R)$ is as well.[15]

Coexistence of differentially capable firms in equilibrium

While the rent schedule provides an analytical solution for identical firms, it cannot tell us where the system will equilibrate when firms do not have homogeneous capabilities. However, it can provide a different and quite valuable insight, by helping us understand when differentially efficient firms *might* coexist in equilibrium. To do so, we consider a situation with any two firms of unequal capability (where $a > a'$) and examine the conditions under which firm a would not want to expand, and a' would not want to shrink, if it is in the sector already and has sunk the customization cost investment for its current scale. The analysis described above still holds as it relates to the most efficient firm, with capability a. What we know for sure is that this firm cannot be in equilibrium if its prevailing rent rate exceeds the value that would give it an incentive to expand its capacity, investing in customizing the resource. We also know that none of the less efficient firms will have an incentive to invest when the most efficient firm is getting close to zero incentive. The open question is when a less efficient firm has an incentive to reduce output, which would have the effect of increasing the profitability of the most efficient firm and reinspiring it to grow. In effect, there is a locus in the (W, P) space where equilibrium might occur, if it is a case where the asymptotic R value of the most efficient firm is approached (gradually) from below. This locus can be obtained by modifying Equation 18, replacing $h()$ by P and $s()$ by W, and taking the a value corresponding to the most

[15] Standard comparative statics techniques can be employed on Equation 18—but subject to an important qualification. If an equilibrium achieved by the dynamic system is disrupted by a decrease in C, then Equation 18 will hold in the new equilibrium at a higher R. The standard techniques will therefore be useful in characterizing it. But suppose the disruption is from an *increase* in C. Then, since the customizing investments are already sunk, R does not decrease as Equation 18 would imply. In that case, equilibrium is established at a higher level of activity than would have occurred if the higher C value had been prevailing all along.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

efficient firm:

$$a \cdot \left(W \cdot \frac{1-b}{b}\right)^{1-b} \cdot P - \frac{W}{b} = i \cdot C \qquad (19)$$

A less efficient firm that has already sunk an investment in capability will not have the incentive to contract as long as it can simply cover its continuing costs, perhaps with room to spare. This requires that its efficiency parameter, a', satisfies

$$a' \cdot \left(W \cdot \frac{1-b}{b}\right)^{1-b} \cdot P - \frac{W}{b} \geq 0 \qquad (20)$$

Moving W/b to the right side in Equations 19 and 20 and dividing yields the 'coexistence condition' for a', namely

$$\frac{a'}{a} > \frac{W}{W + i \cdot b \cdot C} \qquad (21)$$

We can further simplify this expression. Given that we have expressed C as an absolute cost value, it is easy to scale it as a fraction of the prevailing price for the resource W, and set c as the 'cost to invest in a unit of resource as a fraction of the cost of the resource itself' (which is a more intuitive measure). Since $C = W \cdot c$, W then conveniently drops out from this fraction,[16] yielding

$$\frac{a'}{a} > \frac{1}{1 + i \cdot b \cdot c} \qquad (22)$$

This equation shows that given the efficiency level of the most efficient firm, survival prospects for less efficient firms tend to be an increasing function of the customization cost, the interest rate, and the resource elasticity of output. Thus, it shows that firms of somewhat inferior skills *might well persist, even in long-run equilibrium* (as long as they happened to appear early on, made the appropriate sunk investments, and were not too inefficient). This provides an evolutionary explanation for heterogeneity in capabilities even in long-run 'selection equilibrium.' It also shows that the extent of dispersion that can be sustained in equilibrium is a function of the extent to which it is difficult to convert 'generic inputs' into 'idiosyncratic resources,' as well as of the interest rate, and of the relative role of these 'idiosyncratic' resources in total output. This provides a new twist to the intuition that in 'commoditized' sectors, differentially capable firms cannot coexist in equilibrium and no profit can be had: the explanation here suggests that the lack of *input customization,* as opposed to lack of product differentiation, can account for competitive intensity.

Note that this coexistence condition does not involve the relative sizes of the differentially capable firms. Thus, the equilibrium logic does not offer a reason why the most efficient firm should not turn out to be the smallest surviving firm in the industry. While there are good reasons to think that result unlikely, they are historical, path-dependent reasons, not equilibrium reasons.[17] In a real-world industry evolution setting, one might well imagine that a very late entrant could have the best methods—but large incumbents have sunk their customization investments, and they are not going to pull back just out of respect for the young winner of the 'efficiency Olympics.'

While the foregoing provides an interesting link between RBV theorizing and equilibrium analysis, it does not give anything near a full picture of how C, or capability dispersion, affects not only the *coexistence* of firms but also profits, resource prices, and quantities. More important, this calculus leaves the significant issues of paths of adjustment and wealth creation, and the industry dynamics we discussed earlier, entirely in the dark. To consider them, we now shift from closed-form solution to numerical investigation, creating an appropriate experimental design to explore the implications of the model, and then return to examine the theoretical implications.

[16] This result is somewhat oversimplified. While C has been treated as a fixed parameter, W is endogenous in our model. Thus, the ratio c cannot in general be treated as a constant. Obviously, Equation 22 is exact when the resource supply is infinitely elastic, so W is a given constant. And it is a good approximation when $s(R)$ is nearly constant in the relevant region. Our computations assume it is exactly constant.

[17] Note that in keeping with microeconometric productivity studies (e.g., Chew, Bresnahan, and Clark, 1990), we do not assume that scale itself confers any absolute advantage in terms of either production or investment/growth. Our analysis is thus in the spirit of Demsetz (1973), where efficiency begets scale rather than scale begetting efficiency (or profitability/concentration). Our formulation is consistent with long-run constant returns to scale as well as with short-run decreasing returns to scale (as observed in practice and expected in theory).

Copyright © 2012 John Wiley & Sons, Ltd.

EXPERIMENTAL DESIGN

Setup and parameterization

For the calculation of illustrative adjustment paths our computational techniques parallel those used in (Jacobides, 2000; 2008). A common background is posited for these calculations, which we interpret as a highly stylized representation of the processes of industry evolution. For guidance in the choice of parameter values characterizing this background, we draw on the extensive literature that has studied the characteristic historical patterns that mark these processes (Gort and Klepper, 1982; Klepper, 1997; Klepper and Graddy, 1990; Utterback and Abernathy, 1975). These studies have focused on technological change, entry and exit, and the evolution of industrial structure. What we draw from them is, first, a general sense of relevant time frames. Major industries show significant evolutionary change over periods of decades. Dramatic episodes lasting a decade or less are common in smaller sectors, and within the lengthier trajectories of large industrial sectors. Consider, for example, the expansion of the U.S. telephone industry from about 100 firms in 1894, when the Bell patents expired, to over 9,000 organizations in 1902 (Barnett, 1995); or consider, more recently, the rapid evolution of the disk drive industry from 1975 to 1989 (Christensen, Suarez, and Utterback, 1998). Second, this literature documents the powerful role of firm heterogeneity in these stories, and recent contributions provide plausible interpretations of the origins of that heterogeneity in terms of firm-level learning and the historical descent of capabilities (Helfat and Lieberman, 2002; Levinthal, 1997).[18]

These facts provide some empirical basis for the background conditions in our experiments. Regarding profitability and wealth, we have less quantitative research. It has long been clear, however, that great fortunes are founded in the early years of great industries. We also know that outstanding stock market performance is strongly associated with 'newness,' and over an industry's history is often closely related to the rate of entry (Foster and Kaplan, 2001). The assumptions we now introduce represent our effort to characterize, in a highly stylized context, a situation consistent with the broad implications of the historical record. More systematic use of that record is a significant item on the agenda for future work.

