Designing Scalable Digital Business Models

Joanne Jin Zhang
Norwich Business School,
University of East Anglia

Yossi Lichtenstein
Norwich Business School,
University of East Anglia

Jonathan Gander
London College of Fashion,
University of the Arts, London.

Abstract

Digital business models are often designed for rapid growth, and some relatively young companies have indeed achieved global scale. However despite the visibility and importance of this phenomenon, analysis of scale and scalability remains underdeveloped in management literature. When it is addressed, analysis of this phenomenon is often over-influenced by arguments about economies of scale in production and distribution. To redress this omission, this paper draws on economic, organization and technology management literature to provide a detailed examination of the sources of scaling in digital businesses. We propose three mechanisms by which digital business models attempt to gain scale: engaging both non-paying users and paying customers; organizing customer engagement to allow self-customization; and orchestrating networked value chains, such as platforms or multi-sided business models. Scaling conditions are discussed, and propositions developed and illustrated with examples of big data entrepreneurial firms.

Keywords: Scalability, scale, business model design, digital business

Designing Scalable Digital Business Models

Introduction

Over the last two decades, advances in digital technologies have created new possibilities for businesses to scale up. Fast growing young entrepreneurial firms such as Google, Facebook, eBay and Alibaba hold dominant competitive positions and have market capitalizations of tens to hundreds of billions of dollars. Despite these notable examples, our understanding of scale and scalability is underdeveloped, and is overly influenced by the presence of cost advantages achieved through economies of scale in production and distribution. Furthermore, analyses of the business models that have enabled such growth are scattered across economic, organization and technology management literature. This is a missed opportunity, as business growth is not just a signifier of success but can be a planned outcome, a consequence of designing business models in such a way as to enable large scale. For example, while Google's technological innovation lies in addressing scalability in web searching (Brin & Page, 1998), it is the design of their business model that has enabled them to achieve immense scale (Battelle, 2005). This paper seeks to integrate these various streams of literature and apply them to the study of digital business models.

In his seminal work on scale, Chandler (1990) proposed that while technological advances may provide opportunities to achieve the cost advantages associated with supply-side economies of scale, it was managers who had to make the necessary choices required to exploit these opportunities fully: and that consideration continues to apply in the digital economy. We argue that, while the digitalized, non-material nature of the goods and services involved provide the potential for digital businesses to achieve high scalability, it is their managers' business model choices - related to economies of scale on both the demand and supply sides - that have helped realize such potential. For instance, while the World Wide Web's communication and presentation technologies have dramatically reduced the cost of creating markets, it is (for instance) eBay's business model design of self-service for both sellers and buyers and an open rather than hierarchical reputation-building mechanism that has enabled it to take full advantage of those technological opportunities.

Business model design can be seen as a configuration of activities that not only creates value for customers but also allows the firms to capture part of that value (Chesbrough, 2003; Zott & Amit, 2007; Teece, 2010). To try to update the notion of business model scalability, we reason in three stages. First, we see scalability in businesses not simply as referring to growth in size (Chandler, 1990). Yes, a scalable business model is one in which operational elements have been organized to promote growth, but importantly this is achieved while also preserving and in some cases increasing the quality and features of products/services. To complement the view of scaling as growth in size (Chandler, 1990), we draw on the software engineering perspective of scalability that refers to the ability of a system to satisfy quality performance goals when user numbers or other characteristics vary (Duboc, Rosenblum & Letier, 2010; Duboc, Letier & Rosenblum, 2013). Thus, we define business model scalability as follows:

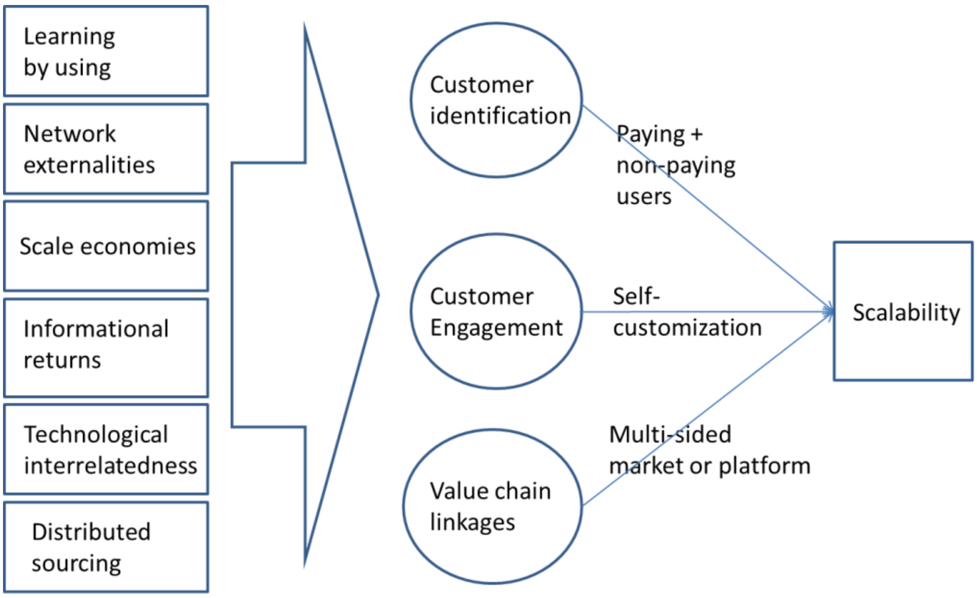
> Business model scalability is the extent to which a business model design may achieve its desired value creation and capture targets when user/customer numbers increase and their needs change, without adding proportionate extra resources.

Second, we examine the sources of digital business scalability drawing on literature on the increased returns to adoption in technological markets (Arthur, 1988). While he focuses on the adopters' perspective, we posit that the same sources that underlie increased returns to the adoption of a technology continue to affect the users and customers who adopt firms' products, and thus create increased returns to scale. Specifically, Arthur (1988) noted that scalability is enhanced by the dynamics of learning by using, network externalities, production economies, informational increasing returns and technological interrelatedness. In addition, we suggest that distributed resourcing (Baldwin & Clark, 2000; Gawer, 2014), that is, the decentralized organization of operational resources, is also an important source of scalability.

Third, we argue that business model design can be seen as the mechanism by which these sources of scalability can be realized, so we examine how the business model design elements can either enable or restrict scalability. Business model design is a distinctive level of analysis (Baden-Fuller & Morgan, 2010), and involves the configuration of a set of interdependent activities that can be grouped around the tasks of customer identification, customer engagement, value chain links and monetization (Baden-Fuller & Mangematin, 2013; Baden-Fuller & Haefliger, 2013). Customer identification involves decisions about the presence of and balance between paying customers and non-paying users. As these authors have pointed out, a distinction can be made between business models involving only paying customers and those that target free users as well. In addition, customer engagement distinguishes between project based offerings - a 'taxi' approach - and generic, predesigned offerings - a 'bus' approach. Value chain linkages consider how hierarchical and controlled vs. how networked and open, governance systems are. The final dimension - monetization - describes how payments are appropriated and organized to cover production costs. We do not consider the monetization dimension separately in this paper, as we believe its effect on scalability mostly depends on the choices taken in the first three dimensions. Instead, issues to do with capturing the value created by the business models are discussed indirectly when we consider the impediments to scalability.

Our analysis results in three principal propositions. We propose that business models that engage both non-paying users and paying customers; that facilitate customer participation in the production of products or services; and that open the value chain to network governance, tend to promote scalability. We also discuss how the interactions between these three business model design elements may affect scalability, as well as the factors that might countervail or offset scalability. Figure 1 illustrates our line of reasoning.

Figure 1: Sources and mechanisms for scalable digital business models



Our paper contributes to the literature in the following ways. First, we update the notion of scalability and highlight its importance in the design of digital business models, an issue often mentioned but rarely discussed in the extant literature. Second, by combining managerial, economic and technological perspectives, we examine the sources of scalability, including both supply and demand-related economies of scale. Although both those types of economies of scale have been discussed before (e.g., Arthur, 1988; Shapiro & Varian, 1999;

Gawer, 2014), framing them in terms of scalability provides insights into how digital businesses scale up. Third, we unpack the characteristics of scalable digital business models by considering three core business model design elements, specifically, customer identification, customer engagement, and value chain linkages. We also provide propositions relating scalability to business model configurations, which we hope will encourage empirical grounded business model research that is so far lacking (Markides, 2013).

In the following sections, we first identify scaling sources, and then examine the relationship between business model design and scaling. We next use examples of business models of big data entrepreneurial firms to illustrate our propositions. Finally, we discuss our contributions, their managerial implications, the limitations of our paper and possible directions for further research.

