

Hyperspecialization and hyperscaling: A resource-based theory of the digital firm

Gianluigi Giustiziero¹  | Tobias Kretschmer^{2,3} 

Deepak Somaya⁴  | Brian Wu⁵ 

¹Frankfurt School of Finance & Management, Frankfurt, Germany

²Ludwig-Maximilians-Universität München, Munich, Germany

³Centre for Economic Policy Research (CEPR), London, UK

⁴University of Illinois Urbana-Champaign, Champaign, Illinois, USA

⁵University of Michigan, Ann Arbor, Michigan, USA

Correspondence

Brian Wu, University of Michigan, Ann Arbor, MI, USA.

Email: wux@umich.edu

Abstract

Research Summary: Digital firms tend to be both narrow in their vertical scope and large in their scale. We explain this phenomenon through a theory about how attributes of firms' resource bundles impact their scale and specialization. We posit that highly scalable resource bundles entail significant opportunity costs of integration (vs. outsourcing), which simultaneously drive "hyperspecialization" and "hyperscaling" in digital firms. Using descriptive theory and a formal model, we develop several propositions that align with observed features of digital businesses. We offer a parsimonious modeling framework for resource-based theorizing about highly scalable digital firms, shed light on the phenomenon of digital scaling, and provide insights into the far-reaching ways that technology-enabled resources are reshaping firms in the digital economy.

Managerial Summary: Why are leading firms in the digital economy simultaneously larger and more specialized than those in the industrial age? Our research explains this phenomenon as being driven by the scalability of digital resources—that is, their capacity to create more value at larger scales when used intensively in a focal activity. We clarify what digital scalability means, and

Authors contributed equally and are listed in alphabetical order.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](#) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Strategic Management Journal* published by John Wiley & Sons Ltd.

highlight trade-offs created by the opportunity costs of not employing scalable digital resources intensively. Digital firms should outsource complementary activities to avoid diverting resources away from their scalable core, and to enhance their ability to grow exponentially. Although resource fungibility and outsourcing costs mitigate these imperatives, digital firms may nonetheless find it profitable to remain specialized despite the challenges of managing outsourcing and sharing value with complementors.

KEY WORDS

digital firms, opportunity costs, resource-based view, scalability, scale and scope

1 | INTRODUCTION

The resource-based view (RBV) originally emerged to explain how unique firm-specific assets—conceptualized as resources or capabilities—lead to sustained competitive advantage (Barney, 1991; Mahoney & Pandian, 1992; Peteraf, 1993; Wernerfelt, 1984), but the RBV has since been extended to explain firm boundaries. Scholarship in the RBV has long held that firms should engage in activities for which they have superior resources (Argyres, 1996; Barney, 1999; Madhok, 1996, 2002), a rationale that has been refined and elaborated by combining it with transaction-costs and property-rights theories (Argyres & Zenger, 2012; Kang, Mahoney, & Tan, 2009; Kaul, 2013; Mayer & Salomon, 2006; Mayer, Somaya, & Williamson, 2012). These conceptualizations broadly fit with Chandler's (1977, 1990) descriptions of successful firms that emerged and grew in the second industrial revolution. Chandler highlighted how these industrial firms increased in both scale and scope by developing and redeploying resources and capabilities; they used new communication and transportation technologies to internalize key parts of the value chain, which enabled them to both avoid transaction costs and take advantage of their capabilities.

In recent years, the digital revolution has profoundly reshaped the way we conduct and organize business (Brynjolfsson & McAfee, 2014; Cusumano, Gawer, & Yoffie, 2019; Siebel, 2019), highlighting a need to reassess our understanding of economic organization and its drivers. In this article, we offer a theory based on the attributes of firm-level resources—in particular those enabled by the digital revolution—to shed light on firms' decisions about which value-adding activities to perform within their boundaries and which ones to outsource. An explanation of these decisions is generally referred to as the theory of the firm (Demsetz, 1988; Rumelt, 1984), underscoring its importance in explaining why firms exist and the role they play in the economy. Thus, *inter alia*, we propose a resource-based theory of the digital firm.

Our theory builds on the concept of “scalability,” by which we mean how the value derived from a firm's resource bundle in a focal activity changes as the size of the bundle increases. We use the term “digital firms” (contrasted with “industrial firms”) to describe firms that participate heavily in the digital economy by either using a significant share of digital resources (e.g., software, algorithms, data) and/or by selling a significant share of digital products and

services (e.g., platforms, software, media).¹ We posit that digital firms tend to have more *scalable* resource bundles due to significant economies of scale in their productive resources and due to markets with low distribution costs and strong network effects (Adner, Puranam, & Zhu, 2019; Autor, Dorn, Katz, Patterson, & Van Reenen, 2020; Hoffman & Yeh, 2018).

The greater scalability of digital firms' resource bundles affects their opportunity costs of integration, which requires allocating resources to multiple value-adding activities, rather than using them more intensively to grow within the focal activity. Consider Scale AI, a startup creating labeled datasets for artificial intelligence (AI) applications, such as datasets of street scenes used to train algorithms for automated driving. The company uses a platform of (contract) gig workers and algorithms to create labeled datasets, and in theory could integrate into developing AI for end user markets (by using its strengths in AI algorithms). However, this would mean that some of Scale AI's resources that cannot be contemporaneously shared across businesses (managers, software engineers, etc.) would have to be diverted from expanding its AI dataset labeling business. Thus, when a resource bundle is scalable, it may be preferable to concentrate resources on a focal activity than to distribute them across multiple activities, making specialization an optimal choice. Moreover, the scalability of the resource bundle creates a substantial push for a firm's growth, which (as our model shows) is further enhanced by specialization in its focal business. This core intuition highlights how digital firms' resource attributes may foster both high specialization and high scale, which consultants have called "hyperspecialization" and "hyperscaling" (McKinsey Global Institute, 2015), labels we adopt and later define in this article.

Many leading industrial firms of the past had both extensive scale and scope (Chandler, 1977, 1990), which we contrast with our theory of digital firms with high scale but narrow scope. For example, industry observers point to the rise of large but highly specialized cloud-based vendors such as Twilio, Stripe, Snowflake, PubNub, and Box who sell digital services tailored for narrow purposes but on a global scale (The Economist, 2021a). Industrial firms' capital-intensive production technologies also supported their growth through economies of scale (Diewert & Fox, 2008), but these scale advantages were limited by two factors. First, the scale economies were more confined by *physical* production capacity limits and further hindered by difficulties in replicating production processes (Knudsen, Levinthal, & Winter, 2014; Penrose, 1959). Second, because scalability is ultimately about the value accruing to the firm, which is in part determined by market demand, it can be reduced by the costs of transporting and distributing physical goods to distant markets, as well as the need to lower prices to sell more output due to downward-sloping demand curves. These limits to demand may have provided additional "inducements" to industrial firms for scope expansion (Penrose, 1959). Penrose (1959), who saw vertical integration as a "special form of diversification" (Penrose, 1959, p. 145), argued that industrial firms tended to expand their corporate scope whenever their capacity for growth was greater than that permitted by their existing markets. Similarly, Chandler (1990) documented that industrial firms which developed capital-intensive production technologies with the capacity to produce more than the market could absorb at reasonable prices often expanded into other value-adding activities that generated scope economies by using the same physical assets.

¹Of course, in practice, companies do not fall neatly into these two categories. Amazon, for instance, employs a combination of digital assets and other resources in its e-commerce business to sell both digital (marketplace) and physical (distribution) services. While we acknowledge this nuance, the broad distinction between (predominantly) digital and industrial firms is adequate to highlight our main ideas and theoretical logic.

Compared to industrial firms, we suggest that digital firms' resource bundles are more scalable *on average* and their scalability is more likely to persist through much higher volumes of output. This is because information technologies have made it possible to replicate and distribute digital goods and services at little incremental cost and allowed digital firms to sell large volumes globally without any significant physical presence. Further, the value of digital goods and services often increases in output due to network effects, which can produce demand-side increasing returns to scale. As a result, the supply-side scale advantages due to the adoption of digital resources are enhanced by the simultaneous effect that digitalization² has had on reducing demand constraints, which may induce digital firms to remain specialized even as they significantly expand their output. Thus, the changes to supply and demand conditions due to digital resources are complements in their effects on the scalability of digital firms, although their impacts on scalability may not be symmetric depending both on how much digitalization increases scale advantages on either side—supply or demand—and on which side acts as the bigger “constraint” to scalability. We discuss these asymmetric effects further in the discussion section.

We demonstrate the logic of our theory with a formal model that generates a set of propositions consistent with many features of modern digital firms (Adner et al., 2019; Cennamo & Santaló, 2019). Throughout the paper, we use the term “integration” to mean a firm's expansion into activities that add value to its focal offering, and “specialization” or “outsourcing” to mean the opposite. This terminology encompasses both value creation along a traditional value chain and in more fluid value networks such as ecosystems of independent firms producing complementary parts of a composite good. Also, we use the label “complementors” for firms offering non-focal value-adding activities, which includes “suppliers,” although there might be meaningful distinctions between these terms in some contexts.³

We contribute to the literatures on the RBV and the theory of the firm in at least three important ways. First, we highlight and explain a growing form of business organization that is highly specialized despite traditional motivations to integrate. We show how such hyperspecialization may stem from a firm's highly scalable resource bundle, unrelated to any effects of digitalization on transaction costs, and how scalability and specialization together can boost the firm's optimal scale of output (hyperscaling). Indeed, the rapid adoption of digital technologies has coincided with the growth of ecosystem forms of organization, whose participant firms can grow very large despite the ecosystem logic of nonintegration (Kretschmer, Leiponen, Schilling, & Vasudeva, 2020). Our theory explains these patterns as the results of a drive for value-adding firm growth in which less (scope) enables more (scale). Second, we contribute to a deeper understanding of the role of resources in (vertical) integration by underscoring a powerful, but often neglected, force: the opportunity costs of not scaling (i.e., increasing size or output) within a specialized activity. If resources are highly scalable, this force can even outweigh

²Two closely related terms—digitization and digitalization—have been used to characterize the impacts of digital technologies on businesses. Gartner (<https://www.gartner.com/en/glossary> [last accessed October 18, 2021]) defines digitization as “the process of changing [information] from analog to digital form” but “without any different-in-kind changes to the process itself.” Digitalization is defined as “the use of digital technologies to change a business model and provide new revenue and value-producing opportunities ... the process of moving to a digital business.” Our theory is fundamentally about the latter phenomenon; therefore, we use the term digitalization in this article.

³Suppliers also perform activities that are complementary to the focal firm's (Balakrishnan & Wernerfelt, 1986; Jacobides, Knudsen, & Augier, 2006; Richardson, 1972): if a downstream stage *a* generates value only with an upstream stage *b*, and vice versa, then the upstream stage and downstream stage are complementary in the sense “*a* doesn't ‘function’ without *b*” (Adner, 2017; Jacobides, Cennamo, & Gawer, 2018).

the traditional RBV logic of leveraging superior resources in multiple value-adding activities. In the limit, it can induce firms to become highly specialized and profit from small margins across very large numbers of customers. Thus, we propose that the adoption of digital technologies may favor specialization even without reductions in transaction costs (or costs of using the market generally), but because of significant increases in the scalability of firms' resource bundles. In turn, we offer an alternative (and complementary) resource-based explanation of why digital technologies lead to specialization, which also distinctively predicts the concomitant phenomena of hyperspecialization and hyperscaling that transaction cost based explanations typically do not. Finally, we advance an approach to formal modeling in the RBV by parameterizing key resource attributes such as scalability, fungibility, and costs of resource accumulation to provide a versatile platform for a broader research program on the impact of resources on corporate strategies and firm performance.

2 | THEORETICAL FOUNDATIONS

The RBV literature has long theorized that firms should integrate into activities whose resource requirements match the profile of their existing valuable, rare, inimitable, and non-substitutable resources (Argyres, 1996; Barney, 1991, 1999; Madhok, 1996, 2002; Montgomery & Wernerfelt, 1988). Research has further illuminated how integration decisions and the development of resources and capabilities are linked (Argyres & Zenger, 2012; Kang et al., 2009; Mayer et al., 2012; Wan & Wu, 2017), and more generally sought to combine the RBV with other theories of vertical integration such as transaction cost economics theory and property rights theory (Kaul, 2013; Mahoney & Qian, 2013; Mayer & Salomon, 2006). The literature has also incorporated intermediate governance modes, such as partial integration (Makadok & Coff, 2009; Parmigiani, 2007; Parmigiani & Mitchell, 2009), and has linked the RBV, evolutionary economics, and the modularity literatures (Baldwin & Clark, 2000, 2008; Helfat, 2015; Helfat & Camporro-Rembado, 2016; Jacobides & Winter, 2005). The key insight from these RBV contributions to the theory of the firm is essentially as follows: In a value chain (or ecosystem) of complementary activities, a firm will integrate into an activity if its resources are more productive in that activity than those of potential outsourcing partners.

With these ideas having evolved in the context of value chains and traditional pipeline business models, however, the current state of the art of RBV theory may not be adequate to fully explain the boundaries of digital firms. Digital firms have several distinctive organizational features and are thus an interesting (and important, due to their rapid proliferation) context for advancing the RBV theory on firm boundaries. Digital resources are often considered scale-free and fungible (Adner et al., 2019; Agrawal, Gans, & Goldfarb, 2018; Brynjolfsson & McAfee, 2014), which could in theory favor more integration. However, this prediction is at odds with the emergence of digital intermediaries and platforms such as Facebook, Google, and Airbnb, whose business models are often based on nonintegration within their platform's ecosystem (Kretschmer et al., 2020). Thus, these digital firms seem to be highly specialized on the vertical dimension, while at the same time growing very large in scale (Adner et al., 2019; Hoffman & Yeh, 2018).