In all cases, we posit an initial state representing an early stage in the process of industry evolution. Firms have already developed their production methods, but industry resource use is a modest fraction of the equilibrium values. The initial resource levels are constant across settings and scenarios. Initial resource use is the same across cases, and in the base case is roughly a third of its equilibrium value. Increasing C makes equilibrium output smaller and in this sense implies that the initial condition is closer to equilibrium. We compute the model outcomes on the basis of quarterly periods, which avoids computational stability issues that can arise when the system is changing quickly. We extend the computation for 250 quarterly periods to allow the slower-moving equilibration processes the time to produce something close to their asymptotic results. (For example, in the scenarios with low but positive heterogeneity, it takes a long time for the modest efficiency differences to express themselves in firm-size differences.) In most cases, the asymptotic equilibrium is closely approximated by period 100; in a few cases it takes until about period 140. By period 250, all our scenarios are essentially at equilibrium.

The profitability conditions of the final period are then extrapolated to infinity for the purposes of present-value calculations, providing assurance that there are no distortions arising from arbitrary truncation of the cash flow series. At reasonable discount rates, however, this adjustment is of negligible consequence. This reflects an important qualification to the strategic significance of 'sustainable competitive advantage,' if we take the shareholders' perspective and consider an industry evolution setting. For our discount rate we use a quarterly rate of 1.8245 percent, which compounds to an annual rate of 7.5 percent. This reasonably approximates the long-run real rate of return on U.S. equities, which seems an appropriate standard for our purposes, given the opportunity-cost interpretation of the interest rate in our model. At that rate, the time zero present value of \$100 at period 101 (start of year 26) is \$16.10.

[18] In the strategic management literature, the significance of firm heterogeneity in various characteristics and in performance has long been recognized (Nelson 1994; Rumelt, Schendel, and Teece 1994). Regarding productivity differences in particular—the locus of heterogeneity in our computations—a substantial body of large-sample evidence has been developed in economics. See, for example, Bloom and Van Reenen 2007; 2010, and references cited therein.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

Experimental design: customization cost settings, heterogeneity scenarios

We vary the experimental conditions in three different ways. There are three values for customization cost C, of which the lowest is effectively zero and the largest produces an equilibrium rent of approximately the same magnitude as the price of the generic resource. Our 'base case' is characterized by the intermediate C value. To develop intuition regarding the plausibility of a C value, it is useful to consider the relationship of this one-time cost to the flow cost of the generic input. In our base case, the one-time cost is of the same order of magnitude (somewhat more than) the cost of a year's services of the generic input. More simply: training costs are comparable to a year's wages.

For each of the three C values, there are 11 heterogeneity 'scenarios.' These range from no heterogeneity to a case where the efficiency ratio between the best and worst firms is three to one. The specific formulation is that there are 11 firms (or types) spread uniformly across an interval centered on $a=1$, and ranging from $1-x/20$ to $1+x/20$, for $x=0,1,2,\ldots,10$.[19] In the online appendix, we present a robustness check confirming that 11 firms provides an adequately close approximation to 'many.'

For this first set of experiments with the model, we chose some key parameters with an eye to the interpretability of results, and also to assure computational stability. The elasticity of demand is set to 1, which means that revenue is a constant and thus not an active factor in the determination of the industry profitability picture. The supply curve of the generic version of the resource is chosen to be perfectly elastic, that is, W is constant.[20]

Also highly relevant to the model results is our specific assumption concerning investment finance. When capability increase passes the profitability test (Equation 11), it is constrained to be a fraction of the level of net profit, that is, the current operating profit less the previous amount of customization expenses (Equation 12). This simple formulation is intended to capture the reality of a feedback from profitability to growth as a factor in industry dynamics, not to model the investment decision realistically. Firms might obviously spend beyond the limits of internal finance by borrowing. On the other hand, given reliance on internal financing, they would spend less if capability increase involved investment costs other than those of customizing the resource—for example, if the ('supporting') plant and equipment had to be owned rather than rented, as is realistic and commonly assumed, but not central to our story. Or there might be other adjustment costs of some sort, or perhaps a dividend payout. The latter sorts of considerations lie behind our choice of 0.5 as the fraction of net profit invested. The important thing, again, is to acknowledge the dynamic linkage between profitability and investment, and to conform roughly to the observed time scale of events.[21]

Extension: considering imperfect competition

As emphasized in the Introduction, our paper focuses primarily on *efficiency profits,* as opposed to profits related to output constraint and price manipulation. As a robustness check, however, we explored one alternative version assuming that firms take a more strategic approach to decisions about capacity expansion, and avoid expanding in a way that would destroy their margins. Specifically, we postulated that all firms in the industry attend to the industry structure and apply higher hurdle rates to investments as concentration increases. As a result, a higher concentration

[19] In most of our heterogeneity scenarios we are, if anything, being somewhat conservative relative to the empirical studies of productivity differences. Syverson (2004), for instance finds that within four-digit Standard Industrial Classification coded industries in the U.S. manufacturing sector, on average, the plant at the 90th percentile of the productivity distribution makes almost twice as much output with the same measured inputs as the 10th percentile plant, and Syverson (2011) notes the existence of more extreme findings.

[20] The Cobb-Douglas exponent on the resource is 0.8, reflecting our intention to give the resource the dominant role in the picture, relative to the generic 'supporting' input. At low output prices, the Cobb-Douglas form implies that very high resource productivity is achieved at very low utilization rates of the supporting input. The result is that a firm with positive capacity—and all firms are always in that condition given that at most a fraction of the resource disappears each period—always chooses a positive output level. Thus, inefficient firms 'fade away' from the marketplace; they do not disappear abruptly. This is convenient in an exploratory exercise that does not allow for continuing entry and thus has no mechanism to balance the effects of permanent exit on the industry dynamics.

[21] The effect of the financing assumption is particularly powerful when customization cost is negligible, as in that condition the only limit on firm growth, the financing constraint, is barely operative and strong disequilibrium disappears quickly. Since it is fundamental to our approach that customization cost is *not* zero in interesting and relevant cases, the distorting effect when C is nearly zero does not constitute a flaw in our results.

Copyright © 2012 John Wiley & Sons, Ltd.