Theoretical Background: Sources of Scaling in Digital Businesses

Digital businesses are those which carry out transactions that are digitally mediated, or involve products or services that are experienced digitally (Weill & Woerner, 2013). It is the digitized, non-material nature of such goods and services that gives them the potential for high scalability, and thus suits our analysis of the scalability of digital business models. The term 'digital businesses' deliberately incorporates a very wide range of firms and business activities. Many businesses can be said to deploy, either in production or distribution, some form of digital technology (Yoo et al., 2012): indeed, the familiar continuum between physical products and intangible services (Rathmell, 1966) is perhaps better described nowadays as a digital continuum - one that is organized according to the degree to which a firm's operations and a customer's experience are made possible by digital technologies - whether the product is almost entirely digitally realized (e.g. social media platform), is an interface for digital products (e.g. a mobile device), incorporates digital elements within a physical object (e.g. fitness gear) or is a physically rendered service (e.g. a home help service). The extent to which a business model might be configured to take advantage of scaling dynamics and benefits will therefore differ depending on their position along this continuum. The emphasis in this paper is on business models where the role of digital technology in the transactions, resources and the customer value proposition can be described as 'pervasive' (Yoo et al., 2012).

Before we conduct this analysis it is necessary to first establish the sources of scalability deriving from the production and consumption of digital products and services. To identify and discuss the sources of scalability that business models are configured to realize, we draw on Arthur's (1988; 1989) work on the sources underlying the adoption of new technologies. Originating from an economic perspective, his arguments describe the adoption or diffusion dynamic as one in which the attractiveness to adopt - for both suppliers and buyers - depends on the degree to which the businesses operations and product/service qualities improve in relation to the number of users. These improvements are generated by a range of factors, including the presence of increasing return effects, production and distribution cost efficiencies, lock-ins and complementarities. While Arthur focuses on the benefits accruing to adopters, we argue that the factors that generate these benefits also relate to sources of scalability. For example, the value created by each additional user under network externalities attracts yet more customers (adoption) and enables the supplier to scale up, presuming it has the capabilities to service more customers. In this paper, we build on these sources of scalability by including more recent work on the strategic, resource-based and modular aspects of digital businesses (Bharadwaj et al., 2013; Zhu, 2004; Yoo et al., 2012).

To organize this literature we have retained Arthur's original headings of 'learning by using', 'network externalities', 'scale economies in production', 'informational increasing returns' and 'technological interrelatedness' and add another - 'distributed resourcing'. As Arthur (1988) acknowledges, these categories overlap, a feature made all the more evident when developments in information technology are incorporated. Nevertheless, they continue to provide a way of structuring what is a very large field of relevant literature.

Learning by using

The greater the scale of a technology's use, the more its users learn about its features, its strengths and weaknesses. Initial interest in learning by using focused on how experiential learning might - by revealing more efficient ways of organizing, designing and producing - reduce the costs of production (Spence, 1981). This learning curve perspective has a supply side focus, involving insights generated by the organization's employees in their role as early users (Rosenberg, 1982). Recognition of the role of users in the development of a firm's products and services (von Hippel, 1976; 1986) extends the source of learning benefits to customers and their ways of discovering new sources of value. Knowledgeable users applying their experience, and their distinctive perspectives as users, can identify and help adapt or enhance the firm's products (von Hippel, 2005; Parmentier and Mangematin, 2014). A good example of this is seen in the user development of Linux software. Using an 'open source' approach enables the establishment of communities of users who can then take part developing the software. The advantages involved in having knowledgeable users review and improve the software are captured in Linux founder Linus Torvald's eponymous 'law': "Given enough eyeballs all bugs are shallow" (Raymond, 1999:29). In many settings, scale is seen as a potential threat to quality, with outcomes such as speed, reliability, accuracy and ease of use deteriorating as a greater number of products or service interactions are delivered. In digital business models it appears that scale can also act, as in this case, to increase quality.

The importance of this type of learning by using has been significantly increased with the development of internet technologies. The ability to accelerate the reach and frequency of such user participation in the development of the firm's offer has seen the emergence of communities of users sharing their knowledge amongst themselves and with firm representatives (Faraj et al., 2011. These social and technological changes have been given a number of labels, such as value co-creation (Prahalad & Ramaswamy, 2004), wikinomics (Tapscott & Williams, 2006) and prosumption - the blending of production and consumption. Whichever term is used, this sharing of efforts between firms and their customers (Tofler, 1980) is central to the potential of digital business practices (Ritzer & Jurgenson, 2010).

Network externalities

The previous source of scalability is linked to the broader phenomenon of network externalities (Rohlfs, 1974; Katz & Shapiro, 1985; 1994). Such externalities occur when the value obtained when using products or services increases with its greater diffusion among populations of users. Direct externalities - such as those that follow from the greater availability and access that accompanies increased scale of adoption - can be found in networks such as communication technologies (e.g. Skype) and social media platforms (e.g. Facebook). Indirect externalities arise when complementary products or services are created to support and enhance customers' use of products, as is seen in the case of smart phones (Church & Gandal, 1992). These complementary products and services enhance the value users obtain from using the primary or leading product, and the ties that may emerge can result in customer lock-in (Katz & Shapiro, 1994).

As there is usually a point of scale - a tipping point or critical mass - at which these externalities become significant, and business models that seek to harness this effect will focus their activities on those users who can initiate or drive the technology's adoption - the innovators and early adopters. These are individuals who can gain value (for example knowledge and social capital) before the utility value produced by externalities (Rogers, 1962). When the business model involves creating a platform where 'sides'- distinct groups of users/customers - interact, the task is arguably more challenging, as one cannot precede the other: both sides must be encouraged to adopt together (Rochet & Tirole, 2003). In this case, platforms are usually built around key contributors who drive the early interactions and help to build scale. For example, Pinterest - the photo sharing social media platform - used an invitation only referral system to spread the word about its service among social media users. Exclusivity of access increased the site's perceived value, and transforming that access into a gift that site members could give their friends encouraged diffusion. This and other incentive systems, such as granting differentiated access to heavy social media users, help both to organize and govern contributions and to encourage new participants (Brousseau & Penard, 2007b).

Once an adoption momentum has been established, network effects can build a self-reinforcing dynamic where additional participants on one side make joining more attractive to the other, and vice versa (Evans, 2003). These direct and indirect externalities create the conditions for a 'winner-takes-all' market, when the positive feedback loop generated by these two scale-derived effects severely reduces the possibility that competitors could effectively offer a rival product or service that would be of equivalent or greater value (Schilling, 2002). So, where network externalities prevail, digital business models may be configured to rapidly diffuse through their markets to reach the point at which the effects gear up and the firm can capture the value involved.

Economies of scale in production and distribution

Supply side effects of scale can include economies of production and distribution. With large production quantities, the unit cost of each product or service encounter falls. This is largely due to the ability to spread fixed costs, advertising budgets and research and design costs across a larger quantity of products, reducing the amount each item has to contribute to their repayment (Chandler, 1990).

Scale economies are particularly obvious in digital businesses, as the development costs of products and services are high, but the marginal cost of adding another customer is negligible. Under these cost conditions, digital business models engage in pre-designed customer engagement, in particular large scale bundling of digital content, such as the provision of a large number of different informational goods (music, video, etc.) as one subscribed unit (Bakos & Brynjolfsson, 1999; Hitt & Chen, 2005). Bundling is especially useful for firms catering to markets made up of customers with highly diverse tastes. Large bundles create value by giving access to a wide variety of products, while curated bundles -

selected and more discriminatory collections of goods - also offer a differentiated source of value. In taste-based markets of uncertain product value, such as cultural and creative sectors like music and film, bundling represents an unusual combination of both scale (number) and scope (variety) based sources of customer value. Thus Spotify, a successful music streaming service, provides curated playlists organized around different themes - morning blues, exercise, energy etc., which can be supplied at zero marginal cost, but reduces users' search costs

Economies of scale in production and distribution may affect also the quality of the product or service provided. Building large numbers of users can improve and sometimes exceed the efficiency and effectiveness of other tasks previously undertaken by the firm. For example, the eBay platform manages quality control through its reputation system. The larger the number of transactions, the more thorough is the feedback mechanism that operates between sellers and buyers. Using internal staff to police users' transactions would bring intolerable costs to the operation, and anyway could never match the quality of information produced by the user-generated data. Numerous examples of the ways large scale of users and transactions can replace or improve on internally provided quality control and product performance can be found. Duolingo, a language learning and translation service, improves its ability to correct and focus learners on their mistakes the more they use it, and Tor's Onion internet encryption tool increases in quality (speed and anonymity) the more users it recruits.