Drawing on several descriptions of the digital transformation of the economy (Adner et al., 2019; Agrawal et al., 2018; Brynjolfsson & McAfee, 2014; Siebel, 2019), our conception of digital firms focuses on two central features. First, these firms employ, to a significant degree, digital resources such as data, software, and AI that are essentially scale free, such that firms'

marginal costs remain low for very large production quantities (Adner et al., 2019; Levinthal & Wu, 2010). Second, digital firms distribute their offerings largely through the internet and cloud platforms (Siebel, 2019), and thus have substantial access to global markets at scale, which mitigates decreasing returns to scale on the demand side. Moreover, many such digital firms are organized as digital platforms or intermediaries (Cennamo & Santaló, 2019; Kretschmer et al., 2020), and may even experience demand-side increasing returns to scale due to network effects (Teece, 2013; Teece, Pisano, & Shuen, 1997). In practice, firms can vary in the extent to which they are “digital,” reflecting different degrees to which they have employed digital resources or sell through digital channels and platforms; thus the resulting scalability of their resource bundles may vary (see Table 1 for definitions of, and the interrelationships between, the concepts of scalability, scale-free resources, and supply-side and demand-side returns to scale). Similarly, some industrial firms may also have resource bundles that enjoy a significant degree of scalability (such as the example of semiconductor fabs below). We posit, however, that digital firms (which are highly digitized along the two dimensions—resources and markets—described above) will *on average* have more scalable resource bundles than industrial firms, and their scalability will likely be sustained over a larger range of output.

Building on the conception of the firm as a bundle of resources (Amit & Shoemaker, 1993; Barney, 1991; Penrose, 1959; Wernerfelt, 1984), we develop a firm-level theory of integration (vs. specialization) that is built on this scalability feature of digital firms. We integrate ideas from the world of technology, where practicing experts often discuss firms' strategies in terms of “scaling laws” (Hoffman & Yeh, 2018; Levie & London, 2018; Wessel, Levie, & Siegel, 2017). Scaling laws are functions of the type $f(x) = kx^\sigma$ which, despite their simplicity, can describe the scaling properties of a variety of complex adaptive systems ranging from organisms (e.g., animals and plants) to organizations (e.g., cities and firms) (Fu et al., 2005; Gabaix, 2016; West, 2017). We use scaling laws to describe the relationship between *value created* (benefits to consumers minus costs) and the size of a firm's *resource bundle*. If this relationship follows a scaling law, we can summarize it with a continuous “scalability” parameter corresponding to the combined effect of supply-side and demand-side returns to scale.

The scalability of a firm's resource bundle reflects in part the types and relative shares of scale free and non-scale free resources in the bundle (Levinthal & Wu, 2010). Digital resources—such as IT systems, cloud platforms, big data, and AI algorithms—tend to be scale free by virtue of almost error-free replication, combined with low-cost global digital distribution and improvements in cost and performance as more users adopt and contribute data to them (Adner et al., 2019; Agrawal et al., 2018; Brynjolfsson & McAfee, 2014).⁴ These digital resources do not face many limits on the extent of their application, so the firm faces fewer capacity constraints as it grows in scale in the focal market (e.g., a cloud service can accommodate numerous additional users at minimal marginal cost). However, to create value, even these scale free resources need some complementary resources (Miller & Yang, 2016; Tambe, Hitt, Rock, & Brynjolfsson, 2020) such as co-specialized human and managerial resources (Castanias & Helfat, 1991; Teece, 1986), which are typically subject to capacity constraints regarding time and attention (Ocasio, 1997; Penrose, 1959). For example, software and AI platforms need experienced engineers to develop, maintain, and improve them, marketers and salespersons to sell their outputs, customer service professionals to improve service quality, and managers to

⁴The scaling properties contributed by digital technologies do not mean that digital firms have no size limit to which they can scale; this limit might simply be quite large and comparable to global demand. The literature on scaling laws also suggests that real world processes are only typically scalable within some range (West, 2017).

TABLE 1 Scalability and related concepts

Construct	Definition	How it relates to scalability
Scalability	Scalability describes how the value accruing to the firm from using its resource bundle in a focal activity changes as the size of the bundle increases. Scalability builds on the idea of <i>value</i> , which captures the interaction between resources and market characteristics—"two sides of the same coin" (Wernerfelt, 1984). As such, scalability corresponds to the net effect of the supply-side returns to scale and demand-side returns to scale. We examine regimes where the resource bundle is scalable (its value per unit of output increases) versus non-scalable (value per unit of output decreases) as the size of the bundle increases.	
Supply-side returns to scale	Supply-side returns to scale describe the change in output quantities due to an increase in inputs. If such an increase leads to a more (less) than proportional increase in output, then there are increasing (decreasing) supply-side returns to scale. Decreasing supply-side returns to scale are often ascribed to the more intensive use of factors whose supply is fixed, such as industrial plants in the short run or land in the long run. Increasing supply-side returns to scale are due to savings in factor requirements per unit of output due to scale-free resources, learning curves, positive feedback loops, or the internal division of labor (see Kim (1997) and Mas-Colell, Whinston, and Green (1995) for a discussion). Depending on input costs, increasing returns to scale can lead to economies of scale—a reduction in the average cost of production for larger output quantities.	The stronger the supply-side returns to scale, the more scalable the firm's resource bundle. However, scalability can be constrained by the demand side (e.g., inelastic demand) even when supply-side returns to scale are high.
Demand-side returns to scale	Demand-side returns to scale describe the change in revenues due to an increase in output. If larger outputs lead to a more (less) than proportional increase in revenues, then there are increasing (decreasing) demand-side returns to scale. Decreasing demand-side returns to scale are often associated with physical goods, for which larger quantities sold correspond to a reduction in prices (law of demand). Increasing demand-side returns to scale are associated with digital goods having network effects (Arthur, 1996; Teece, 2013; Teece et al., 1997), for which the value from consumption increases with the installed base.	The stronger the demand-side returns to scale, the more scalable the firm's resource bundle. However, scalability can be constrained by the supply side (e.g., non-scale free resources or decreasing returns to scale) even when demand-side returns to scale are high.
Non-scale free and scale free resources	"A scale free resource, such as brand name, faces limits on the breadth of its fungibility (i.e., how broadly fungible is a given brand name) but not on its extent of application (i.e., the number of markets in which a given brand can be applied for a given level of fungibility). In contrast, the application of those non-scale free capabilities is driven by the logic of opportunity costs." (Levinthal & Wu, 2010, p. 784)	The larger the share of scale free resources in the bundle, the more scalable the resource bundle, at least on the supply-side.

oversee and direct the enterprise. Often, physical resources such as factories, offices, and warehouses, and even hardware and telecommunication infrastructure to host and deliver digital products, are also required. Using these complementary non-scale free resources incurs opportunity costs (Levinthal & Wu, 2010); for example, an engineer working on a project cannot simultaneously work on another, and warehouses or application servers face congestion costs when they are used to serve too many customers simultaneously.

In this way, combinations of co-specialized scale free and non-scale free resources in a firm's resource bundle (Teece, 1986) interact and contribute to scalability on the supply side. Similarly, the extent to which digital distribution and network effects contribute to scalability on the demand side may also vary. Thus, scalability captures the aggregate effect of all the resources in the firm's resource bundle and measures the extent to which the value created increases with the extent of the resource bundle employed in a particular "activity" (equivalently a "stage of production"). The resource bundle will likely exhibit a tendency to create progressively more value at larger output volumes when the firm employs a substantial share of resources such as data, software, and AI that are virtually scale free (Adner et al., 2019; Levinthal & Wu, 2010) and, in addition, has access to global markets with network effects that can sustain (or even increase) prices as sales grow. By contrast, the resource bundle will likely have a tendency to create progressively less value at larger scales when the firm faces physical production limits, difficulties in replicating capital-intensive processes (Knudsen et al., 2014; Penrose, 1959), and substantial costs associated with the transportation and distribution of physical goods. At the same time, a downward-sloping demand curve would also force the firm to reduce prices to sell larger output volumes, which generates decreasing demand-side returns to scale.

Our categories of digital and industrial are meant to represent archetypes of firms in each category. In practice, the degree to which a firm is "digital" will in turn affect the degree to which its resource bundle is scalable. Consider Amazon's online retail business and its Amazon Web Services (AWS) cloud business. Clearly Amazon's retail business uses a greater share of non-digital resources such as warehouses, inventory, packers, and other employees, and has a greater share of non-digital outputs (e.g., physical products, delivery), and comparatively AWS has a more digitalized (and therefore scalable) resource bundle than Amazon's online retail business. While acknowledging these heterogeneities in the extent to which a firm is digitalized, we use the term "digital firm" to indicate an archetypal firm that has largely digital resources and outputs, and contrast it with an archetypal "industrial firm" that does not. Further, we designate a firm's resource bundle to be "scalable" (non-scalabe) if its value per unit of output increases (decreases) as the size of the bundle increases. We posit that archetypal digital and industrial firms are likely to have scalable and non-scalable resource bundles, respectively, even though not all digital firms may have scalable resource bundles, and not all firms with scalable resource bundles might be digital firms.

To examine the implications of these attributes of the firm's resource bundle, we develop a formal model to explain the firm's integration choices and degree of scaling. Our core intuition is that when a scalable resource bundle can be used to either scale within a focal activity or to increase scope into complementary ones, it leads to specialization because of the high opportunity cost of not focusing on the focal value-adding activity as intensively as possible, even if the resource bundle is fungible to other activities. Figure 1 visualizes this intuition. Let τ be the percentage of the resource bundle, r , allocated to a given application and $V(\tau \times r)$ be the value produced as a function of $\tau \times r$. When the resource bundle is non-scalable (as in an archetypal industrial firm), the opportunity cost of withdrawing resources from the focal activity is

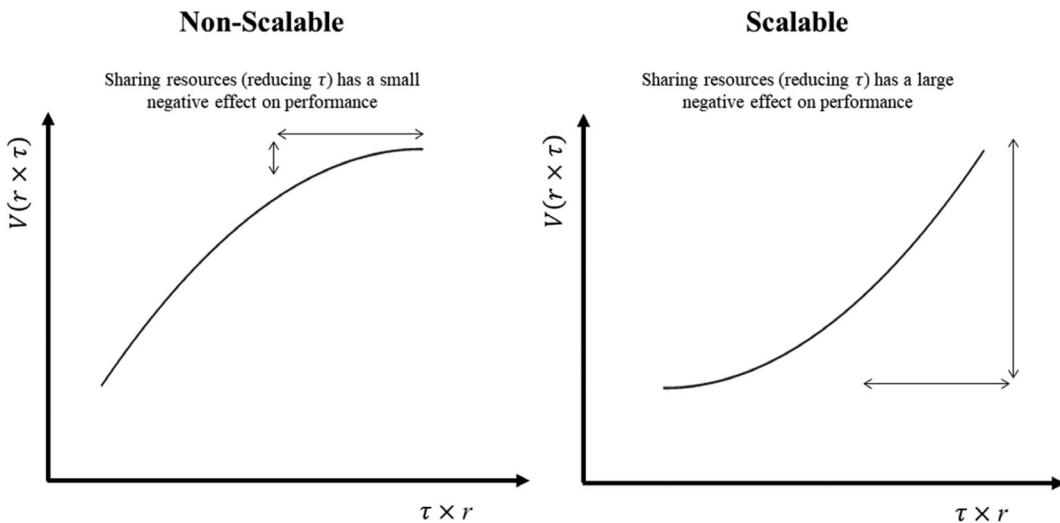


FIGURE 1 Non-scalable versus scalable resource bundles. The figure is a stylized representation of opportunity cost of sharing resources when the resource bundle is non-scalable (left side) and scalable (right side). When the resource bundle is non-scalable, the opportunity cost of withdrawing resources from one application is relatively low because reducing τ (i.e., the percentage of resources allocated to a given application) has a minor impact on performance. When the resource bundle is scalable, the opportunity cost of redeploying the same amount is higher and can result in a significant penalty

relatively low; reducing τ has only a small impact on performance. When the resource bundle is scalable (as in an archetypal digital firm), however, the opportunity cost of redeploying the same amount away from the focal activity is high and can significantly reduce performance.

We present five propositions derived from our model. Our first proposition contrasts the overall integration patterns between archetypal digital and industrial firms. Specifically, for an industrial firm, a marginal change in the cost of outsourcing typically leads to gradual changes in integration, with varying degrees of partial integration as intermediate modes between full specialization and full integration (Makadok & Coff, 2009; Parmigiani, 2007; Parmigiani & Mitchell, 2009); however, for a digital firm, such a marginal change may trigger a discontinuous shift from full specialization to full integration. Propositions 2 and 3, respectively, examine the contrasting roles of scalability and fungibility for digital and industrial firms. We show that the propensity to integrate increases with both scalability and fungibility for the archetypal industrial firm, but they have contrasting effects for the archetypal digital firm. Scalability induces a digital firm to remain specialized even at higher costs of using the market, while fungibility increases the propensity to integrate.

Having examined scope decisions, Proposition 4 formalizes the relationship between scale and scope choices of the firm. The result shows that scale and scope (integration) go hand in hand for industrial firms (Chandler, 1990), whereas scale and specialization are mutually reinforcing for digital firms. Put simply, digital firms have a greater incentive to increase their scale (by augmenting their resource bundle) if they specialize, and the fact that they can grow significantly within their focal activity is a key reason to specialize. In Proposition 5, we extend our model by endogenizing value capture and co-operation with the complementor firm(s). Our results paint a stark contrast between our theory and the canonical proposition of the RBV

literature that firms will integrate if their resources are more productive than those of their complementors. We show that for digital firms, superior resources may not necessarily lead to integration. Instead, up to a point, resource superiority leads the digital firm to forego a greater share of the total value created together with the complementor(s) so as to incentivize supply of the complementary product or service, thus increasing its own returns by specializing and scaling within its focal activity.