Table 2. Parameter values and robustness checks

	Parameter	Parameter Value or Range	Robustness Checks	Observations
a	Production coefficient (capability)	$a = (1 - x/20$ to $1 + x/20)/4$, or $a = (1 - x/20$ to $1 + x/20)$ on annual basis, $x = 0, 1, 2, \ldots, 10$		Breadth of a dispersion produces expected result
b	Cobb-Douglas prod. elasticity for R	0.8	0.5 to 0.95	No significant changes to qualitative results
D	Demand constant	262.5, or 1050 on annual basis	N/A	N/A
W_o	Supply constant for resource	1.75 quarterly (7 on annual basis)	N/A	N/A
ε_D	Elasticity of demand	1	0.5 to 2.5	Interesting results, unrelated to reported ones
ε_S	Elasticity of supply	infinity	0.5, 1 and 2.5	Interesting results, unrelated to reported ones
C	Customization cost	1, 10, 100	1 to 200	Dynamic instabilities for $C = 1$; small values onwards work well
i	Interest/required ROR	0.018245, or 0.075 on annual basis	0.0375 to 0.25	No significant changes to qualitative results; impact as expected
i'	Interest/required ROR for oligopoly	$(1 + 0.075 \cdot (1 + 2H))^{0.25} - 1$ or $0.075 \cdot (1 + 2H)$ on annual basis	$0.075 \cdot (1 + H)$	No significant changes to qualitative results; impact as expected
d	Weight on observations to derive $z_j(t)$	0.5	0 to 1	No significant changes to qualitative results; impact as expected
θ	Fraction of net cash flow invested in C	0.5	0.2 to 1	No major impact; higher θ leads to rapid movement toward equilibrium with erratic overshooting
λ	Fraction of unused resource lost	0.5	0.5 to 1	No significant changes to qualitative results
N	No of firms (firm types) per run	11	6 to 21	No significant changes to qualitative results
v	Price of resource K	0.25, or 1 on annual basis	N/A	K chosen to have unit price, no robustness checks necessary

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

leads to less capacity and thus to greater output constraint. This formulation has the advantage of capturing the main consequence of imperfect competition, at least for some part of the wide range of behaviorally plausible conditions. Specifically, in our 'oligopolistic behavior' setup we set the required rate of return that a firm uses in its expansion calculus to be $i' = i \cdot (1 + 2 \cdot H)$, where H is the prevailing value of the Herfindahl index. So, given that we set i to 7.5 percent on an annual basis, this means that for a heavily concentrated industry with a Herfindahl of 0.5, the expected return for *all* firms competing would be 15 percent—double the return in a totally competitive sector, and a monopolist (with $H = 1$) would expect a return of 22.5 percent before they expand. We reran all of our scenarios in this 'oligopolistic expansion' mode, and we report the differences for some key scenarios that illustrate the intuitions derived from these results. Table 2 provides a summary of parameters, permutations, and robustness checks undertaken.

COMPUTATIONAL RESULTS

We rely primarily on graphical methods to present our results, since these convey a strong sense of the qualitative features that are the main interest. Since our model is deterministic, there is no need for multiple runs to provide a clear picture of the model's logic.[22] We rely particularly on a 3D display with the value for a particular variable shown over a plane; the two dimensions of the plane are the time period and the heterogeneity scenario (0 = no heterogeneity, 10 = maximum heterogeneity). While we run the model for 250 quarterly periods, almost all the action happens between periods 1 and 100, except in the settings with the highest customization cost. Our graphs provide the results of periods 1 to 100, and then periods 246 to 250—to show the extent of convergence and allow us to consider the results 'further out' in the future, without overly compressing the periods 'where the action is.' The results displayed in this manner include some key industry aggregates and descriptive statistics, as well as variables that actually enter the model logic. We first describe the results with our basic competitive model and then consider the 'oligopolistic expansion' extension noted above, as well as other robustness checks.

Industry evolution: profits, prices, quantities, dispersion

The familiar logics of industry evolution are visible in common features of the computed industry trajectories; these form a qualitative picture that is the common background for the experimental effects that are of central interest. Figures 1a–1e show the trajectories, in the base case,[23] for output price, total operating profit, the average capability in use, the variance of capability, and total resource use. The computations for average capability in use and variance use input share weights.

In the first two panels, the most obvious feature is the decline in price and profitability associated with the approach to equilibrium from an initial condition at far below equilibrium capacity. Growth decelerates as profits fall, and the transition to a near-equilibrium condition is less abrupt when firms are different and thus stop growing at different times. Note that the operating profit total is clearly approaching an asymptotic value that is above zero; this is the rent consequence of the customization cost and receives more extensive comment below. In Panel 2c, the effects of heterogeneity and selection are portrayed. The average (share-weighted) resource productivity increases with heterogeneity, and to an increasing extent over time, as selection has time to enhance the market shares of the most efficient firms. Panel 2d shows the decline of the share-weighted variance of productivity as share becomes increasingly concentrated in the most productive firms. Note, however, that the variance does not appear to be approaching zero. This, again, is a consequence

[22] There are minor, quasi-random fluctuations visible in the model output. These are a reflection of its nonlinear dynamics and particularly of the discontinuity in investment behavior created by the profitability test. These fluctuations are not large enough to affect our conclusions substantially. The closest thing to an exception is the small effect seen in the zero heterogeneity condition, where the discreteness of the final expansion decision, and its sensitivity to timing, produces results barely visible in our charts. This happens because all firms are doing precisely the same thing.

[23] Our baseline has a modest customization cost C of 10, which implies a quarterly interest cost of 0.18245. The price of a quarter's service of the input is 1.75, so customization increases input cost by 10.4 percent on a continuing basis. For a customization cost of 100, customization costs about as much as the market price of the generic input. We also ran our results with a very low customization cost ($C = 1$) and have the full set of results available in the online appendix, along with additional graphical results.

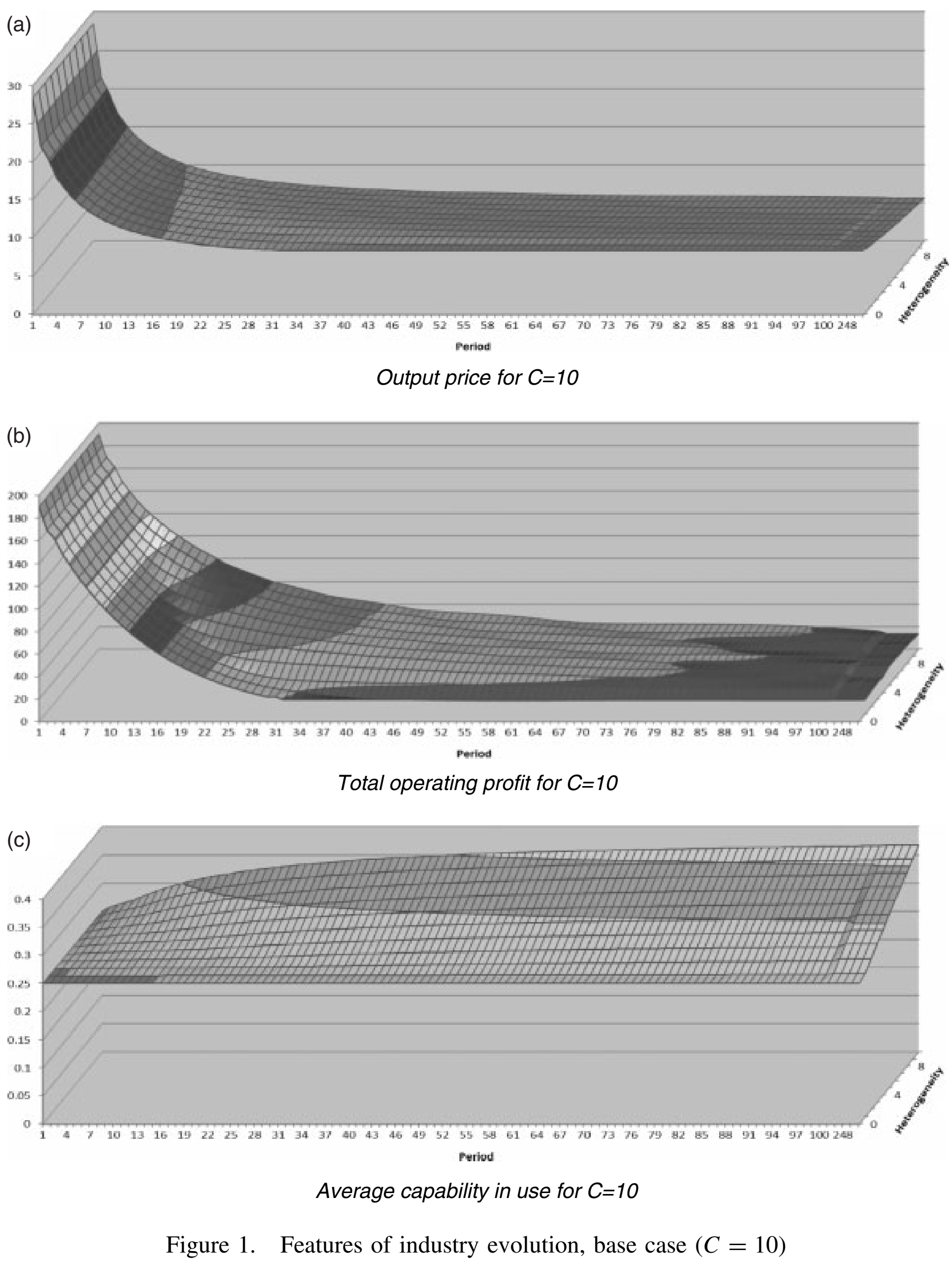
Copyright © 2012 John Wiley & Sons, Ltd.