Informational increasing returns

Arthur's (1988) fourth source of scalability is informational increasing returns, which describes the reduction in users' perceptions of the risk of adopting a product or service following its use by others. The more widely adopted a technology is, the better it is understood and the less risky it appears. Increasing returns may also result from this effect on the psychology of buyers, as the decision to use/adopt/buy a product/service becomes easier as its use becomes more widespread (Rogers, 1962). This is especially relevant for digital products and services, whose complexity and intangibility can make it more difficult for potential users to understand what the offer is, how to use it and what value it brings. This is a form of social contagion (Burt, 1987), where adoption can follow from the legitimizing effect of scale, and digital businesses may attempt to construct a 'bandwagon' effect (Leibenstein, 1950), where perceptions as to the value and ease of adopting a product or service are increased by its popularity among peers. The more popular it becomes, the more effectively the word of mouth adoption dynamic drives further recruitment and use (Chandrashekaran et al., 2010).

These effects are supported by business models that leverage the popularity of a product with non-paying users, and network value chains that use the influence of other suppliers or complementors to create informational increasing returns and scale. For example, Hadoop, a software product for processing large data sets, has become the most popular Big Data technology due largely to the effect of informational increasing returns. Its functionality and reliability have been much enhanced by thousands of individual developers and hundreds of companies that provide training, deployment, customization and other products and services that complement Hadoop[1], a particularly successful example of which is Hortonworks, a Yahoo! spinoff that provides services for the Hadoop technology: at its IPO in December 2014 it was valued at over $1bn.

Technological interrelatedness

The next source of scaling is technological interrelatedness (Frankel, 1955, Yoo et al., 2012). This effect describes contexts where the more a particular technology is adopted, the greater the number of supporting, sub-technologies that are developed and become part of its technological infrastructure. Central to the scalability of digital businesses is a modular design that describes the way software architects structure interactions between different aspects of a product or service (Baldwin & Clark, 2000). Modules allow engineers to effectively 'hide' components, preventing them from creating interconnections with unnecessary modules and producing a complex and unwieldy network (Yoo, 2012). Rendering modules opaque reduces the 'friction' in the network by simplifying communication and exchange routes. A modular approach thus allows the development of products and services to be flexible and open, as updated components can be more easily built and 'plugged' into the system. Firms create Application Programming Interfaces (APIs) to enable this modularity, allowing external partners to develop products and services that are compatible with the focal technology and don't interfere with its other aspects. For example, as of February 2014, Google had published a total of 51 APIs for almost all of its popular consumer products such as Google Maps, YouTube and Google Search. These APIs are a key resource, giving Google access to external developers and enabling them to benefit from the value they create by producing complementary products. In one way this can be seen as an

[1] http://wiki.apache.org/hadoop/PoweredBy;
http://wiki.apache.org/hadoop/Distributions%20and%20Commercial%20Support

economy of scope (Teece, 1980; Gawer, 2014), as the cost of providing new functions and related services are in effect taken on by partner firms.

The recruitment of partner firms through APIs can result in the development of an ecosystem of firms (Gawer & Cusumano, 2002), which can create opportunities to extend the technology's functionality and value and create barriers to entry for rival technologies. The ultimate result is more customers, revenues and scale. Digital business models are accordingly configured to take advantage of these economies of scope and scale through network value chains and a pre-designed ('bus') approach to customer engagement.

Distributed resourcing

Our final source of scalability – the one we have added to Arthur's (1988; 1989) list - is distributed resourcing. This condition refers to the way resources are decentralized across value chains and ecosystems. The distributed design of a digital business's model's informational resources enables the rapid growth in use of a product or service without sacrificing performance or causing congestion in the system which degrades the customer experience.

The requirement for digital business models to use a distributed approach to resource deployment and development follows from the strong network effects identified earlier. The effects of zero marginal costs, the dynamic of informational increasing returns, the prospect of bandwagons and the development of interrelated technologies and complimentary products and services which 'piggy-back' on primary technologies or platforms represent a very favorable set of scaling conditions. Examples of distributed resourcing include the Internet Protocol itself, which was designed to use redundancy and distributed control to create reliable communication using unreliable communication nodes (Baran, 1960). Skype is another good example - its communication services depend on users' own communication gear, as the company itself does not own communication networks. A recent example that demonstrates the importance of distributed data and communication sources occurred during a TV program in Japan, when the number of new tweets suddenly rose from an average of around 6,000 per second to over a 140,000[2]. Despite this being a far greater than previous use levels, Twitter users did not experience any reduction in performance. This is the flipside to positive scaling conditions, the possibility of negative externalities produced by the very dynamics that favor scale.

[2] https://blog.twitter.com/2013/new-tweets-per-second-record-and-how. Accessed 23rd September 2014

An important negative externality is the potential for congestion to occur - an overloading of capacity following a rapid unanticipated, and sizeable increase in user numbers - which can result in service failure. In response, scalable business models feature distributed rather than centralized systems of control points and resource utilization, which help avoid bottlenecks, or failure points in supply that can result from hierarchically ordered decision protocols and centrally controlled resources (Garud & Kumaraswarmy, 1995). A distributed system thus allows operational capacity to grow by simplifying communication between the technology's different distributed components. This involves the replication of data in different components and the design of algorithms that can be used to allow multiple, numerous and simultaneous access to and use of information exchange and transaction systems.

Business Model Design Elements and Scalability

In order to analyze how the above sources of scale advantage might be operationalized, we consider how these generative factors align to the activities that comprise a firm's business model. Research on such activities views a firm as a system made up of a set of interdependent activities that combine to produce the whole (Miller, 1986; 1996, Porter, 1996). These activities are configured in such a way that they are mutually reinforcing and complementary, thus ensuring that they add up to a coherent, understandable and manageable whole (Siggelkow, 2002). At the business model level of analysis (Baden-Fuller & Morgan, 2010), this notion of configuration is applied in a more discrete manner, involving the identification of key activities that take place between customers and the organization (Zott & Amit, 2010; Chesbrough, 2010; Baden-Fuller & Mangematin, 2013).

In a development of this approach, scholars have proposed four dimensions of activity that they argue capture the essence of a business model and, in so doing, enable researchers to examine such models and their relationships to technological developments (Baden-Fuller & Haefliger, 2013). As noted above, these dimensions are labeled 'customer identification', 'customer engagement', 'value chain linkages' and 'monetization'. We focus on the first three, suggest some new ways of considering their characteristics and propose that different ways of configuring each of these elements are associated with high or low scalability. We consider indirectly how the business model monetizes the value created through actions taken on these three dimensions when we discuss possible obstacles to scalability, and the tensions that can be created when firms attempt to capture value whilst also encouraging adoption

(Boudreau, 2010).

We illustrate (rather than verify) our propositions using examples from entrepreneurial Big Data firms which seek to leverage newly available internet, mobile and video data assets, as well as the new technologies of large-scale data retrieval, analysis and management, to develop and commercialize new products and services (McKinsey, 2011; McAfee & Brynjolfsson, 2012). We use this sector for our illustrations because Big Data is an important technological innovation in the digital sector, and the scalability of their business models is central to these firms' survival and growth.

Customer identification and scalability

This business model design dimension refers to the identification of customer groups and specifically whether to engage or target free users. A distinction can be made between business models involving only paying customers and those that target both non-paying users and paying customers (Baden-Fuller & Haefliger, 2013). This decision is central to a business model's design, because it incorporates either two value delivery systems (Baden-Fuller & Haefliger, 2013) or other mechanisms that complement the free offering (McGrath, 2010), and thus goes beyond simple price discrimination between different customer groups. Business models that are organized around both paying customers and non-paying users are often realized as multi-sided or two sided platforms (Baden-Fuller & Haefliger, 2013) - but not always: both paying customers and free users can also exist within single sided business models. This is the case in many 'freemium' business models, such as that operated by Rovio, the makers of the 'Angry Birds' mobile phone game. Employing a 'bait and hook' tactic, a free version of the game is offered to stimulate demand by increasing awareness and understanding of its value: options to purchase extra features and additional functionality are then offered to game users.

While the identification of paying customers and free users has been described in the literature as a decision that plays a key role in the adoption of new products or services in the digital sector, it also has a significant effect on scale, mostly through factors related to demand. First, assuming network externalities, a digital business model that includes both non-paying users and paying customers increases its client-base, thus increasing the size of the network related to a new product and ultimately increasing the product's value. Second, a digital business model that engages with paying customers and non-paying users creates considerable potential for learning by using, because a large number of different users with a correspondingly greater variety of needs are likely to adopt a new product, and to engage in its further development and improvement. Third, such a configuration can make a product part of the technological infrastructure, creating potential for scale through technological interrelatedness. As a technology becomes more adopted - because it has many free users - a wider range of sub-technologies become part of its infrastructure, further increasing adoption and scale. Finally, as mentioned earlier, these demand factors may also influence economies of scale, possibly in production but mostly in distribution. As free users create demand and increase returns because of peer influence and viral marketing, marketing and distribution costs per customer usually fall.