3 | FORMAL MODEL

We formalize our arguments in a parsimonious decision-theoretic model that casts scope expansion strategies as the solution to a resource-allocation problem involving two complementary value-adding activities.⁵ The model borrows elements from the literature on non-scale free resources and resource redeployment (Levinthal & Wu, 2010; Sakhartov & Folta, 2014, 2015) as well as from the value-based literature in strategic management (Chatain, 2014; Chatain & Zemsky, 2007, 2011; Jia, 2013; Postrel, 2018; Wan & Wu, 2017). By design, the model does not incorporate coordination costs, technological interdependencies, or supermodular complementarities. Therefore, if integration occurs, it does so due to resource characteristics, even in the absence of these often-cited drivers of integration decisions (Alchian & Demsetz, 1972; Garicano & Wu, 2012; Helfat & Campo-Rembado, 2016; Postrel, 2009).

For ease of exposition, we first assume perfectly elastic demand so as to focus our attention on the properties of internal resources. Later, we generalize this approach to include richer characterizations of the demand environment, and show that our results can be readily extended to incorporate the combined effects of resource-driven and demand-side returns to scale.

3.1 | Resource attributes and production

To illustrate the basic mechanisms at play, consider an economy with a simple demand environment where the final customer pays a constant amount V for the final product $a \wedge b$,⁶ consisting of two components, a and b , in one-to-one proportion, with the value of the components in isolation being normalized to zero. The bundle $a \wedge b$ can be thought of as the output of an ecosystem whose components exhibit strict complementarities of the type that “ a doesn’t ‘function’ without b ” (Adner, 2017; Jacobides et al., 2018) as in the case of processing units and operating systems for personal computers (Casadesus-Masanell & Yoffie, 2007); or of an ecosystem whose components exhibit complementarities of the type that “ a functions better with b ” as in the case of smartphones and compatible applications (in this latter case, the value of a smartphone without applications is nonzero, but it can be normalized to zero so that V represents the value added by the applications (Postrel, 2018)).

To produce a and b , the focal firm (firm i) allocates its resource bundle, r , to the two activities by allocating $\tau \in [0, 1]$ to activity a and $(1 - \tau)$ to b , generating outputs $Q_{ia}(\tau r)$ and

⁵A decision-theoretic model allows us to highlight the core results around the internal resources of the focal firm. We later study a game-theoretic model endogenizing competition and value capture, both showing the robustness of our main results and generating additional insights.

⁶The demand for the bundle $a \wedge b = \min\{a, b\}$ is perfectly elastic, meaning that V is constant and independent of the quantity of $a \wedge b$ supplied to the market.

$Q_{ib}((1-\tau)r)$, respectively. The production functions Q_{ia} and Q_{ib} follow scaling laws with scaling exponent $\sigma > 0$ characterized by the relation $Q_{ia}(tr) = t^\sigma Q_{ia}(r)$ and $Q_{ib}(tr) = t^\sigma Q_{ib}(r)$ for any scalar t .⁷ The scaling exponent here, σ , corresponds to the supply-side returns to scale of the firm's resource bundle. However, as we later demonstrate, σ can more generally be interpreted as a parameter that captures the combined effects of supply-side and demand-side returns to scale, that is, as the overall scalability of the firm's resource bundle. In this interpretation, if $\sigma > 1$, then the resource bundle is scalable, which represents the archetype of a digital firm in which supply-side returns to scale due to economies in software development (Arthur, 1996) and positive feedback loops in data analytics (Agrawal et al., 2018) are reinforced by demand-side network effects (Adner et al., 2019; Arthur, 1989, 1996; Sutton, 1997; Wessel et al., 2017). If $\sigma < 1$, the bundle is non-scalable, which would be consistent with an archetypical industrial firm for which decreasing returns from a downward-sloping demand curve combine with supply-side returns to scale that are either decreasing, constant, or only slightly increasing.⁸ The case of $\sigma = 1$ separates the scalable and non-scalable ranges of the scalability parameter.⁹

The focal firm's resources have a greater baseline productivity in a , the main activity, than in the complementary activity b , so that $\frac{\partial VQ_{ia}(1)}{\partial r} = \varsigma(1) \geq \frac{\partial VQ_{ib}(1)}{\partial r} = \varphi(1) > 0$. As in prior work (Levinthal & Wu, 2010), the relative magnitude of the constants $\varsigma(1)$ and $\varphi(1)$ defines the fungibility of the firm's resources, that is, the degree to which one unit of the firm's resources performs worse when redeployed from the main activity to the complementary one (Anand & Singh, 1997; Montgomery & Wernerfelt, 1988). Our parameter space also contains the limit case $\varsigma(1) = \varphi(1)$, allowing for resources to perform equally well in both applications (i.e., resources can be perfectly fungible).

The cost of accumulating resources is given by $C(r)$. The function $C(\cdot)$ goes through the origin, is twice differentiable, monotonically increasing, and convex, and satisfies the condition $\frac{\partial C(r)}{\partial r} \leq \varsigma(r)$ for $r \leq r'$ and $\frac{\partial C(r)}{\partial r} > \varsigma(r)$ for $r > r'$, where $\varsigma(r) = \frac{\partial VQ_{ia}(r)}{\partial r}$. This condition ensures that the firm will not accumulate an infinite amount of resources, because at some point the productivity of the firm's resource bundle at the margin does not justify additional investments in resource accumulation (Viner, 1932; West, 2017). Indeed, it is reasonable to expect that the accumulation costs of resources may be convex due to time compression diseconomies, which eventually outweigh the benefits of internally developing additional resources (Dierickx & Cool, 1989; Giustiziero, 2020; Giustiziero, Kaul & Wu, 2019; Pacheco-de-Almeida & Zemsky, 2003, 2007). Ultimately, firms accumulate resources through a coordinated process of acquiring resource factors externally and combining them with internally developed firm-specific components, both of which are difficult to ramp up quickly. For example, the scaling of a digital resource bundle may be limited by the co-specialized non-scale-free resources in the bundle, which become increasingly costly to accumulate rapidly. Examples of such co-

⁷The functions Q_{ia} and Q_{ib} are homogeneous of degree σ . Homogeneity is a property of a large family of production functions, including the Cobb-Douglas production function.

⁸Because scalability approximates the reduced form of supply-side and demand-side returns to scale, it is reasonable to assume σ to be the same for the two components a and b in the current situation, allowing us to focus on core constructs. Demand-side returns to scale are the same in the two activities because the demand from consumers is for the complete product $a \wedge b$, not for the individual components. Moreover, returns to scale on the supply side are also likely to be similar. Were they to differ significantly, the underlying productive resources would have to be different, thus violating a key premise of the resource allocation problem in the model that the resource bundle is applicable to both value-adding activities. Future research can explore situations where σ differs across components.

⁹Mathematically, the case $\sigma = 1$ generates equivalent results to the case $\sigma > 1$ as far as the propositions in the paper are concerned. We use term "scalable resource bundles" for $\sigma > 1$ for expositional purposes.

specialized resources include human capital to create, maintain, and upgrade the digital resources, and complementary resources in functions like logistics, marketing, and customer service, as well as complementary physical assets like servers and warehouses.

3.2 | Integration and outsourcing regimes

The allocation of the resource bundle to value-adding activities maps onto three regimes: full integration (perform both a and b in-house); partial integration (perform a and some b in-house and source additional b from complementors); and specialization (perform only a in-house and fully source b from complementors). Formally, given the focal firm's resource allocation τ_I such that equal quantities of both components are produced in-house, $Q_{ia}(\tau_I r) = Q_{ib}((1 - \tau_I)r)$, then: (a) full integration is equivalent to choosing $\tau = \tau_I$, (b) partial integration to $\tau = \tau_C \in (\tau_I, 1)$, and (c) specialization to $\tau = \tau_S = 1$.

If the focal firm outsources b , then αV , with $\alpha \in (0, 1)$, is the share of the total value (created by the firm and complementor(s) together) that is captured by the complementor(s) for each unit of b . The parameter α measures the per unit value foregone by the focal firm by outsourcing, and thus represents the cost of using the market (which can be interpreted more broadly than simply value sharing).¹⁰ This resembles an *ad valorem* contract (Hagiu & Wright, 2019; Wang & Wright, 2017) typical of platform intermediaries (e.g., Amazon, Uber) and licensing deals (e.g., Netflix, Spotify), but also a parameter proportional to the price paid by a buyer to a supplier (Bennett, 2013). For now, we consider α as exogenous to the firm's decision, which lets us focus on the role of internal resources in the focal firm's choices and outcomes. We later relax this restriction and develop a more general model where the "realized α " is the result of a bargaining process between the focal firm and one or more complementors.

Together with the allocation of resources, the firm also determines how much to invest in its resource bundle according to the cost function $C(r)$, thus facing a two-variable optimization problem in τ and r . Because the order of the maximization does not affect the outcome (Athey, Milgrom, & Roberts, 1998), we consider a two-stage maximization where the firm first decides how much to invest in its resource bundle and then how to apportion its resource bundle between the two activities a and b (Kretschmer & Puranam, 2008; Levinthal & Wu, 2010). Working backward, the optimal resource allocation in the second stage maximizes the objective function $\pi(\tau, r) = VQ_{ia}(\tau r) - \alpha V(Q_{ia}(\tau r) - Q_{ib}((1 - \tau)r))$, where $VQ_{ia}(\tau r)$ corresponds to the revenues from the final good and $\alpha V(Q_{ia}(\tau r) - Q_{ib}((1 - \tau)r))$ to the costs of outsourcing component b . After rearranging, the objective function can be written as:

$$\pi(\tau, r) = \underbrace{(1 - \alpha)VQ_{ia}(\tau r)}_{\text{Benefits of Specialization}} + \underbrace{\alpha VQ_{ib}((1 - \tau)r)}_{\text{Opportunity Cost of Specialization}} \quad (1)$$

The two terms on the right-hand side of Equation (1) highlight the tradeoff faced by the firm when allocating its resources. When $\tau = 1$, the firm specializes in a and generates revenues

¹⁰As we show in our model extension in the Online Appendix, the cost of using the market can include transaction costs.

equivalent to the first term, which corresponds to the benefits of specialization. However, if $\tau=1$, then $(1-\tau)=0$ and the firm foregoes any share of value accruing from the second term, which represents the opportunity cost of specialization. If the benefits of specialization always outweigh the opportunity cost of not capturing all the value through integration for at least some of the output, then the firm will allocate all of its resources to a . If not, the firm will opt to capture more (partial integration) or all (full integration) of the value created by allocating at least some of its resources to the complementary activity b .

3.3 | Solution space

Figure 2 describes the solution space of our model, showing the relationship among the scalability of the resource bundle, the value distribution, and sourcing regimes. (An analytical solution to the firm's maximization problem is provided in the Online Appendix.) Darker colors correspond to lower values of τ , and consequently to more integration. The black solid line traces the value of α beyond which collaborating with the complementor(s) is too costly, triggering full integration. We denote this value as “critical α ” or α^* . The critical α is a function of the value of the firm's outside option, integration, relative to outsourcing.¹¹ As Figure 2 demonstrates, when α is below the critical threshold α^* , the firm benefits from outsourcing, but in different ways for industrial and digital firms. This contrast is summarized in the following proposition.

Proposition 1. (Discontinuous integration): *For industrial firms, an incremental change in the cost of outsourcing triggers a gradual change in integration; that is, integration choices are evaluated at the intensive margin. In contrast, for digital firms, an incremental change in the cost of outsourcing can trigger a discontinuous change between full integration and full specialization; that is, integration choices are evaluated at the extensive margin.* (Proof in Online Appendix.)

Proposition 1 demonstrates that the optimal strategy follows different logics depending on whether the firm's resource bundle is non-scalable ($\sigma<1$) or scalable ($\sigma>1$). When the resource bundle is non-scalable (industrial firms), productivity in the firm's main activity eventually plateaus and the firm can improve resource utilization by redeploying some resources from the main activity to the complementary one. So, profits increase along the intensive margin via a partial integration strategy that equates the (marginal) performance of the firm's resources in a to that in b . Partial integration generates the distinctive “fan-shaped” region on the left side of Figure 2, which is the result of tradeoffs between two underlying mechanisms: scalability facilitates the use of the firm's resource bundle in multiple applications, which leads to more integration, but it also makes some specialization attractive because performance in the main application deteriorates more slowly.

As in prior research (Parmigiani, 2007; Parmigiani & Mitchell, 2009), partial integration in our model requires resources to be fungible. However, even perfect fungibility is insufficient; the resource bundle also needs to be non-scalable. Thus, we expect partial integration to be more common in industrial settings, a result consistent with both within-industry studies

¹¹The critical α ($=\alpha^*$) can asymptotically get close to 1, but cannot reach (or exceed) 1; otherwise, the focal firm would incur a loss or make no profits with specialization, an outcome inferior to what it could attain with integration.