of the long-term coexistence of firms of different efficiency, a possibility discussed above in abstract terms (see Equation 21 and related discussion).

Total resource use rises steeply in the initial phase, but then stabilizes as total output growth slows. While output grows and price declines continue as a result of relatively slow selection effects, resource use is effectively constant. This is because with unitary demand elasticity and price hanging close to average cost, the decline of output price exactly offsets the effect of rising productivity on input demand. Figure 1e shows the amount of total resource use, and confirms this. The impact of the elasticity of demand on profits and resource prices is an important topic for subsequent research.

Sustainable advantage and persistent profits: causes and consequences

The profitability test for capability investment ultimately checks the growth of capacity and output, and produces an industry equilibrium in which



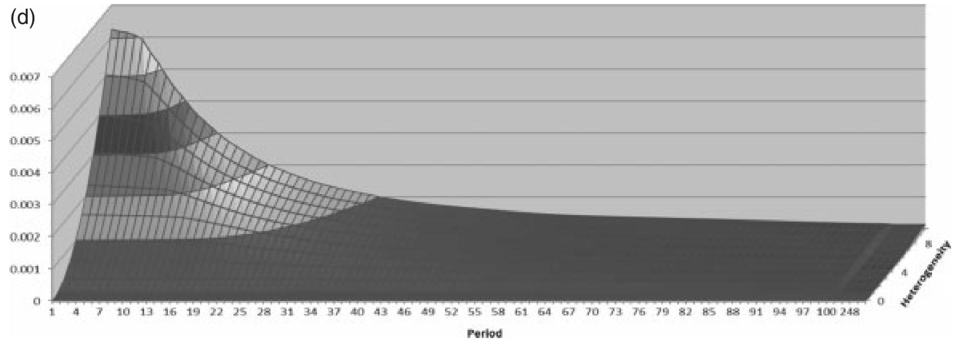
Output price for C=10

Total operating profit for C=10

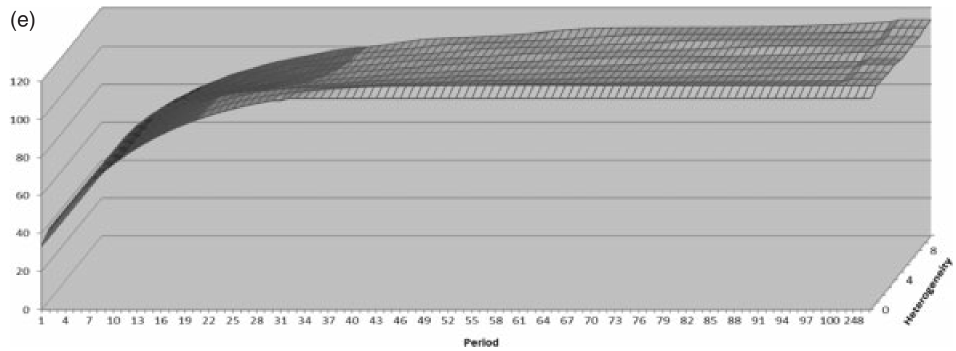
Average capability in use for C=10

Figure 1. Features of industry evolution, base case ($C = 10$)

Copyright © 2012 John Wiley & Sons, Ltd.



Share-weighted variance of capability for C=10



Total resource use for C=10

Figure 1. (*Continued*)

price exceeds the continuing production costs for the most efficient firms. If we compare our $C = 10$ baseline case to the $C = 100$ case, the final value of the output price is increased by over 68 percent in all heterogeneity conditions. Equilibrium output is correspondingly reduced; nevertheless, equilibrium is approached much less rapidly because of the slowing of adjustment associated with the higher customization cost. At equilibrium, the operating profit (or rent) is in the order of 37–45 percent of revenue, as against five–seven percent in the base case, with variation across scenarios. Operating profits are high, but recall that in making investment decisions, firms compare this rent with the value of $C \cdot i$ per unit of the resource. At the margin of the capability investment, returns are close to 'normal' in all cases.

Coexistence of firms in equilibrium, and dynamic paths

Also as analyzed above, positive customization costs promote the long-term coexistence of firms of different efficiencies. When such costs are negligible ($C = 1$), the familiar selection logic is fully in effect. The output share of the most efficient firm in the highest heterogeneity scenario is within rounding error of 100 percent by period 91. In the corresponding scenario for the base case, the top *two* firms are headed for long-term survival, with output shares of 72 percent and 28 percent. And in the high customization cost case, it is *seven* firms with shares from 38 percent down to four percent.

The important role of customization costs in affecting long-term industry dynamics can also be confirmed by looking at the Herfindahl concentration indexes in the case $C = 10$, compared with the case $C = 100$. In the former, the two most effective firms expand, taking the entire market. In the latter, concentration increases very gradually, as even the most efficient firms are checked in their expansion. From about period 200, the index is essentially level at a value slightly below 0.23, implying a 'numbers equivalent' of fewer than five firms. Thus, the asymptotic situation is one of relatively loose oligopoly—but under our assumption

Copyright © 2012 John Wiley & Sons, Ltd.

Table 3. Profit level and sustainability, under competition and oligopoly

(Scenario 5)	$C = 10$	$C = 100$	$C = 10$, oligopoly	$C = 100$, oligopoly	$C = 10$, $\theta = 0.2$ (slow expansion)	$C = 10$, $\theta = 1$ (rapid expansion)
Total profit NPV, seen from period 1	2855	7796	2947	7973	4099	2178
Sustainable profit NPV in period 250 as percent of total NPV	0.08%	0.17%	0.11%	0.20%	0.05%	0.11%
Sustainable profits (from period 1) over total NPV	28.9%	65.9%	42.1%	74.9%	18.8%	41.2%

of price-taking, it is one in which firms exert no market power.