Table 1: Sources of scalability for business models that engage both non-paying users and paying customers

Source	Assumptions	Required Capabilities	Scalability Effects
Learning by using	The technology is not fully developed and learning to use it reveals opportunities for further development that increases its appeal to paying customers	Managerial attention and resources related to the needs and improvements required by non-paying users	Large number and variety of non-paying users increase development, appeal and scale
Network externalities	Larger number of non-paying users increases the value for paying customers	Know-how and resources to cater for both paying customers and non-paying users	Non-paying users increase the size of the network and ultimately the product's value and scale
Scale economies in production & distribution	The cost of serving non-paying users is relatively small and decreases as their number grows. Non-paying users influence adoption by paying customers	Resources to service non-paying users; know-how about leveraging non-paying users' experience for marketing	As the number of non-paying users grows, marketing and distribution costs decrease
Informational increasing returns	Non-paying users make the product better known and understood by potential paying customers	Know-how about leveraging non-paying users' experience in promoting the product	As the number of non-paying users grows, the product is understood better and adopted by paying customers
Technological interrelatedness	The technology may benefit from sub-technologies provided by other suppliers to cater for non-paying users	Modular design and APIs are required; orchestration of the ecosystems of suppliers, paying customers and non-paying users	Non-paying users create incentives for suppliers of sub-technologies that create infrastructure for the focal product adoption and scale
Distributed resourcing	Resources and practices are distributed	Design of distributed resources; ability to add resources rapidly; ability to utilize user's own resources	Growth in the number of users and customers does not degrade performance

Table 1 details these considerations. We present first the assumptions and capabilities required to achieve scalability through each of the sources, and then the effects on scalability given these assumptions and capabilities. The scaling mechanisms of the six sources of scalability outlined above can be read as a list of propositions about their antecedents and effects. Proposition 1 summarizes this discussion:

> Proposition 1: Business models that engage both paying customers and non-paying users promote scalability in digital businesses.

The next step in our reasoning is to consider the interactions between the customer identification element and the other two business model dimensions we consider - customer engagement and value chain linkages. In the case of customer engagement, the logic of Table 1 is consistent with the 'bus' characterization of customer engagement, which is designed to scale up (Baden-Fuller & Mangematin, 2013). But 'taxi' (project based) customer engagement, is not scalable, even when non-paying users create externalities or returns based on use and learning sources of scalability: these sources may reduce costs or increase appeal, but project-based engagement continues to require additional resources proportionate to the increased numbers of new paying customers. As far as value linkages are concerned, they too are subject to the scaling effects of engaging with non-paying users. The sources of scalability influence a network of suppliers, in the case of a networked value chain, in the same way as they influence a single focal supplier, in the case of a hierarchical value chain. For example, learning by using influences a network of suppliers - each with respect to its part of the product or service - in ways that are similar to its influence on single suppliers. These considerations can be summarized in:

> Proposition 1a: Business models that engage both non-paying users and paying customers are a scaling mechanism when a standardized *'bus'* form of customer engagement is used.

> Proposition 1b: Business models that engage both non-paying users and paying customers are a scaling mechanism when either a hierarchy or a network value chain is used.

Finally, we conclude our reasoning about customer identification and scalability by considering the countervailing factors of including non-paying users - those factors that either prevent the inclusion of non-paying users or that offset the positive effects of such users on scalability. The first and obvious factor is that while marginal costs may be close to zero, providing the fixed costs to support the extremely large number of non-paying users can be significant. For example, according to a trade publication, Google spent $7.35 billion in capital expenditures on its internet infrastructure during 2013 - the largest construction effort in the history of the data center industry[3].

The second counteracting factor relates to appropriation – non-paying users are included either to attract paying customers or to reduce unit costs. Given the cost of serving non-paying users, these complements are critical for the survival of the business. McGrath (2010) identifies six types of transactions that complement free offerings: advertising or other multisided mechanisms (non-paying users attracting paying customers); cross-subsidization (something is given away for free in the interest of profiting from another part of the business); promotion (a low cost good is given away to promote something else); freemium (a basic version is given free with the hope of customers paying for an advanced version); barter (a good is given away to customers who provide something of value in return); and the gratis or gift model (something is provided for free simply because those involved enjoy interacting or making a contribution). Without such transactions, serving non-paying users is not sustainable, even given the resulting scaling up.

The effect of non-paying users is illustrated by Splunk, a fast growing entrepreneurial company that provides technology infrastructure for big data analysis.

Illustration

Splunk was founded in San-Francisco in 2003, has revenues of US$268 million (2014), about one thousand employees and market capitalization of about US$8 billion. Splunk provides products that monitor and analyze data from existing IT infrastructures, such as websites, applications, servers, networks, sensors, and mobile devices. Splunk products capture, index and correlate real-time data in a searchable repository from which graphs, reports, alerts, dashboards and visualizations can be created so that clients can make better decisions based on the data, and troubleshoot operational problems rapidly.

Industry experts say that Splunk has about fifty non-paying users per each paying customer. Its products, which offer up to 500 megabytes of data per day for indexing, are available for free download. Free users integrate Splunk products into their IT infrastructures and self-customize them to fit their specific indexing and search needs, which provide Splunk with immense opportunities for learning and creating informational increasing returns. The company encourages non-paying users to report problems in using its products, so it is continually collecting data on new usage patterns and experiences with its new interfaces. The free user community has made Splunk well-known and a de facto standard for some indexing and search tasks for a variety of IT infrastructures.

This position, where Splunk has created a practical standard, increases the value of its products, as IT professionals often know them well, and find them easy to download and use.

Customer engagement and scalability

The customer engagement element of business model configuration concerns the value proposition the firm offers to its customers. This requires taking a customer perspective in order to identify the nature of their needs more accurately, and the quality of the value they experience. It also often involves taking decisions as to the degree of customization involved in the value proposition. This is important, as the more diverse or heterogeneous customers' needs and tastes are, the less scalable a business model will be if specific product features for specific needs are required. At one extreme is the bespoke service or product, created in a project style around particular customers' specific needs and wants. This 'taxi' style of business model contrasts with its opposite, the 'bus' system, involving a pre-designed and mass produced/delivered, one-size-fits-all, approach to the provision of customer value (Baden-Fuller & Mangematin, 2013; Baden-Fuller & Haefliger, 2013).

As with the previous (customer identification) business model configuration element, the way in which digital business models engage with customers is also distinctive. Prosumption or value co-creation (Prahalad & Ramaswamy, 2004) - where the customer participates in the production of the product/service they also consume (Toffler, 1980) - offers an alternative to the 'bus' vs. 'taxi' choice of customer engagement. This happens when the product or service itself is essentially generic - a 'bus' – but where mechanisms are made available for the user to customize it, and make it more bespoke – a 'taxi' – for themselves. The resulting type of customer engagement is perhaps analogous to that of a 'rental car', where resources are provided in an all-purpose way, but the customer is given the facility to direct and use them in ways specific to their needs. The Spotify music streaming service provides a good example of this development in customer engagement dimensions: through an iterative process of recommendations provided by the software and their own explorations, listeners co-create their own bespoke playlists.

We propose that such 'prosumption' cases relate to scalability through the operation of network effects on both the demand and supply sides. Firstly, involving the customer in producing value has the potential to create significant learning by using, improving both the generic form of the product or service and its specific components, via the tasks carried out by the customer. Such learning should increase the number of users, the value produced and ultimately the scale of future adoptions. A similar reasoning is true from an informational increasing returns perspective - that prosumption makes a technology better understood, resulting in a larger number of users and a greater scale of operation.

Secondly, prosumption promotes scale economies of production: as the customer is required to do some of the production work themselves, the cost of satisfying specific or expensive needs and tastes are minimized. Passing on the costs of customization to the user enables the firm to avoid the expensive and difficult- to-scale knowledge work that would be required. So such 'rental car' engagement systems can make such scalable customized value propositions achievable. Table 2 illustrates this discussion by detailing the assumptions and capabilities involved, and the effects that follow. This discussion can be summarized thus:

Proposition 2: Business models that engage customers using a prosumption (self-customization) approach promote scalability in digital businesses.