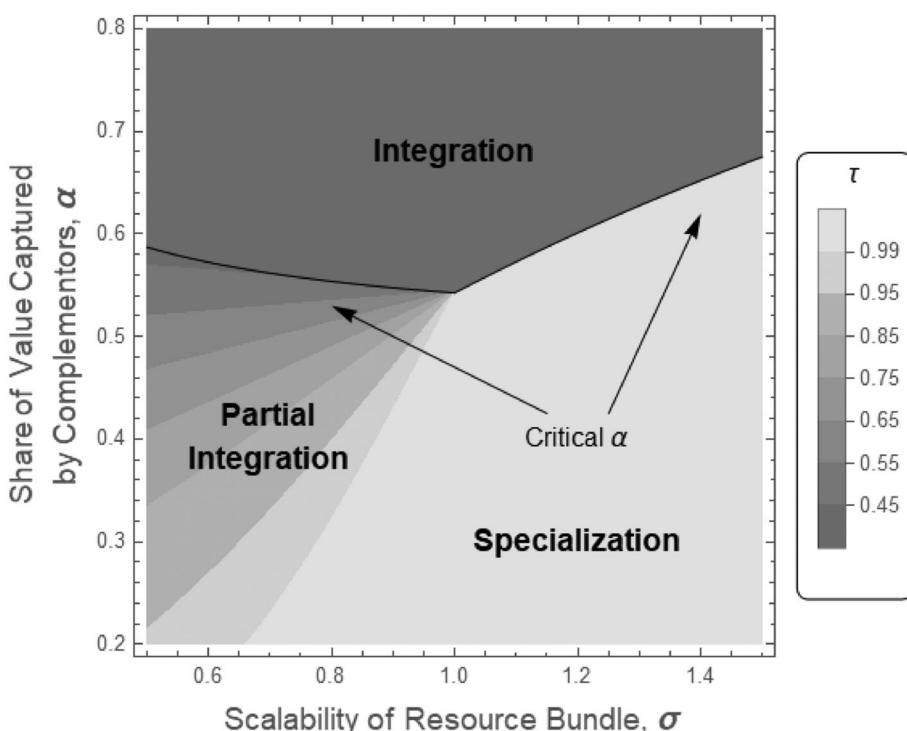


FIGURE 2 Scalability, value distribution, and sourcing regimes. The figure reports the results of a numerical simulation of the relationship among the scalability of the resource bundle, the value distribution, and sourcing regimes, represented by the parameters σ , α , and τ , respectively. Darker colors correspond to more integration. The black solid line traces the critical value of α ($= \alpha^*$) beyond which collaborating with the complementor is too costly, triggering full integration. The optimal strategy follows different logics depending on whether resources are non-scalable (left side, $\sigma < 1$) or scalable (right side, $\sigma > 1$)

(Parmigiani, 2007; Parmigiani & Mitchell, 2009) and cross-industry ones (Atalay, Hortaçsu, & Syverson, 2014). In contrast to the equilibration logic that drives partial integration for non-scalable resource bundles, the resource allocation of scalable resource bundles ($\sigma > 1$, as in the right panel of Figure 1) leads to “bang–bang” outcomes that skip intermediate concurrent sourcing. Because the performance of a scalable resource bundle never plateaus, the firm cannot maximize along the intensive margin. It must examine alternatives at the extensive margin, comparing the overall profits of the two corner solutions: (full) integration and (full) specialization. If the value captured by the complementor (α) is not high enough to outweigh the benefits of growth in the main activity, the firm specializes fully in pursuit of exponential growth in the main activity. Such outcomes mirror the “full stack” (integration) versus “no stack” (specialization) debate in Silicon Valley. Full-stack firms, like Peloton, an exercise equipment and media company, “build a complete, end-to-end product or service. ... [which requires that] ... you need to get good at many different things: software, hardware, design, consumer marketing, supply chain management, sales, partnerships, regulation, etc.” (Dixon, 2014), while no-stack companies, like Twilio, a cloud communication platform, “focus on the last mile of value they provide ... focus on doing only one thing—hopefully well—and utilize other services for everything

else” (Weissman, 2015). As Fred Wilson (2015), a venture capitalist, puts it, “the best approaches are at both ends of the spectrum. Either go full stack or go no stack.”¹²

Interestingly, the contrast highlighted in Proposition 1 is also consistent with the Silicon Valley mantra for early-stage ventures: “to scale, do things that don’t scale” (Hoffman & Yeh, 2018). Consider the case of Airbnb. When Airbnb was a tiny startup, well before it became a digitally scaled firm, its founders realized that the chances of renting a room on Airbnb depended greatly on the quality of photographs. Because Airbnb was operating a rudimentary website and had not yet invested much in digital technologies, the founders did this activity in-house for a significant portion of their hosts (those in New York). Chesky (2017), CEO and co-founder of Airbnb, recounts: “We literally would knock on the doors of all of our hosts [in New York]. We had their addresses and [...] we’d show up at their door and they’re like ‘Wow. This company is pretty small.’” However, once the company reached a critical mass and increased scalability by investing in its digital platform, the company’s strategy progressively shifted, first to managing many independent freelance photographers and eventually to an automated system for a global network of freelance photographers (Hoffman & Yeh, 2018), all driven by Airbnb’s increasing opportunity costs of allocating resources to the complementary activity of photography. As this example illustrates, even highly scalable digital firms often begin with less scalable non-digital resources, and this transition over time reshapes the firm’s integration choices as it moves from the non-scalable region (left side) to the scalable region (right side) of Figure 2.

3.4 | Resource attributes and hyperspecialization

Our second result outlines the relationship between the scalability of resources and the critical α ($= \alpha^*$). It provides a rationale for referring to outsourcing in digital firms as *hyperspecialization*, which we define to mean the firm’s propensity to specialize and not integrate even when the cost of “using the market” (captured here by the value shared with the complementor) is quite high. Thus, hyperspecialization is very different from ordinary specialization that arises when the resource bundle is non-scalable, which can be sustained only when the cost of using the market is low. We denote as critical α ($= \alpha^*$) the maximum value share the focal firm is willing to forego by outsourcing, i.e., the maximum cost of using the market tolerated by the firm before it integrates. Proposition 2 characterizes its relationship to scalability.

Proposition 2. (Scalability and hyperspecialization): *The critical α is increasing in scalability for digital firms and decreasing in scalability for industrial firms.* (Proof in Online Appendix.)

Proposition 2 demonstrates how the effect of scalability on the decision to integrate differs in industrial ($\sigma < 1$) and digital ($\sigma > 1$) firms. Consistent with our understanding of industrial firms (Chandler, 1977, 1990), higher scalability improves the attractiveness of full integration relative to outsourcing when $\sigma < 1$, causing the critical α ($= \alpha^*$) line on the left side of Figure 2 to slope downward. This happens because higher scalability creates “excess capacity” and facilitates the leveraging of the firm’s resource bundle in multiple applications: the more scalable it is, the easier it is for the firm to scale operations into both activities. In digital firms ($\sigma > 1$), however,

¹²Although transaction cost and coordination cost logics can also be used to suggest these full stack versus no stack arguments, they would not suggest a choice between the two polar alternatives, as our theory does.

scalability has the opposite effect. In this region, even if resources are perfectly fungible, the firm can achieve greater productivity by focusing on just one activity because the firm avoids breaking up its resource bundle, and thus sustains its growth along the exponential trajectory. Moreover, because the firm can offset the value foregone in using the market by significantly increasing output volume, the more scalable the firm's resource bundle, the lower the minimum share of value the firm is willing to accept in order to specialize rather than integrate. Put differently, for digital firms with highly scalable resource bundles the impetus to specialize can be incredibly strong, so much so that it overpowers typical drivers of integration—these firms are simply propelled toward *hyperspecialization*. In Figure 2, this effect is reflected in the upward sloping critical α ($= \alpha^*$) line for digital firms ($\sigma > 1$). Proposition 2 finds real-world corroboration in the digital businesses of firms like Airbnb, in which high scalability goes hand in hand with a relatively small share of value (only 3% commissions) but for a huge number of transactions, whereas traditional hotel chains like Hilton and Marriott with less scalable resource bundles remain partially integrated and yet receive 8–12% commissions from third parties who, like Airbnb hosts, own and manage vacation properties (McNew, 2018).

Next, we examine the fungibility of the firm's resource bundle and its relationship with integration choices. The prior RBV literature has heavily focused on fungibility as a resource attribute when examining corporate scope (Levinthal & Wu, 2010; Montgomery & Wernerfelt, 1988). To understand its role more completely, we examine how fungibility works in industrial and digital contexts, and especially how it compares to scalability.

Proposition 3. (Fungibility and hyperspecialization): *For both industrial and digital firms, the critical α is decreasing in the fungibility of the firm's resource bundle.*
 (Proof in Online Appendix.)

Proposition 3 demonstrates that, unlike with scalability, the critical α ($= \alpha^*$) decreases with fungibility for both scalable and non-scalable resources (see Figure 3). This is because fungibility boosts the productivity of the firm's resources in the complementary activity, thus favoring the sourcing regime that uses resources in both activities, that is, integration. Propositions 2 and 3 together show an interesting contrast between industrial and digital environments. In traditional industrial firms, the two resource attributes commonly associated with the “value” of a resource, scalability and fungibility (Schmidt & Keil, 2013), have the same implications for the firm: both facilitate leveraging the firm's resources in multiple applications, allowing the firm to take its competitive advantage from one stage to create value in others. In digital firms, however, scalability and fungibility have *opposite* implications. While higher fungibility renders integration more attractive, higher scalability favors specialization even when the cost of using the market is substantial.

The evolution of the vertical scope of Netflix helps to illustrate our theory. Netflix started in 1997 with a DVD rental-by-mail business model, and relentlessly invested in digitalization and software-driven automation from the outset. The company's datasets, back-end software, and recommendation algorithms were key digital resources that (along with digital distribution through a platform) likely put the company in the scalable region of our model ($\sigma > 1$).¹³

¹³Included in these scaling advantages were the low marginal costs of providing a wide DVD assortment in an internet retailing model, relative to a bricks and mortar movie rental company like Blockbuster. Netflix's software, including its algorithms and customer DVD queue feature, also enabled efficient management of its inventory and distribution, which further improved with scale. On the demand side, the scalability of Netflix resources was further enhanced by highly elastic demand due to the rapid adoption of DVD players among U.S. households.

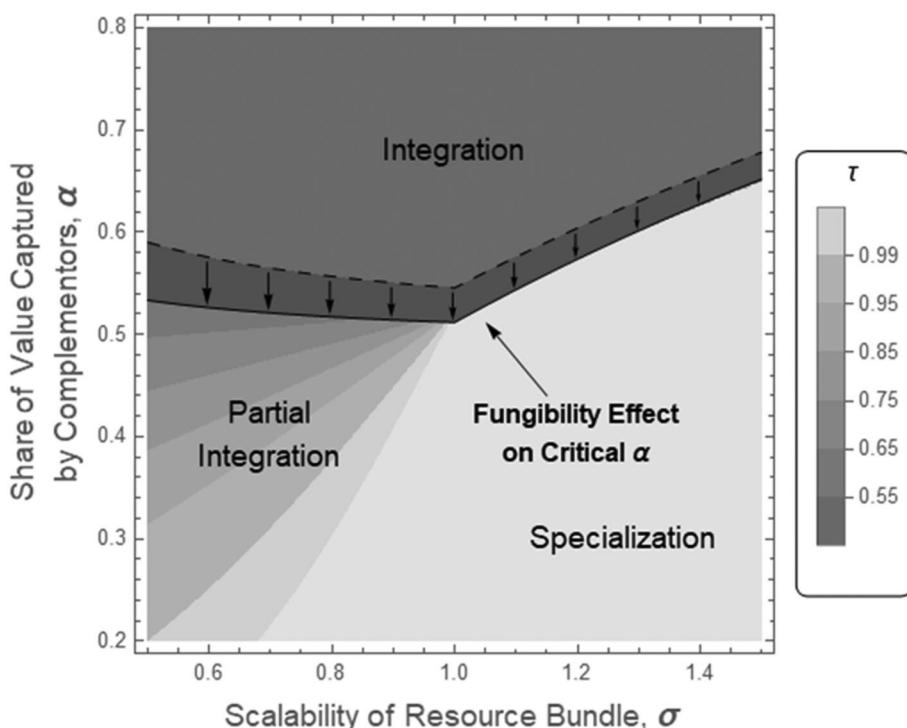


FIGURE 3 Fungibility's effect on critical α . The figure illustrates the downward shift in the critical α ($= \alpha^*$) line with an increase in fungibility. Fungibility also changes the degree of outsourcing undertaken in the partial integration region, which is not illustrated in the figure (and is difficult to discern in a single graph). The boundary between the partial integration region and the specialization region remains unchanged

Consistent with our model, Netflix remained focused on taking advantage of the growth opportunities in its DVD rentals business, and not integrating into upstream content production. Indeed, “Singular focus” were the first two words of the company’s second-ever annual report, in 2003. In its DVD rental business, Netflix bought DVDs and—under the first sale doctrine of copyright law—simply rented them out without further payments to the movie studios. In the late-2000s, as it switched to online streaming, Netflix faced a significant increase in its content costs. Without the first sale doctrine to rely on, Netflix needed to take licenses for content from studios, who began demanding higher prices, thus increasing the value share (α) that Netflix was forced to forego. Netflix’s content costs reportedly went from \$229 million in 2010 to more than \$1 billion in 2011 (Kafka, 2010, 2011a). In response, Netflix integrated into commissioning and owning original content. Ted Sarandos, Netflix’s chief content officer at the time, explained (Kafka, 2011b): “[W]ould [I] prefer to license previous seasons of HBO, Showtime, Starz shows? Sure. And if those shows are not going to be made widely available in decent [price] windows, then my other alternative would be to compete with those guys for those shows.” Netflix’s fungible digital resources, which provided insights into customer preferences, facilitated its vertical integration by providing the company with an information advantage in content development. When acquiring the rights to *House of Cards*, Netflix outbid networks including HBO, Showtime, and AMC, and even made a two-season commitment (instead of the typical commitment to a pilot), based on customer insights from its proprietary data. Since then, original content as a share of new content in Netflix’s portfolio has steadily increased (except when content production was

interrupted by the COVID-19 pandemic). Thus, Netflix has been integrating in the direction predicted by our model given the scalability and fungibility of its resource bundle, as well as changes in the value share captured by content owners.¹⁴

3.5 | Hyperscaling and hyperspecialization

When the resource bundle is scalable ($\sigma > 1$), scalability begets specialization, which begets even more scale. This occurs because the focused allocation of resources to a given activity can increase output exponentially, which incentivizes firms to accumulate larger resource bundles and grow their resource base exponentially, an outcome we refer to as *hyperscaling*. Therefore, hyperscaling is not just a function of scalability but also influenced by hyperspecialization. To show this, we move back to the first stage of the model, turning our attention to the firm's optimal investment in its resource bundle, which depends on the sourcing regime chosen. Formally, the firm's optimal resource investment satisfies the first-order condition $\frac{\partial \pi(\tau_j, r)}{\partial r} = \frac{\partial C(r)}{\partial r}$ for $j \in \{S, C, I\}$, which is reached when the cost of additional investment equals the (marginal) productivity of the firm's resource bundle. Similar to Chandler's (1990) focus on assets, we focus on the size of the firm's resource bundle (which in turn also determines the size of revenues and profits in our model) as the representative measure of firm size.¹⁵ As shown in the following proposition, hyperscaling in firm size is only partly explained by scalability; it is also explained by hyperspecialization.