Wealth creation: NPV of cash flows

We now turn to the important observation that, from the wealth-creation viewpoint, sustained profitability plays a minor role in creating the NPV shown by the trajectory as a whole. For instance, consider the contribution of equilibrium profits (considered in all scenarios at period 250, which is a close approximation to the asymptotic result), as a proportion of the full NPV. Table 3 provides this fraction for $C = 10$ and $C = 100$ in its second row, where we extrapolate the equilibrium profits to infinity for the purposes of the NPV calculation. We provide the figures for Scenario 5 (medium heterogeneity) across different settings. (Results are robust to scenario choice.) As we can see, the fraction is pretty negligible; even for $C = 100$, it is 0.17 percent in terms of appropriately discounted figures.

Some might argue that this comparison is simply the result of discounting returns far off in the future. Strictly speaking (and more simply still) it is the result of allowing consistently for the time value of money in all comparisons across time. The results would be the same if the present values were taken as of period N, since this involves multiplying both the numerator and denominator by the same (large) constant 1.018245^N. And this is precisely the point—the emphasis in the strategy literature on sustainable returns 'sometime in the future' tends to produce a rather elementary distortion of economic perspective, in many cases implying a radical understatement of the importance of the adjustment path in the wealth picture. Of course, correcting that distortion does not involve abandoning the insight that a sustainable profitable position is more valuable than an unsustainable one. Also, if one 'tunes in' only toward the end of an industry's evolution, the cross-sectional differences in sustainable profits may look impressive—much more impressive than they do in the broader perspective of wealth creation. This shows why we question the focus on flow profitability favored in RBV literature.

We can complement this analysis by giving sustainable profits a somewhat artificial advantage in the comparison. In the third row of Table 3, we compute present values for sustainable profits on the assumption that the ultimately sustainable portion counts as 'sustainable' from the start. If the point of the usual emphasis on 'sustainability' is to direct attention to strategic actions that promote it, these early-stage profits should not count, for they are realized without any such actions. We compare this expanded measure with the total NPV generated with the far-from-equilibrium start, counting the equilibrium 'share' of the profits from the first period. Even with this criterion, only in the $C = 100$ case do sustainable profits account for the majority of the total, and this is partly because (as noted previously) the industry starting position is closer to its ultimate equilibrium when customization cost is high. In that case, the equilibrium rent per unit is of approximately the same magnitude as the annual service price of the generic input, which seems like a generous allowance for the proportion of returns from idiosyncrasy.

While the first two columns of Table 3 show the base case and the $C = 100$ case, the next two columns provide the equivalent results on the 'oligopolistic expansion' setup. Interestingly, even in the case of oligopolistic output restraint, the ratio of the sustainable portion (including market

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

power effects) to total wealth creation is not substantially different from the competitive case. Again, sustainable profit is not the most important contributor to wealth creation, at least by most metrics.

A more direct comparison of how different factors affect the NPV of cash flow at the level of the industry comes from comparing what happens when we change the factors driving sustainable profits (i.e., when we change the conditions surrounding customization cost) and compare them to other changes in terms of competitive dynamics. Focusing on the top row of Table 3, we see first that NPV increases by 173 percent, as C changes from 10 to 100. Columns 5 and 6 offer yet another comparison. They provide the NPV figures for the $C = 10$ case and a robustness check with the reinvestment parameter θ set to 0.2 (Column 5) or to 1 (Column 6) as opposed to 0.5 in the base case (Column 1), which increases operating profit by 30 percent or reduces it by 24 percent, respectively. These substantial changes can be compared to a mere three percent increase in NPV when we admit oligopolistic calculation to the expansion conditions.

This suggests that a research focus on the factors that increase or decrease sustainability may be substantially less valuable than a focus on factors that affect the dynamic adjustment process. It also suggests, again, that a high-confidence assessment of the relative significance of transient and sustained profits will require better calibration to the broad historical record. The online appendix to this paper digs down into the evolutionary struggle between more and less effective firms that underpins the aggregate figures.

DISCUSSION

On the one hand, our model seems highly consistent with the theory of sustainable advantage presented in the RBV literature. On the other hand, it also seems to resolve some persistent puzzles about the relationship of that theory to standard microeconomics. We show how profitability that is *sustainable in long-run equilibrium, even under competitive conditions* may derive from the possession of idiosyncratic rent-earning resources, which, however, are not owned by the firm and which receive a 'market price.'

In our formulation, 'idiosyncrasy' matters not just because it might have something to do with the feasibility of imitation by rivals, but because it has everything to do with the input market conditions that allow a firm to trap the rent stream from a resource that it does not necessarily own. In the interest of framing this conceptual issue as sharply as possible, we have posited that a firm's specialized resource is literally unique—in the sense that the customizing investments that create it produce no value in any other firm. This extreme formulation is only a signpost, but it is valuable in understanding the complex reality in which the degree of idiosyncrasy varies with the routines of individual firms.

While the heritage of Ricardian rent theory remains highly visible, we depart from it in ways that seem plainly congenial to RBV thinking. The ultimate source of the value that resources contribute is not acquired from the market, but is something that the firm itself creates through a time-consuming process—and creates to an extent that is guided and constrained by market competition (see, again, Dierickx and Cool, 1989). Our formulation is, similarly, consistent with Barney's (1986) critique of the idea that superior returns can be captured simply by buying firms that possess superior resources. It is also broadly consistent with the VRIN/VRIO (value, rarity, inimitability, non-subtitutability/value, rarity, inimitability, organization) analysis of resource characteristics required for sustainability; in particular, we have presumed that such resources are 'inimitable.' But in general, we have suppressed important aspects of RBV thinking in the interests of simplicity.

The only resource 'leveraging' that is reflected in the model is the simple kind featured in basic evolutionary models—the profit-motivated effort to do more of the (profitable) same, which is accomplished by replicating routines (Winter and Szulanski, 2002). We have consigned to the prehistory of our process the creation of the idiosyncratic routines themselves, which are typically the fruit of a complex interweaving of creativity, learning, and investment over extended periods (Dierickx and Cool, 1989; Montgomery, 1995; Winter 1995; Levinthal 1997).

As we have suggested, the interpretation of any 'rent-as-profit' story must address the key question, 'can this return legitimately be called "profit"?' Ultimately, this is a matter of semantic taste—but we argue that the persistent returns in

Copyright © 2012 John Wiley & Sons, Ltd.

our model can reasonably be termed profits. They reflect continuing benefits to which no continuing outlay corresponds, and standard accounting will unquestionably show such a return as a part of net income. Similarly, the costs of customizing a generic input are typically expensed rather than capitalized; the books of real firms show no asset corresponding to most capability investments. Thus, our analysis of profit is conceptually consistent with empirical studies of profitability based on accounting measures. We believe that it is consistent with most uses of the profit concept in the strategic management field—though it points to the need for greater attention to returns on investment.