Table 2: Sources of scalability for prosumption (rental car) business models

Source	**Assumptions**	**Required Capabilities**	**Scalability Effects**
Learning by using	Customer needs are idiosyncratic and they wish to customize	Support customization; disseminate learning between users	Self-customization creates learning, improves value and increases adoption
Network externalities	Customers accept additional value created by others through customization	Manage and disseminate value created by customers' customization	Customization by some users increases value of the product to other users and increases adoption and scale
Scale economies in production & distribution	The product can be divided into a generic base and a range of open, customisable features	Provide intuitive and easy to use tools to users to self-customize	Reduced production costs of the generic base enable lower prices and more adoption and scale
Informational increasing returns	Customization requires user confidence is increased through adoption and customization by others	Disseminate customization know-how for example through online communities	Customization increases understanding of the product, its value, adoption and scale
Technological interrelatedness	Other technologies emerge to support the task of customization	Modular design and APIs are required; should enable interoperability with related products/services	An eco-system of related technologies increases the ease of adoption and use; it increases barriers to entry and raises switching costs
Distributed resourcing	Self-customization enables users to distribute resources	Allow customization of resources	May reduce resourcing bottlenecks

We now consider the interaction between prosumption customer engagement and the two other business model elements - customer identification and value chain linkages. Both paying customers only and a mix of non-paying users and paying customers are possible structures for prosumption business models. Non-paying users can participate in the production of the service and increase learning, network externalities, etc.: and those who self-customize increase the possibility of scalability. On the other hand, non-paying users are not essential for prosumption by paying customers, so the prosumption model is viable for both customer identification configurations. Similarly, for hierarchical and network value chains - both are possible, but not essential for prosumption by customers to be part of the business model design. Network value chains may also allow different varieties of prosumption or self-customization related to different complementary products or services that are provided in conjunction with the focal technology or platform. These arguments are summarized by Propositions 2a and 2b.

> Proposition 2a: A prosumption customer engagement is a scaling mechanism when the business model allows access and use to either paying customers only or to a mix of paying customer and non-paying user.

> Proposition 2b: A prosumption customer engagement is a scaling mechanism when the business model uses either hierarchical or network value chains.

Finally, we conclude the discussion about customer identification and scalability by noting that there are counteractive factors that can make prosumption or co-creation difficult to pursue. First, this configuration assumes that customers have the high level of competence required to engage in self-customization. Increased participation requires skill and knowledge, and users need to have good sectoral knowledge combined with sufficient familiarity with the various software tools and interfaces involved. In addition, the ‘rental car’ approach is viable when it competes on price with project-based (taxi) models, and with pre-designed (bus) systems via the flexibility of its design and use. Second, Prahalad and Ramaswamy (2004) emphasize that co-creation requires that the focal firm has new managerial attitudes, processes and products in place, which allows value creation to be located in the consumer-company interaction and not just in the company itself. For digital businesses, forming the “information infrastructure must be centered on the consumer and encourage active participation in all aspects of the co-creation experience, including information search, configuration of products and services, fulfilment, and consumption.” (Prahalad & Ramaswamy, 2004; p. 11). (These ideas are exemplified by the empirical illustration that follows.) Finally, recent research about such service co-creation shows that it increases service complexity and the possibility of service failure. Customers are also likely to formulate higher-quality expectations and thus experience correspondingly high levels of disappointment if performance is poor (Heidenreich et al., 2014).

The following example illustrates proposition 2. The ability to self-customize is central to the business model of EDITED - without self-customization, its millions of data items would be useless to its customers.

Illustration

EDITD was founded in 2009 and positions itself as the world's biggest data warehouse for clothing. It has 20 employees and it has recently secured £2.6m from private equity in order to expand (FT, 2014). EDITD monitors the websites of over 4,000 fashion retailers to track the life cycles of more than 50 million garments. It also follows 2,000 blogs, Twitter and Tumblr accounts in order to combine "the knowledge of over 800,000 thought-leaders, key influencers and fashion experts giving you an instant source of inspiration and insight into the hottest trends and opinions".

EDITD customers are fashion professionals who utilize the large variety of data the company provides by using EDITED pre-designed interactive tools to customize the data to their own specific business needs. The vast amount of data involved (e.g. prices of more than 50 million garments) is worthless without detailed customization to fit each customer's specific fashion product lines: so it is important that customers can do this themselves without engaging with EDITD personnel. Traditional consultancies customize similar information by selling bespoke reports that are developed by editorial and consultancy staff. In contrast, EDITD leverage the customers' own knowledge of their businesses and the now universal computer literacy and web-based interactive tool know-how to allow them to prosume this service.

The result is high scalability through learning by using, creating positive network effects and informational increasing returns on the demand side, as well as reducing development and production costs on the supply side. EDITD reports that large retailers such as Gap and Target use its services, a significant achievement for such a young and small company.

Value chain links and scalability

Value chain linkages - also called information flow architecture or governance systems - refer to mechanisms the firm uses to deliver its products or services to its customers. Platforms play a prominent role in digital businesses (Yoo et al., 2012), and the literature has distinguished three different types of platforms: supply-chain, industry (Gawer, 2014) and multi-sided platforms (Eisenmann, et al., 2006, 2011; Haigu & Wright, 2011). This body of work focuses on innovation and competition, and on economies of scope (Gawer, 2014). In this section we argue that scalability, or economies of scale, is also central to platform based business models.

When considering the configuration of value-chain linkages, it is useful to make a distinction between hierarchical - or vertically integrated - and network governance approaches. Hierarchical governance systems use administrative fiat and decision-making protocols (policies), while network governance approaches adopt a more open style of decision-making that extends the task of organizing the value chain beyond the firm's boundaries to include firms producing complementary products and services (Lorenzoni & Baden-Fuller, 1995). We argue that the network governance mode of platforms promotes scalability, while hierarchical governance does not.

Network, or open, value chains have the potential to generate scale in a number of interlinked ways. First, opening up the firm's value chain brings in greater diversity in terms of engaging more partner firms, suppliers and user communities: and when network effects act on a wide range of participants their impact is magnified. Second, network value chains often allow technical openness that promotes learning by using and informational increasing returns (von Hippel, 1976, 2005). An extreme example of open governance is open source software - with no barriers to its use, larger numbers of users interact and generate informational increasing returns by sharing their knowledge and the results of their work (West, 2003; Dahlander & Magnusson, 2008). The positive reputational rewards that can be accrued by firms that open up their value chains in this way further enhance scalability: the costs of picking a partner are not trivial, so a firm with a positive reputation is likely to attract more partners and thus support the creation of an ecosystem in which both it and they are embedded.

Third, network value chains can reduce the costs of production and distribution by enabling large numbers of new customers to be serviced without having to make major investments in enlarging capacity. For example, the owner of a multi-sided market benefits from products and services produced and distributed by one side– sellers – to the other – buyers – at minimal cost to the platform owner, whose activity is simply matching the two sides (Haigu & Wright, 2011). Finally, an open model can reduce the cost of improving and developing a product or service by engaging a user community (von Hippel, 2005), adopting modular design principles (Baldwin & Clark, 2000) and supporting combinatorial innovation (Yoo et al., 2012). Table 3 details these considerations by illustrating the assumptions, capabilities and effects for each of these scalability sources, while Proposition 3 summarizes this discussion.

Proposition 3: Business models that orchestrate network value chain linkages promote scalability in digital businesses.

Table 3: Sources of scalability for networked value chains business models

Source	Assumptions	Required Capabilities	Scalability Effects
Learning by using	The platform has the potential to facilitate additional suppliers and improved products	Managerial attention and resources for all partners; quick support for improvements by suppliers	Increase learning by suppliers, resulting in improved products and further appeal of the platform
Network externalities	Larger numbers of suppliers and products increase the appeal of the platform	Reputation and know-how about managing platforms and orchestrating eco-systems	Increase the number of suppliers and expectations for a large and high value network
Scale economies in production & distribution	Suppliers provide non-generic functionality	Design and maintain a generic platform; prevent non-scalable features becoming part of the platform	Allow the platform owner to provide generic and scalable functionality
Informational increasing returns	Suppliers disseminate information about the platform	Well-designed technical and commercial interfaces; accurate, up-to-date information about them	As the number of suppliers grows the platform becomes better known and understood
Technological interrelatedness	External suppliers can create complementing technologies	Modular design and APIs are required; orchestration of the platform-related ecosystem	Enables the introduction of sub-technologies that increase the platform's appeal and scale
Distributed resourcing	Suppliers may provide distributed resources and practices	Design of distributed resources and practices; ability to add resources rapidly or to orchestrate the suppliers' resources	Takes advantage of the distributed resources of multiple suppliers to maintain performance when scaling up

We now consider the interactions between network value chains and the two other business model elements - customer identification and engagement. Business models that engage with non-paying users as well as paying customers are popular configurations for platform based business models. The two elements enhance scalability – one by greater openness towards non-paying users and the other by greater openness towards suppliers. The interaction between them involved in having non-paying buyers and sellers - an interaction that has been characterized as the search engine business model (Baden-Fuller & Haefliger, 2013) - enhances scalability via both mechanisms. The other possibility of a platform based on a simple business model (excluding non-paying users) is also feasible: cinemas and dating sites fit this configuration.