Proposition 4. (Hyperscaling and hyperspecialization): *Specialized digital firms are larger than integrated digital firms as well as fully or partly integrated industrial firms; however, the relative scale difference between specialized and integrated digital firms decreases with the fungibility of the integrated firms' resource bundle.* (Proof in Online Appendix.)

This proposition highlights that, unlike industrial firms in which—to borrow from Chandler's famous book title (Chandler, 1990)—scale and scope went hand in hand, digital (scalable) firms have a propensity toward greater scale that is instead supported by their specialization. Our comparison in Proposition 4 pits specialized firms against both integrated firms with the same scalable resource bundles, and specialized or integrated firms with non-scalable resource bundles. If two firms utilize identical and scalable ($\sigma > 1$) resource bundles, but one is specialized and the other is not, the specialized firm is larger. In contrast, for non-scalable ($\sigma < 1$) resource bundles, the integrated firm is larger than the specialized one, consistent with Chandler's thesis (Chandler, 1990). These results suggest that digital firms not only have a propensity for hyperscaling, but also that hyperspecialization and hyperscaling are mutually reinforcing. Figure 4 illustrates the relationship between scalability of the resource bundle and the scale of the firm when firms fully integrate or fully specialize. When resources are non-scalable, the solid line

¹⁴Original programming as a share of new content in Netflix's portfolio has been increasing, but it has not gone to 100% as our model would predict. One reason Netflix still licenses content from media companies is that the cost of using the market did not increase as much in some cases—for example, legacy content, or content from media companies that need a good digital distribution partner. Also, Netflix often licenses content through longer term contracts, and the firm's integration is increasing as these contracts expire (e.g., with the Walt Disney Company in 2019).

¹⁵Because the extent of labor employed in production can be lower when digital resources are used, our results cannot be extended to make predictions about employment.

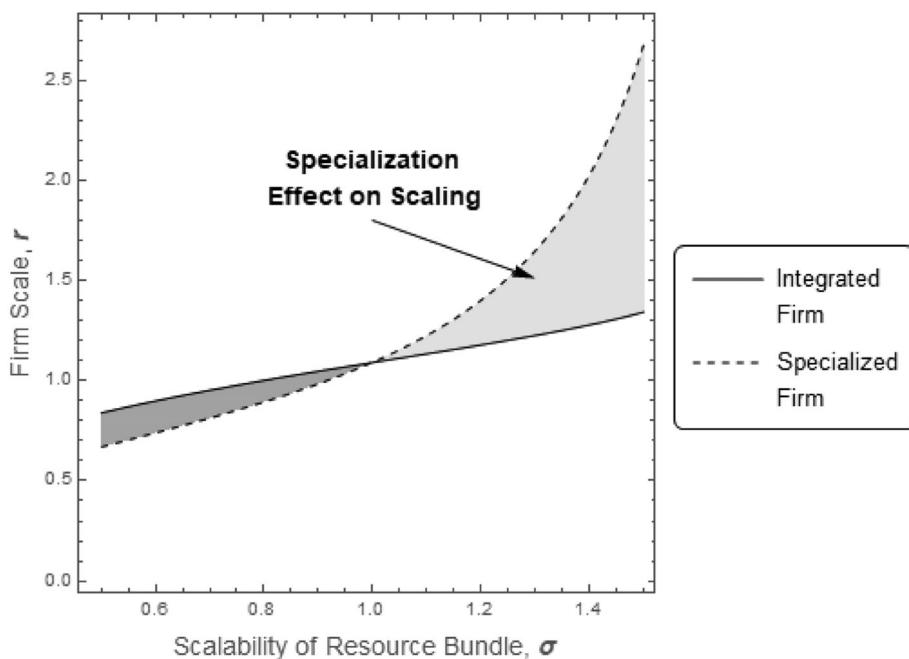


FIGURE 4 Specialization's effect on scaling. The figure shows the relationship between the scalability of the firm's resource bundle σ and the scale (size) of the firm (captured by the optimal size of the firm's resource bundle r) at the two ends of the sourcing continuum, integration and specialization. When the firm's resource bundle is non-scalable (left side, $\sigma < 1$, as in an archetypal industrial firm) the solid line corresponding to firm size under integration dominates the dashed one corresponding to firm size under specialization. This changes when resources are scalable (right side, $\sigma > 1$, as in an archetypal digital firm), with the difference between the two lines corresponding to an enhancement of the resource scaling effect to one of hyperscaling—that is, the optimal scale of the firm becomes increasingly larger due to the complementarity between scalability and specialization

corresponding to integration dominates the dashed one corresponding to specialization. This pattern changes when resources are scalable, with the difference between the two lines corresponding to a specialization effect—that is, extra scaling due to the complementarity between scalable resource bundles and specialization.¹⁶ These findings reveal an important strategic trade-off inherent in the allocation of scalable resource bundles in digital firms: when they shrink their range of activities, they are able to grow more in size. When resources are non-scalable, however, we revert to the organizational landscape of traditional industries, with integrated firms growing larger than their specialized counterparts.

Another feature illustrated in Figure 4 is that the specialization effect increases more than proportionally with scalability for digital firms. If two firms both have scalable resource bundles, then the firm with a more highly scalable resource bundle is not only larger, but the

¹⁶As an interesting subtlety, an increase in the scalability exponent increases scale (as shown in Figure 4) only if the optimal r exceeds a certain threshold. This cut-off value for r identifies a point whose surpassing can lead to sustained growth and can be interpreted as a tipping point, a critical mass or minimum resource base that must be attained in order to trigger hyperscaling. This is often the case with digital goods, which typically require a substantial upfront investment, but then scale at almost no cost. As noted by Arthur (1996, p. 103) “the first disk of Windows to go out the door cost Microsoft \$50 million; the second and subsequent disks cost \$3.”

additional boost to firm size from specialization is disproportionately higher for this firm. Thus, digital firms that have very highly scalable resource bundles experience an outsized specialization effect; that is, they become extremely large when they specialize. Further, Proposition 4 compares two integrated firms with resource bundles that have different levels of fungibility with a specialized one, positing that the integrated firm with a more fungible resource bundle has a smaller relative size penalty from integration. This result follows from the fact that the integrated firm with more fungible resources is more efficient *on average* in its secondary activity, which justifies investment in greater scale (more resources). Thus, our model predicts that the size penalty associated with integration may be smaller for firms whose vertical integration is facilitated by fungible resources (like the Netflix example above). That said, the scaling penalty associated with integration for digital (scalable) firms does not converge to zero at any level of fungibility.

Proposition 4 is consistent with the observation that the largest companies by market capitalization in the 1950s and 1960s, such as GM, AT&T, DuPont, and Standard Oil, are all classic examples of vertically integrated corporations, which are featured in Chandler's historical accounts. By contrast, many of the largest companies by market capitalization today are successful as specialized super-intermediaries, whose business models involve mediating between complementors in ecosystems. To date, firms like Google, Facebook, and Microsoft "continue to derive the bulk of their revenues and, for the most part, profits from the [core] businesses which made them into trillion- or near-trillion-dollar companies" (The Economist, 2021b). These firms have surpassed their industrial counterparts in terms of output volumes, market capitalization, and asset values, despite being more specialized.

Proposition 4 is also consistent with changes over time in the semiconductor industry. The impact of Moore's (First) Law on the semiconductor industry—exponentially increasing miniaturization and reducing costs per unit—is well known. However, the industry has also been shaped by "Moore's Second Law," which asserts that the cost of an integrated circuit (IC) fabrication plant (fab) will double every 4 years (Moore, 1995) and has sharply increased the returns to scale in IC fabrication over time.¹⁷ In turn, these changes have led to more consolidation among firms (scale) and the emergence of dedicated fabs (specialization) as the dominant type of IC firm.¹⁸ This example illustrates the applicability of our theory (in some cases) to industries beyond digital businesses, so long as scalable resource bundles are present.

3.6 | Endogenizing the firm's value share

Our main model sets the fraction of value shared by the focal firm with complementors (α) to be exogenous. We now endogenize α by modeling two key aspects of the complementor market: the number of complementors (Brandenburger & Nalebuff, 1996; Porter, 1980) and their productivity (Barney, 1999; Jacobides & Winter, 2005). We model the complementor market as consisting of N identical firms. Each generates outputs $Q_{ja}(\tau_j r_j)$ and $Q_{jb}((1-\tau_j)r_j)$, where Q_{ja} and Q_{jb} are scaling laws with scaling exponent $\sigma_j > 0$. Complementor productivity is analogous

¹⁷In addition to supply-side scale economies, IC fabs also benefit from demand-side network externalities due to the availability of complements in the form of validated semiconductor design modules that meet their manufacturing design rules (Linden & Somaya, 2003).

¹⁸Prior research has primarily emphasized the role of transaction costs and their decline through modularization in explaining the vertical disintegration of the industry (Linden & Somaya, 2003; Macher & Mowery, 2004); however, declining transaction costs alone would not explain increasing consolidation among IC fabrication firms.

to firm i 's, with complementors having a greater baseline productivity in b , the complementary activity, such that $\frac{\partial VQ_{jb}(1)}{\partial r_j} = \varphi_j(1) \geq \frac{\partial VQ_{ja}(1)}{\partial r_j} = \varsigma_j(1) > 0$. A complementor's cost of acquiring resources is given by $C(r)$, defined analogously to firm i 's cost function. For added realism we also incorporate transaction costs, measured as the value (θ) lost in coordinating market exchanges (Coase, 1937; Williamson, 1975), such that whenever the component b is outsourced the complementors only receive $\theta\alpha$, with $\theta \in (0, 1)$.¹⁹ The role of transaction costs is qualitatively similar to that of sharing value (α); they increase the costs of using the market but do not fundamentally alter the insights derived from our model.

Our extension shows that the value share captured by complementors increases when the number of complementors shrinks. The intuition behind this result is simple: when competition among complementors decreases, their relative bargaining power with respect to the focal firm increases, resulting in the remaining complementors receiving a larger share of the value pie (formally, $\frac{\partial \alpha}{\partial (-N)} > 0$). This confirms a key finding of the IO-based strategy literature (e.g., Brandenburger & Nalebuff, 1996; Porter, 1980), which suggests that outsourcing tends to be more favorable in more competitive complementor markets (Hecker & Kretschmer, 2010). This baseline result serves as a “reality check” for our model.

Perhaps more counterintuitive is the result that the value share captured by complementors increases when their productivity decreases (formally, $\frac{\partial \alpha}{\partial (-\varphi_j(1))} > 0$). Metaphorically speaking, this result indicates that the focal firm is willing to “compensate” the complementors in a manner proportional to their “incompetence.” This happens because the focal firm benefits from the scale of complementors' output. Therefore, it is willing to concede a larger share to less-productive complementors to offset their inferior productivity and incentivize them to invest in capacity. In industrial firms, however, any increase in α due to a decrease in complementor productivity also induces the focal firm to expand its vertical scope through partial integration (as illustrated on the left-hand side of Figure 2). This effect is stronger the more fungible and productive the focal firm's resources; thus less productive complementor resources generate a dynamic akin to the canonical prediction of the RBV literature that firms expand their vertical scope if their resources are fungible and more productive than those of complementors. For digital firms, however, resource superiority does not lead to more integration unless α reaches the critical α ($= \alpha^*$) line (as illustrated on the right-hand side of Figure 2); instead, the focal firm foregoes increasing shares of the value created as a way to incentivize supply of the complementary products or services.²⁰

Consider for instance the case of Shopify, an e-commerce platform that connects merchants operating online stores with developers creating specialized apps (Dushnitsky & Stroube, 2021). Shopify's ecosystem of third-party app developers captures a much larger share of revenues than Shopify itself. In 2019, Shopify's revenue was around \$1.5 billion, whereas the partner ecosystem generated more than \$6.9 billion (Holmes, 2020). As Tobias Lütke (2020), co-founder and CEO of Shopify, explains: “What we did to get the platform off the ground is to basically leave all the economics for Shopify on the table and give it to the

¹⁹Although we assume that transaction costs are directly borne by complementors, they are indirectly borne by the focal firm i since α is increasing in $-\theta$ (see proof in the Online Appendix), and would have a similar effect if they were borne by the focal firm or split between the focal firm and the complementor(s).

²⁰This happens despite the absence of administrative and bureaucratic costs, which would tilt the firm calculus toward outsourcing (Coase, 1937; Williamson, 1975). Indeed, the existence of increasing returns at the technological level is equivalent to assuming that administrative and bureaucratic costs are not substantial (Coase, 1937).

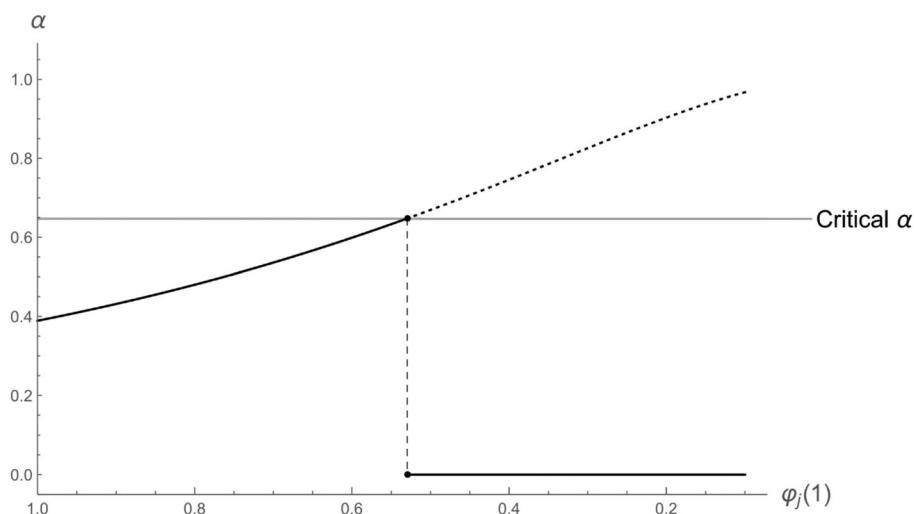


FIGURE 5 Complementor productivity and value capture. The figure illustrates how the value share that a digital firm is willing to forego (α) changes as the relative capability of the complementors ($\varphi_j(1)$) decreases. The focal firm's resources are set to be perfectly fungible, with productivity in both applications normalized to 1. There is a region on the horizontal axis (from 1 to approximately 0.54) in which the focal firm is more competent than the complementors, but does not integrate. As the complementors become less productive, not only does the focal firm continue to outsource to them but the value share it is willing to forego also increases. Once this value share reaches the critical α ($= \alpha^*$), the firm finds it optimal to integrate and no longer shares any value with the complementors. We note that a complementor cannot profitably undercut the competition. If a complementor offered a lower α to the focal firm, such complementor would not capture the whole market. On the contrary, it would scale less because a lower α would reduce the returns of its resources and its incentives to invest in capacity

third-party app developers. [...] I think Bill Gates said this, I think it's almost called the 'Gates line'—You are not a platform until the people who are building on you make more money than you do. [...] It's hard to do, because you are leaving a lot of economics that you could easily take for yourself on the table—or actually, you are investing it into your own future by giving it to other people."