One possible objection is that our profits 'merely' reflect the normal return on investment. Actually, 'normal return' plays a narrow but critical role in our model; it controls the *margin* of the process of capability creation for the most capable firm. On infra-marginal units in the resource creation process, the most capable firm typically makes above normal returns. Other firms may benefit from their capability investments as well, even though they fail to exact rents corresponding to normal returns in the long run. Here and generally, the industry evolution framework serves to underscore the point that timing 'details' matter in economic competition, just as they do in some types of athletic competition—in the 100 meters, but not in the broad jump, for example. In the existing strategy literature, we have drawn inspiration from the 'uncertain imitability' model of Lippman and Rumelt (1982), which offers an analysis of sustained profitability with key similarities to our own. Their model, like ours, examines competition among heterogeneous firms. The *ex ante* perfect symmetry among potential firms is broken in their case by a chance mechanism that assigns different, and permanent, efficiency levels to different firms; in our case this mechanism is implicit in the learning contingencies that give rise to the differing routines and efficiency levels at the start of our dynamic process. Both mechanisms involve costs that must be incurred in order to establish a potential firm as an actual firm; since these are not continuing production costs, yet are voluntarily incurred only in anticipation of returns, the equilibrium picture is one of firms earning differential rents. In Lippman and Rumelt (1982) further entry is ultimately blocked by the upfront cost of acquiring capability and the uncertain result of attempting to do so. In our model, 'customization cost' plays a similar role, but it is an obstacle to growth rather than to entry.

Our analysis also differs from Lippman and Rumelt (1982) in a number of other respects, mostly having to do with its grounding in evolutionary economics. First, like most evolutionary models, ours respects the additivity axiom of production theory. By contrast, the production functions assumed by Lippman and Rumelt are not merely inimitable, but also non-replicable by the firm itself. Returns to scale are diminishing even in the long run, and this is the basic explanation for the fact that firms of different efficiencies can coexist in the equilibrium that the authors describe. The question of why the scale of application of information should be limited *even for its initial possessor* is left dangling—a shortcoming that can also be seen in more recent efforts to explain and model the coexistence of heterogeneous firms in competitive equilibrium (see Syverson, 2011). Our model allows for efficient firms to replicate their methods *while paying an investment price to do so.*

Second, our model shares the common evolutionary premise that mistaken decisions not only take place but also have lasting consequences for the evolution of the system—in fact, these negative consequences can be viewed as the obverse side of the benefits that behavioral variety brings to the system. Lippman and Rumelt (1982), by contrast, adhere closely to the full-rationality paradigm of mainstream economics, and manage with great analytical skill to describe an equilibrating process in which all of the worst mistakes (i.e., the investments with no returns) perish promptly and without consequence for other actors.

Third, we view the competitive process as occurring in real time, and compute the implications of behavioral rules embedded in time. Behavior is represented as driven by knowledge that actors might plausibly possess. Finally, we maintain that the notion of one-time costs of capacity expansion, as in the creation of a new establishment, has the virtue of greater realism than the idea of an entrepreneur's one-time purchase of a 'recipe,' or purely informational resource.

Our perspective also has some commonalities with Sutton (1991), who shows that 'endogenous sunk costs' (i.e., the need to commit funds irreversibly to activities such as advertising or research and development, which can improve efficiency or attract customers) can account for patterns

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

of industry structure and for sustained profitability. However, our concept of 'customization cost' relates to sunk costs that are *not independent of scale*, which is the crucial consideration that underlies our highlighted possibility of persistent heterogeneity in a model involving price-taking firms. More generally, our paper shows the importance of complementing the analysis of flow profitability at one point in time (and of long-run equilibrium) with the analysis of the NPV of cash flow (and of the adjustment path). That is what mainly matters to the appraisal of outcomes, from the viewpoint of either the focal private actors or society at large. Short-run and equilibrium perspectives, by contrast, primarily serve the interests of analysts and academics, whose lives are thereby made simpler.

More important, our paper advances a clear distinction between inputs, which are traded in the market, and idiosyncratic resources, which require a customization to make them part of the productive fabric of a firm. We argue that it is this customization, and its dynamics, that drives profits and competitive dynamics in a sector.

Limitations

Our exploration of the dynamic consequences of efficiency differences has not, of course, produced a 'rule for riches'—or even a promising set of hints. We have not accounted for the mechanisms that separated the winners from the losers in our arrays. The ultimate sources of superior efficiency—whether in creativity, persistent effort, or luck—remain beyond the scope of our analysis. The very diversity of these potential sources suggests that the prospects for future insights through the application of the sorts of techniques used in this paper should be assessed with restrained optimism. We have employed strong simplifications to clarify the economic logic of 'semi-permanent attachment' and 'sustained profitability.' These simplifications unfortunately tend to obscure other highly relevant aspects—the implications of entry and exit; the time-consuming nature of capability building; the financial implications of owned assets; and the consequences of secondary markets for partially customized assets (which may exist, even though thin).

Our experiments involve a stylized start to the industry evolution process, and invoke a rough judgment about how far from equilibrium such a start might realistically be. We did not explore how the 'distance from equilibrium' issue is entangled with the complexities of ongoing technological change and the demand changes associated with aggregate economic growth. Most importantly, our effort to clarify the subtle issues involving pricing of resources that are 'semi-permanently attached' to the firm led us to exclude the more substantial and straightforward 'attachment' of asset ownership. In so doing, we also excluded from the present-value analysis the wealth changes attributable to changes in the market prices of owned assets—which in many cases form an important part of the picture. Whereas the positioning school has tended to discuss timely purchase of key resources in terms of the creation of entry barriers, we would argue that this view must, at a minimum, be complemented by an acknowledgment of the straightforward wealth consequences of changing market valuations—changes that the combination of input scarcity and industry growth will very commonly produce (Jacobides, Knudsen, and Augier, 2006; Jacobides and Winter, 2007; Winter, 1995).

Our analysis is based on a very simple representation of the question of who appropriates the benefit of customization. We have assumed conditions under which firms can plausibly keep the near totality of the benefit of the customization costs they incur. Ranging from this extreme to the opposite one, where there is a single idiosyncratic actor who is not an owner but is of great significance to the firm, there is an interesting variety of situations that deserve more attention.

Finally, while we have discussed 'wealth creation,' we have not examined that concept itself in depth. We did not address its relationship to the 'destruction' aspect of 'creative destruction'—that is, the destruction of private wealth that often accompanies wealth creation of the kind we discuss. This is important not only because the dynamics of new and old are often closely linked but also because the broader perspective is essential to understanding the implications for the 'wealth of nations' as opposed to private wealth. Also, we did not make the link to shareholder returns and wealth at the investor level. Since the stock market registers opinions about the future as well as current results, it is reasonable to assume that wealth creation at the shareholder level is, in the aggregate, shifted toward the early stages by an even greater amount than our calculations suggest.

Copyright © 2012 John Wiley & Sons, Ltd.

However, the uncertainty that attends these situations implies that the net wealth creation at this level is the net result of a lot of ups and downs in the prospects of individual firms. Individual stock prices are moved, in both directions, as news of the evolutionary struggle flows in. If, as we recommend, the assessment of wealth creation moves up on the agenda of the strategy field, there will be a need for strong empirical methods appropriate for such long-term appraisals and for interpreting shorter-term movements in strategic terms. For such work, we urge attention to the inspiring example provided some time ago by George Baker's research on Beatrice (Baker, 1992).