Where the customer engagement element is considered, the logic of network value chains – platforms, multi-sided markets – works fine with a 'bus' type engagement as well with the 'rental-car' or prosumption configuration. However, a platform that engages on a 'taxi' or project basis cannot contribute to high scalability. It is possible to match buyers to sellers who then engage on a project basis (with the former selling bespoke services to the latter). However, the logic of a platform itself is that of a bus – matching buyers and sellers on the basis of a pre-designed offering - as Propositions 3a and 3b summarize:

> Proposition 3a: A network value chain is a scaling mechanism when either i) only paying customers or ii) a mix of paying customers and non-paying users is used.
>
> Proposition 3b: A network value chain is a scaling mechanism when either a standardiz*ed '*bus*' or 'rental car'* customer engagement is used.

Finally, we conclude our reasoning about networked value chain linkages and scalability by discussing the counteractive factors and tensions that present themselves when trying to build scalability into a business model. These relate to the capabilities required, the reduced opportunities for value appropriation and the increased competition that can occur with open value chain linkages.

Iansiti and Levien (2004) discuss the difficulties that platform owners face in creating and maintaining healthy ecosystems, which include developing mechanisms to allow network participants to connect easily and to create new products and services efficiently. Although not a standard example - given its dominance as a seller - the difficulties related by Amazon's founder Jeff Bezos in making its shop into a multi-sided market are telling (Kirby & Stewart, 2007). His account reveals how persistence and relentlessness were necessary to overcome fears of cannibalization, and the long process of experimentation and failure involved in finding a way to design the market to include both Amazon and external sellers.

In addition, a platform owner must learn how to share value between ecosystem participants. For example, eBay realized early on that it needed to charge its sellers well below the typical margins that most retailers would charge (Iansiti & Levien, 2004). In contrast, Groupon struggled to sustain its revenue and profit following its initial rapid scale up, as it failed to share value with the companies providing discounted offers by insisting on deep discounts and charging high participation fees (Edelman et al., 2011).

A further case of tension between monetizing the increased value created by scale and ensuring ease of adoption and continued diffusion occurs when open value chains introduce intra-system competition. Such competition between platform owners and their own complementors can depress profits and deter future investment (Katz & Shapiro, 1985 1994). A good example of this danger is evident in the case of Facebook Home. In 2013, Facebook launched a new service using Google's Android software to establish a home page designed social media use on smartphones that included a search functionality which competed against Google's search engine.

A final case of the adoption vs. appropriability trade-off (Gawer, 2014; Boudreau, 2010; West, 2003) facing highly scalable business models is the attempt to translate use into revenue by charging existing users for a service. Online newspapers, for example, continue to struggle with the design of their paywalls - payment systems that restrict or meter access to their content. Shazam is a good illustration of open value chain linkages as it builds on both mobile and music technology platforms to create a growing multi-sided market.

Illustration

Shazam, a music recognition service, was launched in 2002 in London - by 2013, it reported 400 million users with £31 million revenue and annual growth of 50%. The service is a multi-sided platform that matches buyers (who wish to identify a music track) with sellers (music owners). Shazam users are asked to hold up their phones to a music speaker and the software identifies the music being played, and includes a link that allows users to buy the track from music service providers such as Spotify, Google Play, and Amazon. In 2009 Shazam reported that around 8% of users purchased the track after it had been identified. The fact that the basic service is free of charge and available as a mobile application has created Shazam's enormous scale. On the demand side, there are indirect network externalities in this type of service – Shazam's central database has over 11 million music tracks which are licensed from music labels. The larger the database, the more probable is accurate music recognition, and the more customers are attracted, thus making Shazam a better partner to the music labels. On the supply side, Shazam is not required to produce, market or distribute music. Its relatively small resources (around 300 employees) can focus on developing its mobile application and on its propriety music recognition algorithms: these activities increase scale economies on the supply side.

Furthermore, with investment, data, technology and a large user base, Shazam not only uses data but also creates it. As usual with IT, Shazam registers all transactions and interactions, so 'informating' or generating new streams of information (Zuboff, 1988). Shazam is used more than 15 million times a day, and uses the data it harvests to predict which new artists will gain mainstream attention the following year.

Discussion and Conclusion

Summary and Contributions

The primary goal of this paper is to draw scholarly attention to issues of business model scalability. What does it mean to be scalable? What are the sources of scalability? How can a business model be designed to achieve scale? Given the importance of scalability, particularly in digital businesses, it is surprising that there has been little discussion of the issue in the business model literature so far: this paper contributes to filling that gap.

First, we have updated the notion of business model scalability to highlight that it is the configuration of business model elements that may increase levels of value creation and appropriation when the number of users and customers increases and their needs change, but without adding proportionate extra resources.

Second, by combining managerial, economic and technological perspectives, we have examined the sources of scalability including both supply and demand-related economies of scale. We suggest that scalability is enhanced by learning by using, network externalities, informational increasing returns, technological interrelatedness, and distributed resourcing. Although these have been discussed before (e.g., Arthur, 1988; Chandler, 1990; Shapiro & Varian, 1999; Gawer, 2014), framing them in terms of scalability contributes to a richer understanding of the drivers of scalable digital businesses.

Third, we have unpacked the characteristics of scalable digital businesses by considering three core business model design elements. Business model design can be seen as the configuration of a set of activities that not only creates value for customers but also allows firms to capture value (Zott & Amit, 2007). Although the organization literature implies that the configuration of these activities needs to be aligned with conditions in the external environment (Siggelkow, 2002), the critical condition of scaling - how a business model can continue to meet changing customer demands as the firm increases its customer base – has been under-explored. We contribute to business model literature by proposing that scalability is increased by business models that engage both paying and non-paying users (whether single or multi-sided); that allow customers to participate in the production process; and that select network governance for value chain linkages. We argue that scalability is not enhanced by simple, single sided or hierarchical business models, and is hampered by project-based (or taxi) customer engagements.

In sum, by looking through this newly articulated scalability lens, we gain further insights into how businesses scale up. As a concluding illustration, consider how Google demonstrates the application of the three business model mechanisms discussed in this paper.

Concluding illustration

Brin and Page's famous paper (1998) 'The Anatomy of a Large-scale Hypertextual Web search Engine' starts with identifying the growing amount of data on the web as well as the number of users as a scalability issue. They contrast Google's algorithm with Yahoo's " human maintained lists" and reason that the latter "cover popular topics effectively but are subjective, expensive to build and maintain, slow to improve, and couldn't cover all esoteric topics" (p. 107). They explain that keyword-matching search engines usually returned too many low quality matches, and that some advertisers mislead search engines to gain people's attention. They predict that the "top search engine will handle hundreds of millions of queries per day by the year 2000" and conclude that "the goal of our system is to address many of the problems, both in quality and scalability, introduced by scaling search engine technology to such extraordinary numbers" (p. 108).

Their suggested solution was to calculate the number and importance of a web page's citations, or links, and to use this PageRank method to prioritize the results of keyword searches. In other words, Brin and Page opened the value chain of their Google search engine to the community of web page authors. As a web page links to another web page, the target page's PageRank is increased, and this increase is proportional to the importance (or PageRank) of the source web page. The quality and scalability of Google's search engine is thus based on this opening of the value chain. However, Google's revenue is also rooted in a business model designed on a relationship between both non-paying users and paying customers and in the quality and scalability of its advertisement services. From January 2000, Google sold text ads, priced per thousand impressions, through sales representatives - but this move didn't produce much revenue and was not scalable (Battelle, 2005). After the dot-com bubble burst, Google introduced a series of upgrades that were copied from competing search engines (e.g., GoTo.com) and then improved. The 2002 version of AdWords already included pay-per-click pricing, and was supported by an auction-based self-service mechanism (Battelle, 2005). Removing the human sales representatives and introducing the possibilities of self-customization - of both content and price - created the opportunities for Google to scale up. There were economies of scale in production and distribution, because each additional advertising customer added only a negligible cost. Indirect network externalities – more users, more focused and thus more useful advertisements – had the opportunity to kick in. Finally, advertisers could start promoting their businesses at minimal cost, learning and better understanding this new media while growing their advertisement budgets. These and other technology and business model design decisions - covering user identification, user engagement and value chain linkages - created Google's current scale, to the point where it handles hundreds of millions of queries per minute.

To summarize, Google's successive search engine improvements opened its value chain to the decisions of web page authors about which pages to hyperlink to. Its business is based on non-paying users targeted by paying advertisers, and it allows these customers to self-customize their ads and their criteria for accessing users.