These tradeoffs are illustrated in Figure 5. Figure 5 reveals how the complementor's value share changes with its relative productivity (on the horizontal axis). While a decrease in complementor productivity initially leads to a continuous increase in its value share, once the realized α reaches the critical α ($= \alpha^*$) line, outsourcing becomes too costly compared to internal production and the focal firm switches to integration.²¹ Thus, the sourcing regime exhibits a combination of continuous and discontinuous change as a result of two conceptual drivers. One driver is the continuous increase in value shared within the outsourcing regime as complementor productivity decreases, whereas the other is the discontinuous transition to integration when sufficiently low levels of complementor productivity force the focal firm to cut out complementors and stop outsourcing. We summarize these insights in the following proposition.

²¹For completeness, note that integration also occurs when the realized α falls below the complementors' critical α , that is, the minimum share of the value created that the complementors will accept to collaborate with the focal firm.

Proposition 5. (Realized α): In digital firms, the value share captured by complementors, α , increases when complementor productivity decreases up to the critical threshold α^* . If complementor productivity declines further, digital firms will integrate and make the input in-house. (Proof in Online Appendix.)

One implication of Proposition 5 emerges from examining different parts of the scalable region in Figure 2. For digital firms, integration is more likely to be triggered when margins are thinner (the value share foregone is larger) because a sudden switch to integration occurs only in the proximity of the critical α ($= \alpha^*$) line, which in turn is increasing in scalability. Consider Amazon's different strategies for its two key businesses, e-commerce (retail) and cloud (AWS). Amazon's margins are thinner in e-commerce, consistent with its retailing complementors being arguably less competent. Amazon has demonstrated an appetite for internalizing adjacent activities in e-commerce, such as logistics (truckling and air freight investments to disintermediate UPS and FedEx) and product development (displacing third-party sellers by introducing its own private labels) (Schreiber, 2016; Zhu & Liu, 2018). By comparison, Amazon has largely refrained from integrating into complementors' businesses in cloud computing (Hoffman & Yeh, 2018). The relationship between AWS and Twilio, a business-to-business cloud communication platform that hosts its services on AWS, is a case in point. As one commentator observed around the time of Twilio's IPO (Seward, 2016): "Twilio customers, in other words, are outsourcing messaging to Twilio, which in turn outsources to Amazon. [...] You could argue this is a precarious position to be in because Amazon could always decide to make messaging a feature of AWS." However, "the two companies are better described as partners rather than competitors" (Sun, 2017), illustrating how superior and fungible resources might not be a sufficient motivation for integration in digital firms.

3.7 | Incorporating the demand environment

Thus far, we assumed perfectly elastic demand, which lets the firm sell unlimited quantities of the final product $a \wedge b$ for a fixed price in the market, and has thus largely isolated the model from demand side features. However, a firm's impetus to specialize and grow ultimately depends on a combination of supply and demand forces (Helfat & Eisenhardt, 2004; Makadok, 1999; Penrose, 1959; Sakhartov & Folta, 2015). For example, decreasing returns on the demand side from a downward-sloping demand curve can limit the benefits of selling larger output volumes. We now incorporate a richer demand environment in our model using the Dixit-Stiglitz system of monopolistic competition (Alfaro, Antràs, Chor, & Conconi, 2019; Dixit & Stiglitz, 1977; Helpman, 1985). In this system, the (inverse) demand function follows the scaling law $V = A Q_{ia}^{(1-\rho)/\rho}$, where parameter $\rho > 0$ determines the demand-side returns to scale.²² A downward-sloping demand corresponds to $\rho > 1$, perfectly elastic demand to $\rho = 1$, and

²²In the Dixit-Stiglitz system, the (inverse) demand function corresponds to the scaling law $V = A Q_{ia}^{(1-\rho)/\rho}$, where V is the value for the final consumer and $A > 0$ is a given term for the firm. Assuming the focal firm captures the full surplus from customers (as in the main model), the resource allocation problem described in Equation (1) generalizes to the maximization of $\pi(\tau, r) = (1 - \alpha) A \rho Q_{ia}(\tau r)^{1/\rho} + \alpha A \rho Q_{ib}((1 - \tau)r)^{1/\rho}$, wherein $Q_{ia}(\tau r)^{1/\rho}$ and $Q_{ib}((1 - \tau)r)^{1/\rho}$ follow scaling laws of degree σ/ρ . The model extension leading to Proposition 5 is robust to the Dixit-Stiglitz demand system of monopolistic competition, in which for every product produced in-house, each complementor faces a demand curve corresponding to $V_j = A_j Q_{jb}^{(1-\rho_j)/\rho_j}$, with $A_j, \rho_j > 0$. If A_j is exogenous, the Dixit-Stiglitz demand system will lead to results that are qualitatively similar to Proposition 5. Alternatively, if A_j is endogenized as a decreasing function of the number of firms producing differentiated goods, both the complementors and the focal firm would be more inclined to cooperate because the monopolistic competition outcome would reduce the baseline level of demand for their products.

increasing returns in demand to $\rho < 1$. This demand environment can augment our model to yield a profit equation similar to (1), wherein the revenues from the two complementary activities follow scaling laws of degree σ/ρ rather than σ . All our results generalize to the range of demand conditions described above by simply substituting σ/ρ as the scalability parameter. Thus, hyperspecialization and hyperscaling will occur only if $\sigma/\rho > 1$. Demand-side returns to scale can therefore be neutral, boost, or even negate the supply-side increasing returns to scale of digital firms. For digital firms with highly elastic demand ($\rho \rightarrow 1$) enabled by the ease of global digital distribution, overall scalability will be primarily determined by supply-side returns to scale (similar to our model above). However, for digital firms that either experience demand-side increasing returns (due to network effects) or conversely face decreasing returns due to downward-sloping demand curves, the demand side will complement or undermine, respectively, supply-side returns to scale of digital firms' resource bundles.

4 | DISCUSSION

This article describes a theory of digital firms that explains how the scalability of their resource bundles raises the opportunity cost of integration, and thus leads to specialization. We also incorporate the fungibility and accumulation costs of resources and derive a set of additional results: scalability will induce specialized firms to focus on volume rather than on capturing a larger share of the value created with complementors, to out-scale integrated firms, and to switch discontinuously from specialization to integration, whereas fungibility will mitigate some of these effects. We now highlight our contributions to the theory of the firm, the RBV theory of (vertical) integration, and future work on the RBV.

4.1 | A theory of the digital firm

Transaction costs have long been the centerpiece of theorizing on vertical integration, also called “the theory of the firm” (Coase, 1937; Williamson, 1975, 1999). A key contribution of our paper is to offer a theory of integration choices in digital firms that relies not on transaction costs but on the properties of digital firms' resource bundles. By highlighting how digitalization affects scalability in both production and demand, our theory of the digital firm differs from prior explanations for how information technologies have impacted economic organization. In this conventional view, scholars have argued that information technology and the internet would reduce transaction costs and thus lead to greater specialization and vertical “disintegration” into smaller firms (Brynjolfsson, Malone, Gurbaxani, & Kambil, 1994). Similarly, business historians have proposed a “post-Chandlerian” form of economic organization in high technology industries arising from greater modularity and more relational contracting (Lamoreaux, Raff, & Temin, 2004; Langlois, 2003, 2004), resulting in less vertical integration and smaller firm sizes. Thus, these transaction costs logics typically predict that digital firms would have reduced scope (specialization) and reduced scale (or at least not hyperscaling). By drawing on the RBV and especially the construct of *scalability*, our theory predicts that digital firms are likely to be *both* highly specialized (hyperspecialization) and very large (hyperscaling), which appears consistent with an emerging class of digital firms (Adner et al., 2019; Hoffman & Yeh, 2018; Parker, Van Alstyne, & Choudary, 2016).

Our theorizing further suggests that a reduction in transaction costs might be neither necessary nor sufficient to explain the rise of ecosystems as a dominant form of economic organization in the digital economy. In an extension in the Online Appendix, we show that transaction costs play a very similar role to resource productivity differences, simply as another cost of “using the market” (Chu & Wu, 2021; Mahoney & Qian, 2013), which can be offset by the advantages of scaling. When digital firms have scalable resource bundles (arising in part due to demand-side network effects in platforms), our theory suggests that they may be driven toward greater specialization even in the presence of significant transaction costs. This is consistent with the observation that in some sectors of the economy ecosystems have arisen not because of lower transaction costs but to manage higher transaction and coordination costs (Baldwin & Clark, 2008; Dosi, Gambardella, Grazzi, & Orsenigo, 2008). Similarly, scholars have noted that ecosystems bind complementors in relationships of significant interdependence and often entail more, not less, interorganizational interactions than arm’s length relationships (Ganco, Kapoor, & Lee, 2020; Postrel, 2009). Thus, our theory also provides insights into the underlying mechanics of platforms and ecosystems, including the strategies of firms that create digital platforms with very thin commissions.

Despite transaction (and bureaucratic) costs not being a focus of our theory, they nonetheless play an important role in setting its boundary conditions, and relaxing those conditions is a fruitful direction for future research. For example, two implicit boundary conditions to our theory are that: (a) there are significant frictions in the market for services arising from firms’ resource bundles (Penrose, 1959) such that the firms must employ resources internally to produce outputs from them, and (b) there are significant frictions in the market for corporate control that prevent firms from simply acquiring others and thus overcoming the limits to integration and growth imposed by their resource accumulation costs. When these market frictions are instead low or when the benefits of overcoming them are significant, our theory offers additional implications that can be explored in future work (e.g., a motivation for mergers and acquisitions in digital firms). Moreover, although we ignore both the transaction costs of using the market and the bureaucratic costs of firm size in our exposition, our theory does suggest that if scalability is sufficiently high digital firms will find it optimal to specialize and scale despite these costs. However, we acknowledge that digital technologies may not only affect the attributes of resource bundles, but also transaction costs in markets and bureaucratic costs within firms (Kretschmer & Khashabi, 2020), which may combine with resource characteristics in shaping economic organization. Following a tradition of integrating governance and resource-based perspectives (Argyres & Zenger, 2012; Kang et al., 2009; Mayer et al., 2012; Wan & Wu, 2017), future work could study the interplay of scalability and transaction costs to develop a richer understanding of digital firms.

In explaining how digitalization affects the scalability of firms’ resource bundles, and thus induces hyperspecialization and hyperscaling, we have thus far treated the roles of supply-side and demand-side returns to scale as more or less equal and symmetric. However, this is not always so, and we graphically illustrate these considerations in the Online Appendix. First, it is important to note that if returns to scale increase on one side (supply or demand) its impact on overall scalability is enhanced by corresponding increases on the other side. Put differently, returns to scale in supply and demand are complementary in their effects on scalability. All else equal, digitalization is likely to make firms more scalable if it increases both supply-side and demand-side returns to scale. Second, enhancing returns to scale on the side that poses a greater constraint tends to have a larger impact on scalability. For example, if firms already have constant or increasing supply-side returns to scale, but demand is inelastic, then digitalization’s

impacts on demand-side returns to scale are more consequential for scalability than equivalent increases on the supply-side. Finally, it is also important to consider the degree to which digitization affects supply-side and demand-side returns to scale. For example, if returns to scale on the supply side increase substantially due to a highly digital resource bundle, its impacts on scalability might be very large even if the greater constraint is on demand side, and vice versa for substantial increases in demand-side returns to scale (e.g., due to platform-driven network effects).

4.2 | The RBV theory of vertical integration

The RBV has long held that relative resource strengths are an important determinant of vertical integration (Argyres, 1996; Madhok, 1996, 2002), a prediction that also comports with managerial experience in traditional industrial firms (Barney, 1999). However, recent advances in the RBV that explicitly incorporate resource attributes (Levinthal & Wu, 2010; Schmidt & Keil, 2013) and the emergence of digital firms (Adner et al., 2019; Siebel, 2019) present opportunities to examine and extend this theory. We provide such an extension by focusing on the opportunity costs of scalable resource bundles that are more productive when intensively deployed within a focal activity, rather than being spread across multiple activities in pursuit of integration. We further examine the role of resource fungibility, account for resource accumulation costs and demand conditions, and incorporate competition with the firm's complementor(s) in characterizing a set of implications that arise from our theory.

Our research contributes to the RBV theory of vertical integration by showing that the opportunity cost of scaling within a specialized activity can sometimes be strong enough to outweigh the benefits of leveraging superior resources across multiple value-adding activities. Our theory thus extends the classic RBV argument that superior resources will trigger integration (Argyres, 1996; Madhok, 1996, 2002) by showing that this is not a sufficient condition. Indeed, our results highlight that a firm may outsource to less efficient complementors and even share more value with them, but only if it has a highly scalable resource bundle that can be scaled in its focal activity. In pursuing a parsimonious and logically consistent model, however, we have conveniently assumed that the firm's resource bundle is equally usable and scalable in both its main and complementary activities. Thus, a potentially valuable extension to our theory would allow for the composition and scalability of the resource bundle to differ between the main and complementary activities, which could produce additional insights about vertical integration under different configurations of scalable and non-scalable resources.