CONCLUSIONS AND FUTURE AGENDA

This paper uses formal modeling to advance our theoretical understanding of some key strategy issues in five ways.

First, we elucidate the nature of the 'semi-permanent attachment' of resources, providing a simple model consistent with the overview of appropriation issues provided by Coff (1999).

Second, we show that heterogeneous firms can coexist in a competitive equilibrium, and establish that the degree of heterogeneity is directly linked to the firms' growth calculus.

Third, we show that sustainable profits in an industry can well emerge without any 'monopolistic' imperfections or information asymmetries in either the product market or the resource market. This strengthens the case for the 'economizing' viewpoint by directing attention to the dynamic logic of competition and industry evolution (cf. Williamson, 1999; Teece, Pisano, and Schuen, 1997).

Fourth, we provide an indicative list of dynamic factors that shape industry evolution and profit distribution (capability heterogeneity; resource customization cost; ease of expansion), and show how these elements interact to drive both short-run and sustainable profits.

Fifth, we illustrate that sustainable profits might represent only a small part of the wealth creation picture, and that changes in the factors that affect the adjustment processes of firm scale and profits over time may be much more important than changes in factors that determine the ultimate sustainability of profits.

In future work, we propose to apply our model (or some descendant) to a wider range of strategic issues. Some involve contextual factors, such as elasticities of demand and supply. Several will be pursued with more attention to analysis at the individual firm level, where the consequences of heterogeneity are played out. The question of how learning and imitation affect the picture is high on the list. Our preliminary explorations tend to confirm the basic insights of the RBV regarding the implications for long-run flow profitability, but also suggest the importance of confronting other aspects, such as the impact of different imitation rates on NPV. The model's simple story of a single customized resource does not, however, illuminate the subtle connections among complexity, idiosyncrasy and imitability—an area where much insight has been gained in recent years with the aid of the NK modeling framework (Kauffman, 1993; Levinthal, 1997; Rivkin, 2000, 2001; Lenox, Rockart, and Lewin, 2006; Csaszar and Siggelkow, 2010). Extending that framework to include the economic logic of customization cost would be an important step. Such an extension might examine in particular the implications that would follow if greater complexity, higher efficiency, and higher customization costs tend to go together.

Significant as this future agenda is, some important issues have already been clarified. This paper provides a bridge between the tenets of the RBV and microeconomic theory, as well as evolutionary theorizing. It puts forward a novel explanation for why differentially capable firms might coexist in competitive equilibrium, focusing on the role and nature of capability development. It considers the nature of profits that accrue to firms, as distinct from the returns attributable to resources. Finally, we propose a shift of emphasis from sustained-flow profitability to the logic of returns on investment, and from comparative static analysis to the dynamic adjustment process and its determinants. The need for such a shift in strategic analysis is this paper's most important message.

ACKNOWLEDGEMENTS

Our paper says, in the end, about what we intended it to say, in the beginning. If, however, it now says it in ways that might be widely understood, such a desirable outcome reflects valuable feedback from a wide range of colleagues, seminar audiences,

Copyright © 2012 John Wiley & Sons, Ltd.

reviewers and editors. In all of these categories, there was marked division of opinion. It is hardly necessary to pronounce the usual exoneration; we thank them all.

Winter acknowledges, more specifically, research support from The Mack Center for Technological Innovation, at The Wharton School, and Jacobides acknowledges support from the RMD Fund at the London Business School.

REFERENCES

Akerlof G. 1970. The market for lemons: quality uncertainty and the market mechanism. *Quarterly Journal of Economics* **84**: 488–500.

Baker G. 1992. Beatrice: a study in the creation and destruction of value. *Journal of Finance* **47**(3): 1081–1119.

Barnett W. 1995. Telephone companies. In *Organizations in Industry: Strategy, Structure, Selection*, Carroll G, Hannan M (eds). Oxford University Press: New York; 278–289.

Barney J. 1986. Strategic factor markets: expectations, luck, and business strategy. *Management Science* **32**(10): 1231–1241.

Barney J. 1991. Firm resources and sustained competitive advantage. *Journal of Management* **17**: 199–220.

Becker GS. 1993 [1964]. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education* (3rd edn). University of Chicago Press: Chicago, IL.

Bloom N, Van Reenen JV. 2007. Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics* **122**: 1351–1408.

Bloom N, Van Reenen J. 2010. Why do management practices differ across firms and countries? *Journal of Economic Perspectives* **24**(1): 203–224.

Brandenburger A, Stuart HW. 1996. Value-based business strategy. *Journal of Economics and Management Strategy* **5**: 5–24.

Chew WB, Bresnahan T, Clark KB. 1990. *Measurement, Coordination, and Learning in a Multiplant Network*. Harvard Business School Press: Boston, MA.

Christensen CM, Suarez FF, Utterback JM. 1998. Strategies for survival in fast-changing industries. *Management Science* **44**(12 Part 2): S207–S220.

Coff RW. 1997. Human assets and management dilemmas: coping with hazards on the road to resource-based theory. *Academy of Management Review* **22**: 374–402.

Coff RW. 1999. When competitive advantage doesn't lead to performance: the resource-based view and stakeholder bargaining power. *Organization Science* **10**: 119–133.

Coff RW. 2010. The coevolution of rent appropriation and capability development. *Strategic Management Journal* **31**(7): 711–733.

Csaszar F, Siggelkow N. 2010. How much to copy? Determinants of effective imitation breadth. *Organization Science* **21**: 661–676.

Demsetz H. 1973. Industry structure, market rivalry, and public policy. *Journal of Law and Economics* **16**(1): 1–9.

Dierickx I, Cool K. 1989. Asset stock accumulation and the sustainability of competitive advantage. *Management Science* **35**(12): 1504–1511.

Dosi G, Marsili O, Orsenigo L, Salvatore R. 1995. Learning, market selection and the evolution of industrial structures. *Small Business Economics* **7**(6): 411–436.

Foster RN, Kaplan S. 2001. *Creative Destruction: Why Companies That Are Built to Last Underperform the Market—and How to Successfully Transform Them*. Currency/Doubleday: New York.

Fudenberg D, Levine DK. 1989. Reputation and equilibrium selection in games with a patient player. *Econometrica* **57**: 759–778.

Gort M, Klepper S. 1982. Time paths in the diffusion of product innovations. *Economic Journal* **92**: 630–653.

Hatch NW, Dyer JH. 2004. Human capital and learning as a source of competitive advantage. *Strategic Management Journal* **25**(12): 1155–1178.

Helfat CE, Lieberman MB. 2002. The birth of capabilities: market entry and the importance of pre-history. *Industrial and Corporate Change* **11**: 725–760.

Jacobides MG. 2000. Capabilities, Transaction costs, information technology and profitability in the unbundling mortgage banking value chain: a study of the drivers and implications of changing vertical scope. Unpublished Doctoral Dissertation, University of Pennsylvania, Philadelphia, PA.

Jacobides MG. 2008. How capability differences, transaction costs, and learning curves interact to shape vertical scope. *Organization Science* **19**(2): 306–326.

Jacobides MG, Knudsen T, Augier M. 2006. Benefiting from innovation: value creation, value appropriation and the role of industry architectures. *Research Policy* **35**: 1200–1221.