Managerial implications

The analysis in this paper presents managers with a framework for considering what scalability could mean for their businesses and how they might achieve it. These considerations have become increasingly important, since venture capital is keen on funding scalable businesses and shuns ventures that appear to lack such potential. Table 4 summarizes these considerations for digital businesses.

Table 4: Summary of managerial implications

Business model design element	**Implications for managers**	
	Sources of Scaling	**Countervailing factors**
<u>Customer identifications</u> Engaging both non-paying users and paying customers	to test the product in a variety of contexts in order to increase reliability and add functionality to reduce marketing costs, for example by viral marketing to create incentives for suppliers to develop complementing products to utilize users' own computing and communication gear as a way to distribute data resources	may be costly because of their large numbers may be costly to support as their preferences may differ from paying customers requires striking a balance with paying customers in terms of managerial attention and resources
<u>Customer engagement</u> Allowing prosumption (i.e. self-customization)	to cater for specific customers' needs to learn about new configurations of the product as a basis for new product features to reduce development costs by allowing customers to develop further functionalities to allow customers to select data resources to fit their computing and communication gear	is not always feasible requires deep understanding of customers' needs and capabilities may reduce reliability and consistency of the product may be costly and complex to support as the number of product configurations grows with each customer
<u>Value chain links</u> Orchestrating networked value chain linkages	to make eco-system partners learn new possibilities for the core technology to create a platform for suppliers to provide complementing products to reduce costs by inducing eco-system partners to participate in marketing and distribution to utilize the resources of eco-system partners to increase resource distribution	requires specific eco-system management capabilities may be costly to develop and maintain these capabilities may create competition to the focal product

A further consideration for managers and entrepreneurs is the dynamic between scaling as 'opportunity' vs. 'imperative'. Whilst we suggest that scaling is a choice, the global scale that some digital businesses have achieved may seem to demonstrate that scaling up is actually a business imperative. Because scale economies in production and distribution are inherent in digital technology, businesses that do not scale up can be seen as inefficient, costly and thus bound to fail when competing against those that do. The economies of scale in demand further increase this imperative. However, scale in digital businesses can be also counterproductive, unprofitable or inconsistent with other goals. For example, scale destroyed the deep price discrimination logic of Groupon: as the number of users increased they weakened the value proposition of offering services only to consumers looking for bargains (Edelman et al., 2011). Thus, we may conclude that scale is not an absolute imperative, but an opportunity that should be seized on when it arises. Further, the extent to which a business model might be configured to take advantage of the scaling dynamics we have identified will necessarily depend on the degree to which the dimensions of its value creation can be operated digitally. Current experimentation in the division between the physical and the digital aspects of business models (for example by companies like Uber) is extending the relevance of our analysis beyond the digital sector of the economy.

Limitations and future research

Given the restricted length of this paper, we have been unable to develop several important issues relating to scale and scalability. For instance, the relationship between scope and scale has been excluded. The literature has acknowledged the role of scope - the variety of complementary products and services that a company offers - in the likelihood of innovations being generated (Gawer, 2014). We have only addressed scope indirectly. Some business models - for example multi-sided markets - may require a range of products to be part of the value proposition in order for the operation to scale, but we have not examined this relation in detail. Another important scalability issue is speed - the time taken to reach scale. It appears crucial in many digital businesses, because of the fast diffusion and adoption rates enabled by new communication technologies and related business models. Is speed a function of a scalable business model design? Is speed a key driver to achieving scale? We suggest this is a fruitful area for future research.

A further limitation of this paper concerns the relationship between firm strategy and business model design for scalability. Important questions remain about how different firm strategies might affect the ability of a business model to scale up: future research could investigate which strategies accelerate and which degrade a firm's ability to realize the scalability of its technology and its business model. Another key issue relates to the value of business model scalability, which will require assessing a business model's sustainability and profitability. For example, the design of self-customization services may involve significant development costs. Similarly, as previously discussed, there may be significant difficulties in creating and maintaining networked business models, as their construction can often also increase the likelihood of competition (Gawer, 2014). These factors should be taken into account while designing scalable business models, as they may limit a firm's ability to both scale up and sustain a profitable business.

Overarching theme

As we conducted the analysis and built our argument it became clear that there is a common feature across each of the three mechanisms by which digital business models scale up: they all implied the replacement of direct control by the focal firm with orchestrated contributions from external actors. So it seems that scalable business models in the digital sector are not constructed entirely within firm boundaries - rather they involve establishing relationships and interactions that extend across them. The mechanisms of scalable business models are articulated through the active participation of a range of different kinds of partners - free users, paying customers, prosumers and suppliers of complementary products and services.

This characteristic of scalability is made possible by the immense reduction in production and distribution costs that follow from the non-materiality of digital products and services. With appropriate configurations of business model resources and managerial attitudes, scale need not be constrained, and can be achieved through the business model design elements we have described. Scalability is also fuelled by the low cost of accessing large markets of users and producers. The pervasiveness of digital communication and the rapidly spreading infrastructure of digital products and services ensure that large networks of interconnected actors exist and can be accessed easily.

One of the consequences of scalability through orchestration is rapid growth. While control requires greater resources and oversight, in many cases orchestration involves the enrolling of resources owned by others. The self-customizing user brings their skill, the platform partner their knowledge and connections and the free user their own communication and computing assets. This creates the opportunity for extremely rapid growth, unencumbered by the requirement to match growth with internal firm specific resources and the creation of global digital near-monopolies.

Acknowledgements

We thank Professor Vincent Mangematin and Professor Charles Baden-Fuller, the editors of this volume, for their motivating and insightful comments and encouragement, and those of the two anonymous reviewers, We are grateful for the feedbacks received at EGOS Colloquium 2013 and 2014, as well as workshops held at Cass Business School, City University of London and from colleagues at Cass, Grenoble École de Management and UEA, especially: Stefan Haefliger, Valerie Sabatier and Alessandro Giudici.

References

Arthur W.B. (1988). Competing Technologies: An Overview, Technical Change and Economic Theory, ed. Dosi, et. al., New York: Columbia University Press, pp. 590-607.

Arthur W.B. (1989). Competing technologies, increasing returns, and lock-in by historical events. The Economic Journal, **99**(394): 116-131

Baden-Fuller C. & Haefliger S. (2013). Business models and Technological Innovation, Long Range Planning, **46**:419-426.

Baden-Fuller C. & Morgan M. (2010). Business Models as Models. Long Range Planning. **43**(2/3):156-171.

Baden-Fuller C. & Mangematin V. (2013). Business Models: A challenging agenda, Strategic Organization. **11**(4):418-427.

Bakos Y. & Brynjolfsson E. (1999). Bundling information goods: Pricing, profits and efficiency. Management Science. **45**(12): 1613–30.

Baldwin, C.Y. & Clark, K.B. (2000). Design Rules: The Power of Modularity, vol. 1. MIT Press, Cambridge, MA.

Baran P. (1960). Reliable digital communications systems using unreliable network repeater nodes. Rand Corporation Published research papers. Accessed 26 Jan 15 http://www.rand.org/pubs/papers/P1995.html

Battelle J. (2005). The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture. Penguin.

Bharadwaj A., El Sawry O.A., Pavlou P.A. & Venkatraman, N. (2013). Digital Business Strategy: Toward a Next generation of Insights, Management Information Science Quarterly, **37**(2): 471-482.

Boudreau K. (2010). Open Platform Strategies and Innovation: Granting Access vs. Devolving Control. Management Science. **56**(10): 1849-1872.

Brousseau E. & Penard T. (2007b). The Economics of Digital Business Models: A framework for Analyzing the Economics of Platforms. Review of Network Economics, **6**(2):81-114.

Brin S. & Page L. (1998). The Anatomy of a Large-scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems, **30**(1-7):107-117.

Burt R.S. (1987). Social Contagion and Innovation: Cohesion versus Structural Equivalence. American Journal of Sociology, **92**(6):1287-1335.

Chandler A. (1990). Scale and scope: the dynamics of industrial capitalism. Harvard University Press.

Chandrashekaran M., Rajdeep Grewal R. & Mehta R. (2010). Estimating Contagion on the Internet: Evidence from the Diffusion of Digital/Information Products. Journal of Interactive Marketing. **24**: 1–13.

Chesbrough H. (2003). Open Innovation: The New Imperative for Creating and Profiting from Technology. Boston: Harvard Business School Press.

Chesbrough H. (2010). Business Model innovation: Opportunities and Barriers. Long Range Planning, **43**(2-3): 354-363.

Church J. & Gandal N. (1992). Network Effects, Software Provision, and Standardization. The Journal of Industrial Economics, **40**(1): 85-103.