Our work also adds to the RBV by developing a formal approach that integrates resource attributes into economic models of firms and markets. Prior research has modeled resource attributes such as fungibility, scale adjustment costs, and redeployability (Knudsen et al., 2014; Levinthal & Wu, 2010; Sakhartov & Folta, 2014, 2015; Schmidt & Keil, 2013; Wu, 2013) in addressing core questions for strategic management. We add to this work by characterizing the construct of scalability of the resource bundle and studying its implications for firm boundaries and scope. *Inter alia*, we define and explain the meaning of scalability, and how digital firms may become more scalable through a combination of digital resources and digital outputs. Last but not least, the concept of resource-based hyperspecialization and hyperscaling we advance is different from Adam Smith's "pin factory" tradition of specialization based on learning, which

is history-dependent and arguably yields a smaller impetus to scale; however, they do share some commonalities in the importance of demand considerations and potential correlation between firm size and specialization (Becker & Murphy, 1992; Stigler, 1951).

4.3 | Future directions for RBV research

Our theory and modeling framework can be a platform for future work on resource attributes and their effects on firm-level outcomes. The framework is parsimonious, but captures important resource attributes such as scalability, fungibility, and resource adjustment costs (Knudsen et al., 2014; Levinthal & Wu, 2010; Schmidt & Keil, 2013) that matter to firm decision-making on resources and their impacts. Our model has also yielded a number of nuanced and counter-intuitive results. Despite these attractive features, our framework also has limitations. While we have examined competition between the focal firm and its complementors, other interesting scenarios remain unexplored. For example, the final product may face competition from other firms, who may also compete to attract away the complementor(s) and thus deny the scaling advantages the focal firm may access through specialization. Hoffman and Yeh (2018) suggest that competitive pressures will only heighten the pressures to scale rapidly, and to hyper-specialize in support of growth. Nonetheless, consideration of such competition, including the case of complementors who can multi-home, provides a rich landscape within which to extend our theory. Further, potential competition through market entry can be affected by both traditional isolating mechanisms (Barney, 1991; Peteraf, 1993; Rumelt, 1984; Wernerfelt, 1984) and by preemption through investments in scalable resource bundles (Wibbens, 2021). Future work can also model the dynamic competition between firms and complementors when relative resource advantages are not stable but depend on prior integration choices (Argyres & Zenger, 2012; Mayer et al., 2012), and both types of firms can compete in the final product market (Wan & Wu, 2017) or the focal firm can compete with the complementors in its main market (Zhu & Liu, 2018).

Because it builds on number of simplifying assumptions, our framework also leaves out several factors of theoretical importance. Among these, future work can examine the elemental drivers of resource attributes such as scalability, which we treat as exogenously determined. Endogenizing scalability can be of interest in the study of early stage ventures where firms can anticipate the opportunities to scale and make resource investments accordingly, which adds further nuance to our understanding of how resources evolve (Helfat & Peteraf, 2003). In their book *Blitzscaling*, Silicon Valley VC investor Reid Hoffman and his coauthor Chris Yeh posit two “growth limiters” firms must anticipate to scale rapidly: the lack of product/market fit and the lack of operational scalability, and suggest that “both can still kill your company,” so that “the wisest innovators design operational scalability into their theories” (Hoffman & Yeh, 2018, pp. 75–76). Future extensions can examine if and when early-stage process and operational improvements are key for firms to achieve a “first-scaler advantage,” which accrues not to the first firm that enters a market but the first firm that serves that market at scale (Hoffman & Yeh, 2018; Levie & London, 2018).

At the other end of firm growth, even highly scalable businesses will eventually saturate their market, and demand-side decreasing returns will kick in. Put simply, the firm runs out of additional customers to serve, even at increasingly low prices. Similarly, larger firms will inevitably incur greater bureaucratic and administrative costs (Williamson, 1975, 1999), which also impose a penalty on their size. However, these limits on specialization and scale need to be

juxtaposed against the magnitude and range of the resource bundle's supply-side returns to scale; if they are substantial, digital firms may sustain specialization and continue to scale for a long time. Understanding these issues would add nuance to the classic literature on the link between vertical integration and industry life cycles (Becker & Murphy, 1992; Stigler, 1951), which suggests that integration is more likely in the early and later stages of the industry life cycle. It is unclear whether digital firms face the same forces and evolutionary patterns. Future work could add precision to the expected changes in resource scalability and fungibility as digital firms grow, which can yield insights into these firms' changing scale, scope, and boundaries over time. Moreover, we focus on the general case where only the net scalability of the resource bundle matters, arising from a combination of scale free resources and favorable demand conditions, which leaves open a number of combinations of resource and demand conditions. Resource and demand conditions can depend on the firm's choices to varying degrees and can vary over its life cycle. These nuances could generate a complex set of strategies and outcomes that can further enrich the theory developed in this article.

To conclude, our analysis of scalable resource bundles has led us to a parsimonious theory that can explain the observed properties of digital firms and their integration strategies. The many extensions we highlight illustrate the potential of our theory and modeling framework for a rich body of future work. Inevitably, the complexity and the vast heterogeneity among firms are such that our theory cannot possibly explain the idiosyncratic features of all firms. Nonetheless, we suggest that it adds an essential dimension to our understanding of the rich tapestry of economic organization in the digital age.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge valuable feedback from Nick Argyres, Dan Levinthal, Marvin Lieberman, Claudio Panico, Arkadiy Sakhartov and participants at the 2nd AI and Strategy Consortium (Indian School of Business, 2021); Academy of Management Annual Meeting (Boston, 2019); Consortium for Research in Strategy Conference (2021); IACMR Webinar Series (2021); Imperial College Business School Strategy Seminar (2020); Institutions and Innovation Conference (Harvard Business School, 2019); Strategic Management Society Special Conference (Frankfurt, 2019); University of Illinois Strategy Proseminar (Urbana-Champaign, 2021); and Wharton Technology and Innovation Conference (2021). Customary disclaimers apply. Kretschmer gratefully acknowledges support from the German Research Foundation (dfg), Grant No. KR 3479/8-1.

[Correction made on 05 March 2022 after first online publication: The Acknowledgement section has been updated in this version.]

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

ORCID

Gianluigi Giustiziero  <https://orcid.org/0000-0003-0649-6599>

Tobias Kretschmer  <https://orcid.org/0000-0002-7830-7825>

Deepak Somaya  <https://orcid.org/0000-0003-0111-4862>

Brian Wu  <https://orcid.org/0000-0003-4213-3808>

REFERENCES

- Adner, R. (2017). Ecosystem as structure: An actionable construct for strategy. *Journal of Management*, 43(1), 39–58.
- Adner, R., Puranam, P., & Zhu, F. (2019). What is different about digital strategy? From quantitative to qualitative change. *Strategy Science*, 4(4), 253–261.
- Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Brighton, MA: Harvard Business Press.
- Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *American Economic Review*, 62(5), 777–795.
- Alfaro, L., Antràs, P., Chor, D., & Conconi, P. (2019). Internalizing global value chains: A firm-level analysis. *Journal of Political Economy*, 127(2), 508–559.
- Amit, R., & Shoemaker, P. J. H. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, 14(1), 33–46.
- Anand, J., & Singh, H. (1997). Asset redeployment, acquisitions and corporate strategy in declining industries. *Strategic Management Journal*, 18, 99–118.
- Argyres, N. S. (1996). Evidence on the role of firm capabilities in vertical integration decisions. *Strategic Management Journal*, 17(2), 129–150.
- Argyres, N. S., & Zenger, T. R. (2012). Capabilities, transaction costs, and firm boundaries. *Organization Science*, 23(6), 1643–1657.
- Arthur, B. W. (1989). *Increasing returns and path dependence in the economy*. Ann Arbor, MI: University of Michigan Press.
- Arthur, B. W. (1996). Increasing returns and the new world of business. *Harvard Business Review*, 74(4), 100–109.
- Atalay, E., Hortaçsu, A., & Syverson, C. (2014). Vertical integration and input flows. *American Economic Review*, 104(4), 1120–1148.
- Athey, S., Milgrom, P., & Roberts, J. (1998). Robust comparative statics. Unpublished research monograph. Retrieved from <https://scholar.harvard.edu/files/athey/files/draftmonograph98.pdf>
- Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2), 645–709.
- Balakrishnan, S., & Wernerfelt, B. (1986). Technical change, competition, and vertical integration. *Strategic Management Journal*, 7(4), 347–359.
- Baldwin, C. Y., & Clark, K. B. (2000). *Design rules: The power of modularity*. Cambridge, MA: The MIT Press.
- Baldwin, C. Y., & Clark, K. B. (2008). Where do transactions come from? Modularity, transactions, and the boundaries of firms. *Industrial and Corporate Change*, 17(1), 155–195.
- Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120.
- Barney, J. B. (1999). How a firm's capabilities affect boundary decisions. *MIT Sloan Management Review*, 40(3), 137–145.
- Becker, G. S., & Murphy, K. M. (1992). The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics*, 107(4), 1137–1160.
- Bennett, V. M. (2013). Organization and bargaining: Sales process choice at auto dealerships. *Management Science*, 59(9), 2003–2018.
- Brandenburger, A. M., & Nalebuff, B. J. (1996). *Co-opetition: A revolution mindset that combines competition and cooperation*. New York, NY: Currency Doubleday.
- Brynjolfsson, E., Malone, T. W., Gurbaxani, V., & Kambil, A. (1994). Does information technology lead to smaller firms? *Management Science*, 40(12), 1628–1644.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. New York, NY: WW Norton & Company.
- Casadesus-Masanell, R., & Yoffie, D. B. (2007). Wintel: Cooperation and conflict. *Management Science*, 53(4), 584–598.
- Castanias, R. P., & Helfat, C. E. (1991). Managerial resources and rents. *Journal of Management*, 17(1), 155–171.
- Cennamo, C., & Santaló, J. (2019). Generativity tension and value creation in platform ecosystems. *Organization Science*, 30(3), 617–641.

- Chandler, A. D. J. (1977). *The visible hand: The managerial revolution in American business*. Cambridge, MA: Belknap Press.
- Chandler, A. D. J. (1990). *Scale and scope: The dynamics of industrial capitalism*. Cambridge, MA: Belknap Press.
- Chatain, O. (2014). How do strategic factor markets respond to rivalry in the product market? *Strategic Management Journal*, 35(13), 1952–1971.
- Chatain, O., & Zemsky, P. (2007). The horizontal scope of the firm: Organizational tradeoffs vs. buyer-supplier relationships. *Management Science*, 53(4), 550–565.
- Chatain, O., & Zemsky, P. (2011). Value creation and value capture with frictions. *Strategic Management Journal*, 32(11), 1206–1231.
- Chesky, B. (2017). Handcrafted: Airbnb's Brian Chesky. *Masters of Scale*. Retrieved from <https://mastersofscale.com/brian-chesky-handcrafted/>.
- Chu, L., & Wu, B. (2021). Designing offline platforms for customized goods and services: A market-frictions based perspective. *Academy of Management Review*, forthcoming. <https://doi.org/10.5465/amr.2020.0247>.
- Coase, R. H. (1937). The nature of the firm. *Economica*, 4(16), 386–405.
- Cusumano, M. A., Gawer, A., & Yoffie, D. B. (2019). *The business of platforms: Strategy in the age of digital competition, innovation, and power*. New York, NY: HarperCollins Publishers.
- Demsetz, H. (1988). The theory of the firm revisited. *Journal of Law, Economics, and Organization*, 4(1), 141–162.
- Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science*, 35(12), 1504–1512.
- Diewert, W. E., & Fox, K. (2008). On the estimation of returns to scale, technical progress and monopolistic markups. *Journal of Econometrics*, 145(1–2), 174–193.
- Dixit, A., & Stiglitz, J. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, 67(3), 297–308.
- Dixon, C. (2014). Full stack startups. Retrieved from <https://cdixon.org/2014/03/15/full-stack-startups>.
- Dosi, G., Gambardella, A., Grazzi, M., & Orsenigo, L. (2008). Technological revolutions and the evolution of industrial structures: Assessing the impact of new technologies upon the size and boundaries of firms. *Capitalism and Society*, 3(1), 1–49.
- Dushnitsky, G., & Stroube, B. K. (2021). Low-code entrepreneurship: Shopify and the alternative path to growth. *Journal of Business Venturing Insights*, 16, e00251.
- Fu, D., Pammolli, F., Buldyrev, S. V., Riccaboni, M., Matia, K., Yamasaki, K., & Stanley, H. E. (2005). The growth of business firms: Theoretical framework and empirical evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52), 18801–18806.
- Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30(1), 185–206.
- Ganco, M., Kapoor, R., & Lee, G. K. (2020). From rugged landscapes to rugged ecosystems: Structure of interdependencies and firms' innovative search. *Academy of Management Review*, 45(3), 646–674.
- Garicano, L., & Wu, Y. (2012). Knowledge, communication, and organizational capabilities. *Organization Science*, 23(5), 1382–1397.
- Giustiziero, G. (2020). Is the division of labor limited by the extent of the market? Opportunity cost theory with evidence from the real estate brokerage industry. *Strategic Management Journal*, 42, 1344–1378.
- Giustiziero, G., Kaul, A., & Wu, B. (2019). The dynamics of learning and competition in Schumpeterian environments. *Organization Science*, 30(4), 668–693.
- Hagiu, A., & Wright, J. (2019). The optimality of ad valorem contracts. *Management Science*, 65(11), 5219–5233.
- Hecker, A., & Kretschmer, T. (2010). Outsourcing decisions: The effect of scale economies and market structure. *Strategic Organization*, 8(2), 155–175.
- Helfat, C. E. (2015). Vertical firm structure and industry evolution. *Industrial and Corporate Change*, 24(4), 803–818.
- Helfat, C. E., & Campo-Rembado, M. A. (2016). Integrative capabilities, vertical integration, and innovation over successive technology lifecycles. *Organization Science*, 27(2), 249–264.
- Helfat, C. E., & Eisenhardt, K. M. (2004). Inter-temporal economies of scope, organizational modularity, and the dynamics of diversification. *Strategic Management Journal*, 25(13), 1217–1232.
- Helfat, C. E., & Peteraf, M. A. (2003). The dynamic resource-based view: Capability lifecycles. *Strategic Management Journal*, 24(10), 997–1010.