Jacobides MG, Winter SG. 2007. Entrepreneurship and firm boundaries: the theory of a firm. *Journal of Management Studies* **44**(7): 1213–1241.

Jovanovic B. 1982. Selection and the evolution of industry. *Econometrica* **50**: 649–670.

Kauffman SA. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press:Oxford, U.K.

Klepper S. 1997. Industry life cycles. *Industrial and Corporate Change* **6**(1): 119–143.

Klepper S, Graddy E. 1990. The evolution of new industries and the determinants of market structure. *RAND Journal of Economics* **21**(1): 27–44.

Koopmans TC. 1957. *Three Essays on the State of Economic Science*. McGraw: New York.

Lenox MJ, Rockart SF, Lewin AY. 2006. Interdependency, competition, and the distribution of firm and industry profits. *Management Science* **52**: 757–772.

Levinthal D. 1997. Adaptation on rugged landscapes. *Management Science* **43**: 934–950.

Copyright © 2012 John Wiley & Sons, Ltd.

Lippman SA, Rumelt RP. 1982. Uncertain imitability: an analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics* **13**: 418–438.

Lippman SA, Rumelt RP. 2003a. A bargaining perspective on resource advantage. *Strategic Management Journal* **24**(11): 1069–1086.

Lippman SA, Rumelt RP. 2003b. The payments perspective: micro-foundations of resource analysis. *Strategic Management Journal*, October Special Issue **24**: 903–927.

Makadok R. 2010. The interaction effect of rivalry restraint and competitive advantage on profits: why the whole is less than the sum of the parts. *Management Science* **56**: 1090–1102.

Malerba F, Nelson R, Orsenigo L, Winter S. 1999. History-friendly models of industry evolution: the computer industry. *Industrial and Corporate Change* **8**: 1–36.

Mas-Collel A, Whinston MD, Green JR. 1995. *Microeconomic Theory*. Oxford University Press: New York.

Montgomery C. 1995. *Resource-Based and Evolutionary Theories of a Firm: Toward a Synthesis*. Kluwer: Boston, MA.

Nelson RR. 1994. Why do firms differ, and how does it matter? In *Fundamental Issues in Strategy*, Rumelt RP, Schendel DE, Teece DJ (eds). Harvard Business School Press: Boston, MA; 247–270.

Nelson RR, Winter SG. 1982. *An Evolutionary Theory of Economic Change*. Belknap Press: Cambridge, MA.

Peteraf M, Barney J. 2003. Unraveling the resource-based tangle. *Managerial and Decision Economics* **24**(4): 309–323.

Porter ME. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press: New York.

Rivkin J. 2000. Imitation of complex strategies. *Management Science* **46**: 824–844.

Rivkin J. 2001. Reproducing knowledge: replication without imitation at moderate complexity. *Organization Science* **12**: 274–293.

Rubin P. 1973. The expansion of firms. *Journal of Political Economy* **81**: 936–949.

Rumelt RP, Schendel DE, Teece DJ. 1994. *Fundamental Issues in Strategy*. Harvard Business School Press: Boston, MA.

Sutton J. 1991. *Sunk Costs and Market Structure*. MIT Press: Cambridge, MA.

Syverson C. 2004. Product substitutability and productivity dispersion. *Review of Economics and Statistics* **86**(2): 534–550.

Syverson C. 2011. What determines productivity? *Journal of Economic Literature* **49**(2): 326–365.

Teece DJ, Pisano G, Shuen A. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* **18**(7): 509–533.

Utterback JM, Abernathy WJ. 1975. A dynamic model of process and product innovation. *Omega* **3**: 639–656.

Walker FA. 1887. The source of business profits. *Quarterly Journal of Economics* **1**(3): 265–288.

Wernerfelt B. 1984. A resource-based view of the firm. *Strategic Management Journal* **5**(2): 171–180.

Williamson OE. 1999. Strategy research: governance and competence perspectives. *Strategic Management Journal* **20**(12): 1087–1108.

Winter SG. 1964. Economic 'natural selection' and the theory of the firm. *Yale Economic Essays* **4**: 225–272.

Winter SG. 1971. Satisficing, selection and the innovating remnant. *Quarterly Journal of Economics* **85**(2): 237–261.

Winter SG. 1995. Four Rs of profitability: rents, resources, routines and replication. In *Resource-Based and Evolutionary Theories of the Firm*, Montgomery C (ed). Kluwer: Boston, MA; 147–158.

Winter SG. 2010. The replication perspective on productive knowledge. In *Dynamics of Knowledge, Corporate Systems and Innovation*, Itami H, Kusunoki K, Numagami T, Takeishi A (eds). Springer-Verlag: Berlin, Germany; 93–119.

Winter SG, Szulanski G. 2002. Replication of organizational routines: conceptualizing the exploitation of knowledge assets. In *The Strategic Management of Intellectual Capital and Organizational Knowledge*, Choo CW, Bontis N (eds). Oxford University Press: New York; 207–221.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj

APPENDIX: ANTICIPATED CHANGE IN VARIABLE PROFIT PER UNIT OF RESOURCE

An important fact about the trajectories displayed in early stages of industry evolution is that initially high-profit prospects gradually give way to the more 'normal' prospects of approximate equilibrium. The manner in which this point is acknowledged in theory is a matter of great importance, by virtue of its strong implications for the actor choices governing system dynamics. It seems obvious that profit-seeking actors, even if assumed to be boundedly rational and imperfectly informed, are not likely to assume that the profit circumstances of the current period will persist indefinitely. The problem is to characterize what they might do instead. In search of a simple way to represent the way that prevailing trends affect actor calculations, we settled on a simple 'adaptive expectations' scheme focused on variable profit per unit of the resource. In effect, this scheme extrapolates recent experience of percentages changes in the quantity defined in Equation 9 above, namely:

$$\begin{aligned} V(t) &= [\Pi^*(t) + 1 \cdot K(t)]/R^*(t) \\ &= P(t) \cdot a \cdot [R^*(t)/K(t)]^{b-1} - W(t) \end{aligned}$$

It is more convenient to do the analysis in the changes of the natural log of $V(t)$, rather than the period-to-period percentage changes. (The convenience has to do with the fact that representing the relevant compounding across time would require use of geometric means, if we were to deal in the percentage changes directly.) Thus, in the spirit of adaptive expectations, we define $Lvdel(t) = \log(V(t)/V(t-1))$, the most recent change in the log of variable profit per R, and $Lvdav(t) = (1-d) \cdot Lvdav(t-1) + d \cdot Lvdel(t)$ is the moving average process for that change, yielding geometrically declining weights for the variable profit series with common ratio $(1-d)$, $0 \leq d \leq 1$. When $d = 1$, this formulation extrapolates the most recent change—*not* the most recent value of $\log(V(t))$. Then the period t estimate of the per-period relative rate of variable profit change is $z = \exp(Lpdav(t)) - 1$. In our calculations, we assume that firms allow for a single period of change at that rate, then extrapolate constancy from there on. In the industry evolution context, this one-period change is expected to be negative, but becoming less so as time goes on. Then $(1+z) \cdot V(t)$ would be the projected variable profit per R for the following period, and $(1+z) \cdot V(t)/i$ is the present value if that variable profit stream is assumed to continue indefinitely.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/smj