Dahlander L. & Magnusson M. (2008). How do firms make use of open source communities? Long Range Planning, **41**:629-649.

Duboc L., Rosenblum D.S. & Letier, E. (2010). Death, Taxes, & Scalability. IEEE Software, **27**(4): 20-21.

Duboc L., Letier E. & Rosenblum D.S. (2013). Systematic elaboration of scalability requirements through goal-obstacle analysis. IEEE Transactions on Software Engineering, **39**(1): 119-140.

Edelman B., Jaffe S. & Kominers S. D. (2011). To Groupon or not to Groupon: The profitability of deep discounts, Harvard Business Review, Working paper, No. 11-063.

Eisenmann T., Parker G. & Van Alstyne M. (2006). Strategies for two-sided markets. Harvard Business Review, **84**(10), 92–101.

Eisenmann T., Parker G. & Van Alstyne M. (2011). Platform envelopment. Strategic Management Journal, **32**(12): 1270-1285.

Evans D.S. (2003). Some Empirical Aspects of multi-sided platform industries. Review of Network Economics, **2**(3).

Faraj S., Jarvenpas S.L. & Majchrzak A. (2011). Knowledge Collaboration in Online Communities. Organization Science, **22**(5): 1224-1239.

Frankel M. (1955). Obsolescence and technological change in a maturing economy. American Economic Review, **45**: 296-319.

Garud R. & Kumaraswamy A. (1995). Technological and Organizational Designs for realizing economies of substitution. Strategic Management Journal. **16**: 93-109.

Gawer A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. Research Policy. **43**:1239-1249.

Gawer A. & Cusumano M. (2002). Platform Leadership: How Intel, Microsoft and Cisco drive industry Innovation. Harvard Business School Press: Boston.

Haigu A. & Wright J. (2011). Multi-sided platforms. Harvard Business School working paper, 12-24.

Heidenreich S., Wittkowski K., Handrich M. & Falk T. (2014). The dark side of customer co-creation: exploring the consequences of failed co-created services. Journal of the Academy of Marketing Science, 1-18.

Hitt, L. M. & Chen, P. (2005). Bundling with Customer Self-selection: A simple approach to bundling low marginal-cost goods. Management Science. **51**(10):1481-1493.

Iansiti M. & Levien R. (2004). Strategy as Ecology. Harvard Business Review, **82**(3): 68-78.

Katz M. & Shapiro C. (1985). Network externalities, competition and compatibility, American Economic Review, **75**(3): 424-440.

Katz M. & Shapiro C. (1994). Systems competition and Network effects. Journal of Economic Perspectives, **8**(2): 93-115.

Kirby J. & Stewart T.A. (2007). The Institutional Yes. Harvard Business Review, **85**(10): 74–82.

Leibenstein H. (1950). Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand. The Quarterly Journal of Economics, **64**(2): 183-207.

Lorenzoni G. & Baden-Fuller C. (1995). Creating a Strategic Centre to Manage a Web of Partners. California Management Review, **37**(3):146-163.

Markides C. (2013). Business Model innovation: What can the ambidexterity literature teach us?

Academy of Management Perspectives, **27**(4): 313-323

McAfee A. & Brynjolfsson E. (2012). Big Data: The Management Revolution. Harvard Business Review, **90**(10): 60-68.

McGrath R. (2010). Business Models: A Discovery Driven Approach. Long Range Planning, **43**(2-3): 247-261.

McKinsey Global Institute, May (2011). Big data: The next frontier for innovation, competition, and productivity.

Miller D. (1986). Configurations of Strategy and Structure: Towards a synthesis. Strategic Management Journal, **7**(3): 233-249.

Miller D. (1996). Configurations Re-visited, Strategic Management Journal, **17**: 505-512

Parmentier G. & Mangematin V. (2014). Orchestrating innovation with user communities in the creative industries, Technological Forecasting and Social Change, **83**: 40-53.

Porter M. (1996). What is strategy? Harvard Business Review, **74**(6): 61-78

Prahalad C.K. & Ramaswamy V. (2004). Co-Creation Experiences: The Next Practice in Value Creation, Journal of Interactive Marketing, **18**(3): 5–14.

Rathmell J.M. (1966). What is meant by Services? Journal of Marketing, **30**:32-36.

Raymond E. (1999). The Cathedral and the Bazaar, Knowledge, Technology and Policy, Fall, **12**(3): 23-49.

Ritzer G. & Jurgenson N. (2010). Production, Consumption, Prosumption The nature of capitalism in the age of the digital 'prosumer'. Journal of Consumer Culture, **10**(1), 13-36.

Rochet J-C. & Tirole J. (2003). Platform Competition in two-sided markets. Journal of European Economic Association. **1**(4): 990-1029.

Rogers E.M. (1962). Diffusion of innovations. Free Press of Glencoe, New York.

Rohlfs J. (1974). A theory of interdependent demand for a communications service. Bell Journal of Economics, **5**: 16-37.

Rosenberg N. (1982). Inside the black box: Technology and Economics. Cambridge, Cambridge University Press.

Schilling M.A. (2002). Technology Success in Winner-Take-All Markets: The Impact of Learning Orientation, Timing and Network Externalities. Academy of Management Journal, **45** (2): 387-398.

Shapiro C. & Varian H.R. (1999). Information Rules. A strategic Guide to the Information Economy. Harvard Business School Press: Boston Mass.

Siggelkow N. (2002). Evolution towards fit. Administrative Science Quarterly, **47**: 125-159.

Spence A.M. (1981). The learning curve and competition. Bell Journal of Economics, **12**: 49-70.

Tapscott D. & Williams A.D. (2006).Wikinomics: How Mass Collaboration Changes Everything. New York: Portfolio.

Teece D. (1980). Economies of Scope and the Scope of the Enterprise. Journal of Economic Behavior and Organisation. **1**:223-247.

Teece D. (2010). Business Models, Business Strategy and Innovation. Long Range Planning, **43**(2-3):172-194.

Toffler A. (1980).The Third Wave. New York: William Morrow.

von Hippel E. (1976). The dominant role of users in the scientific instrument innovation process. Research policy, **5**: 212-239.

von Hippel E. (1986). Lead users: A source of novel product concepts. Management Science, **32**(7): 791-805.

von Hippel E. (2005). Democratising Innovation. MIT Press: Cambridge, MA.

Weill P. & Woerner S. (2013). Optimising your digital business model. MIT Sloan Management

Review, Spring: 71-78.

West J. (2003). How Open is Open Enough? Melding Proprietary and Open Source Platform Strategies. Research Policy, **32**(7): 1259-1285.

Yoo C.S. (2012). Modularity Theory and Internet Policy. Faculty Scholarship. Paper 470. http://scholarship.law.upenn.edu/faculty_scholarship/470. Accessed 24 Sept 14.

Yoo Y., Boland R.J., Lyytinen K. & Majchrzak A. (2012). Organizing for Innovation in the Digitized World. Organization Science, **23**(5): 1398-1408

Zhu K. (2004). The Complementarity of Information Technology Infrastructure and E-Commerce Capability: A resource-based assessment of their business value. Journal of Management Information Systems, **21**(1): 167-202.

Zott C. & Amit R. (2007). Business model design and the performance of entrepreneurial firms. Organization Science, **18**: 181–199

Zott C. & Amit R. (2010). Business Model Design: An Activity System Perspective. Long Range Planning, **43**(2-3):216-226.

Zuboff S. (1988). In the Age of the Smart Machine: The future of work and power. Basic Books.

Biographies

Joanne Jin Zhang is a Lecturer in Strategy & Entrepreneurship at UEA. She has worked at Cambridge University and City University of London. Her research and teaching interests focus on building scalable and sustainable businesses. She has gained her PhD and MBA from City University, London**.** Before joining academia, she co-founded a business in China. Norwich Business School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK tel: +44 (0) 16 03597310 e-mail**:** joannejinzhang@gmail.com

Yossi Lichtenstein has held posts at the University of East Anglia, Israel's College of Management, University College Dublin, IBM Research and Hewlett-Packard Laboratories. He has published about economic aspects of software development and has co-authored patents and research papers about hardware design automation. Norwich Business School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ UK tel: +44 (0) 16 03592263 e-mail: yossi.laor@gmail.com

Jonathan Gander has held posts at the Universities of East London, Westminster, Kingston and East Anglia. His research interests focus on the organisation, management and competitive strategies of firms within the creative economy. He is currently Director of the EMBA at the London College of Fashion. London College of Fashion, University of the Arts, London. 20 John Prince's Street, London W1G 0BJ UK tel: +44 (0) 20 75147400 e-mail: j.gander@fashion.arts.ac.uk