- Helpman, E. (1985). Multinational corporations and trade structure. *Review of Economic Studies*, 52(3), 443–457.
- Hoffman, R., & Yeh, C. (2018). *Blitzscaling: The lightning-fast path to building massively valuable businesses*. New York, NY: Currency.
- Holmes, T. (2020). Statistics and figures about Shopify. *Shopify & You*. Retrieved from <https://www.shopifyandyou.com/blogs/news/statistics-about-shopify>.
- Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, 39(8), 2255–2276.
- Jacobides, M. G., Knudsen, T., & Augier, M. (2006). Benefiting from innovation: Value creation, value appropriation and the role of industry architectures. *Research Policy*, 35(8), 1200–1221.
- Jacobides, M. G., & Winter, S. G. (2005). The co-evolution of capabilities and transaction costs: Explaining the institutional structure of production. *Strategic Management Journal*, 26(5), 395–413.
- Jia, N. (2013). Competition, governance, and relationship-specific investments: Theory and implications for strategy. *Strategic Management Journal*, 34(13), 1551–1567.
- Kafka, P. (2010). Those bits aren't free: Netflix could be racking up a \$2 billion content tab. *All Things D*. Retrieved from http://allthingsd.com/20101027/those-bits-arent-free-netflix-could-be-racking-up-a-2-billion-content-tab/?mod=ATD_rss.
- Kafka, P. (2011a). "House of Cards" could cost Netflix big—and still save it money in the end. *All Things D*. Retrieved from <http://allthingsd.com/20110316/house-of-cards-could-cost-netflix-big-and-still-save-it-money-in-the-end/>.
- Kafka, P. (2011b). Netflix bets big on "House of Cards" but swears it's not a strategy shift: Q&A with content boss Ted Sarandos. *All Things D*. Retrieved from <http://allthingsd.com/20110318/netflix-bets-big-on-house-of-cards-but-swears-its-not-a-radical-departure-qa-with-content-boss-ted-sarandos/>.
- Kang, M., Mahoney, J., & Tan, D. (2009). Why firms make unilateral investments specific to other firms: The case of OEM suppliers. *Strategic Management Journal*, 30(2), 117–135.
- Kaul, A. (2013). Entrepreneurial action, unique assets, and appropriation risk: Firms as a means of appropriating profit from capability creation. *Organization Science*, 24(6), 1765–1781.
- Kim, J. (1997). *Three sources of increasing returns to scale* (p. 18). Washington: Finance and Economics Discussion Series of the Federal Reserve Board.
- Knudsen, T., Levinthal, D. A., & Winter, S. G. (2014). Hidden but in plain sight: The role of scale adjustment in industry dynamics. *Strategic Management Journal*, 35(11), 1569–1584.
- Kretschmer, T., & Khashabi, P. (2020). Digital transformation and organization design: An integrated approach. *California Management Review*, 62(4), 86–104.
- Kretschmer, T., Leiponen, A., Schilling, M., & Vasudeva, G. (2020). Platform ecosystems as meta-organizations: Implications for platform strategies. *Strategic Management Journal*, 1–20 (forthcoming).
- Kretschmer, T., & Puranam, P. (2008). Integration through incentives within differentiated organizations. *Organization Science*, 19(6), 860–875.
- Lamoreaux, N. R., Raff, D. M., & Temin, P. (2004). Against Whig History. *Enterprise & Society*, 5(3), 376–387.
- Langlois, R. N. (2003). The vanishing hand: The changing dynamics of industrial capitalism. *Industrial and Corporate Change*, 12(2), 351–385.
- Langlois, R. N. (2004). Chandler in a larger frame: Markets, transaction costs, and organizational form in history. *Enterprise & Society*, 5(3), 355–375.
- Levie, A., & London, S. (2018). From start-up to scale: A conversation with Box CEO Aaron Levie. *McKinsey Podcast*. Retrieved from <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/from-start-up-to-scale-a-conversation-with-box-ceo-aaron-levie>.
- Levinthal, D. A., & Wu, B. (2010). Opportunity costs and non-scale free capabilities: Profit maximization, corporate scope, and profit margins. *Strategic Management Journal*, 31(7), 780–801.
- Linden, G., & Somaya, D. (2003). System-on-a-chip integration in the semiconductor industry: Industry structure and firm strategies. *Industrial and Corporate Change*, 12(3), 545–576.
- Lütke, T. (2020). Be a platform: Tobi Lütke, founder & CEO, Shopify. *Masters of Scale*. Retrieved from <https://mastersofscale.com/tobi-lutke-be-a-platform>.
- Macher, J. T., & Mowery, D. C. (2004). Vertical specialization and industry structure in high technology industries. *Advances in Strategic Management*, 21, 317–355.

- Madhok, A. (1996). The organization of economic activity: Transaction costs, firm capabilities, and the nature of governance. *Organization Science*, 7(5), 577–590.
- Madhok, A. (2002). Reassessing the fundamentals and beyond: Ronald Coase, the transaction cost and resource-based theories of the firm and the institutional structure of production. *Strategic Management Journal*, 23(6), 535–550.
- Mahoney, J. T., & Pandian, J. R. (1992). The resource-based view within the conversation of strategic management. *Strategic Management Journal*, 13(5), 363–380.
- Mahoney, J. T., & Qian, L. (2013). Market frictions as building blocks of an organizational economics approach to strategic management. *Strategic Management Journal*, 34(9), 1019–1041.
- Makadok, R. (1999). Interfirm differences in scale economies and the evolution of market shares. *Strategic Management Journal*, 20(10), 935–952.
- Makadok, R., & Coff, R. (2009). Both market and hierarchy: An incentive-system theory of hybrid governance forms. *The Academy of Management Review*, 34(2), 297–319.
- Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*. Oxford: Oxford University Press.
- Mayer, K. J., & Salomon, R. M. (2006). Capabilities, contractual hazards, and governance: Integrating resource-based and transaction cost perspectives. *Academy of Management Journal*, 49(5), 942–959.
- Mayer, K. J., Somaya, D., & Williamson, I. O. (2012). Firm-specific, industry-specific, and occupational human capital and the sourcing of knowledge work. *Organization Science*, 23(5), 1311–1329.
- McKinsey Global Institute. (2015). Digital America: A tale of the haves and have-mores. December 2015: 1–120. Retrieved from <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/digital-america-a-tale-of-the-haves-and-have-mores>.
- McNew, B. S. (2018). Do not disturb: Why Marriott and Hilton's franchise model works. *The Motley Fool*. Retrieved from <https://www.fool.com/investing/2016/12/21/do-not-disturb-why-marriott-and-hiltons-franchise.aspx>.
- Miller, D. J., & Yang, H. (2016). Product turnover: Simultaneous product market entry and exit. *Advances in Strategic Management*, 35, 49–87.
- Montgomery, C. A., & Wernerfelt, B. (1988). Diversification, Ricardian rents, and Tobin's q. *RAND Journal of Economics*, 19(4), 623–632.
- Moore, G. (1995). Lithography and the future of Moore's law. Paper presented at SPIE's 1995 Symposium on Microlithography, Santa Clara, CA. Retrieved from <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/2437/0000/Lithography-and-the-future-of-Moores-law/10.1117/12.209151.short>.
- Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal*, 18, 187–206.
- Pacheco-de-Almeida, G., & Zemsky, P. (2003). The effect of time-to-build on strategic investment under uncertainty. *The Rand Journal of Economics*, 34(1), 166.
- Pacheco-de-Almeida, G., & Zemsky, P. (2007). The timing of resource development and sustainable competitive advantage. *Management Science*, 53(4), 651–666.
- Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. New York, NY: WW Norton & Company.
- Parmigiani, A. (2007). Why do firms both make and buy? An investigation of concurrent sourcing. *Strategic Management Journal*, 28(3), 285–311.
- Parmigiani, A., & Mitchell, W. (2009). Complementarity, capabilities, and the boundaries of the firm: The impact of within-firm and interfirm expertise on concurrent sourcing of complementary components. *Strategic Management Journal*, 30(10), 1065–1091.
- Penrose, E. T. (1959). *The theory of the growth of the firm*. Oxford, England: Blackwell Publishers.
- Peteraf, M. A. (1993). The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal*, 14(3), 179–191.
- Porter, M. E. (1980). *Competitive strategy: Techniques for analyzing industries and competitors*. New York, NY: Free Press.
- Postrel, S. (2009). Multitasking teams with variable complementarity: Challenges for capability management. *Academy of Management Review*, 34(2), 273–296.
- Postrel, S. (2018). Transaction surplus superiority in canonical market segments: Using the profit map to guide positioning and investment choices across price-rivalry regimes. *Strategic Management Journal*, 39(6), 1573–1602.

- Richardson, G. B. (1972). The organisation of industry. *Economic Journal*, 82(327), 883–896.
- Rumelt, R. P. (1984). Toward a strategic theory of the firm. In R. B. Lamb (Ed.), *Competitive strategic management*. Englewood Cliffs, NJ: Prentice-Hall.
- Sakhartov, A. V., & Folta, T. B. (2014). Resource relatedness, redeployability, and firm value. *Strategic Management Journal*, 35(12), 1781–1797.
- Sakhartov, A. V., & Folta, T. B. (2015). Getting beyond relatedness as a driver of corporate value. *Strategic Management Journal*, 36(13), 1939–1959.
- Schmidt, J., & Keil, T. (2013). What makes a resource valuable? Identifying the drivers of firm-idiosyncratic resource value. *Academy of Management Review*, 38(2), 206–228.
- Schreiber, Z. (2016). Amazon logistics services—The future of logistics? *Supply Chain 24/7*. Retrieved from https://www.supplychain247.com/article/amazon_logistics_services_the_future_of_logistics.
- Seward, Z. (2016). The most interesting tech IPO of the year. *Quartz*. Retrieved from <https://qz.com/695430/the-most-interesting-tech-ipo-of-the-year/>.
- Siebel, T. M. (2019). *Digital transformation: Survive and thrive in an era of mass extinction*. New York, NY: RosettaBooks.
- Stigler, G. (1951). The division of labor is limited by the extent of the market. *Journal of Political Economy*, 59(3), 185–193.
- Sun, L. (2017). The 3 biggest misconceptions about Twilio Inc. *The Motley Fool*. Retrieved from <https://www.fool.com/investing/2017/10/04/the-3-biggest-misconceptions-about-twilio-inc.aspx>.
- Sutton, J. (1997). *Technology and market structure*. Cambridge, MA: MIT Press.
- Tambe, P., Hitt, L., Rock, D., & Brynjolfsson, E. (2020). Digital capital and superstar firms (Working Paper No. w28285). National Bureau of Economic Research.
- Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, 15(6), 285–305.
- Teece, D. J. (2013). In C. R. Thomas & W. S. Shughart (Eds.), *The new managerial economics of firm growth* (pp. 279–301). Oxford: Oxford University Press.
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533.
- The Economist. (2021a). How to thrive in the shadow of giants. Retrieved from <https://www.economist.com/business/2021/05/20/how-to-thrive-in-the-shadow-of-giants>.
- The Economist. (2021b). The new rules of competition in the technology industry. Retrieved from <https://www.economist.com/business/2021/02/27/the-new-rules-of-competition-in-the-technology-industry>.
- Viner, J. (1932). Cost curves and supply curves. *Journal of Economics*, 3(1), 23–46.
- Wan, Z., & Wu, B. (2017). When suppliers climb the value chain: A theory of value distribution in vertical relationships. *Management Science*, 63(2), 477–496.
- Wang, Z., & Wright, J. (2017). Ad valorem platform fees, indirect taxes, and efficient price discrimination. *The Rand Journal of Economics*, 48(2), 467–484.
- Weissman, A. (2015). No stack startups. Retrieved from <http://blog.aweissman.com/2015/05/no-stack-startups.html>.
- Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, 5(2), 171–180.
- Wessel, M., Levie, A., & Siegel, R. (2017). Why some digital companies should delay profitability for as long as they can. *Harvard Business Review*. Retrieved from <https://hbr.org/2017/05/why-some-digital-companies-should-resist-profitability-for-as-long-as-they-can>.
- West, G. B. (2017). *Scale: The universal laws of growth, innovation, sustainability, and the pace of life in organisms, cities, economies, and companies*. Westminster, England: Penguin.
- Wibbens, P. D. (2021). The role of competitive amplification in explaining sustained performance heterogeneity. *Strategic Management Journal*, 42, 1769–1792.
- Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. New York, NY: Macmillan.
- Williamson, O. E. (1999). Strategy research: Governance and competence perspectives. *Strategic Management Journal*, 20(12), 1087–1108.
- Wilson, F. (2015). The no stack startup. Retrieved from <https://avc.com/2015/05/the-no-stack-startup/>.

- Wu, B. (2013). Opportunity costs, industry dynamics, and corporate diversification: Evidence from the cardiovascular medical device industry, 1976–2004. *Strategic Management Journal*, 34(11), 1265–1287.
- Zhu, F., & Liu, Q. (2018). Competing with complementors: An empirical look at Amazon.com. *Strategic Management Journal*, 39(10), 2618–2642.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Giustiziero, G., Kretschmer, T., Somaya, D., & Wu, B. (2023). Hyperspecialization and hyperscaling: A resource-based theory of the digital firm. *Strategic Management Journal*, 44(6), 1391–1424. <https://doi.org/10.1002/smj.3365>