C

Capability and Multinational Enterprises (MNEs)

David J. Teece
Berkeley Research Group, LLC, Emeryville, CA, USA
Haas School of Business, University of California, Berkeley, Berkeley, CA, USA

Abstract

Academic approaches to the activities of the MNE have been dominated by the internalization theories developed in the field of ▶ international business (IB), including the envelope approach known as the ▶ eclectic paradigm. These theories are helpful but have several shortcomings, including an emphasis on the cost of investment decisions rather than on the opportunities they can create, a limited focus on managerial decision-making, and no model of the creation and maintenance of firm-level advantage. These problems can be addressed by a capability-based approach, such as the dynamic capabilities framework.

Definition The activities of the multinational enterprise (MNE) frequently involve the extension, leveraging, and creation of capabilities across borders. However, the literature on cross-border investment is dominated by theories based on concepts that ignore capabilities, such as transaction costs. Non-capabilities-based theories reveal little about the use of strategy to build competitive advantage. Capability-based explanations for MNE activities received relatively little attention until the 2000s.

Since the 17th-century appearance of globe-spanning trading companies such as the Dutch East India Company, the global economy has been knit ever more tightly together by multinational enterprises (MNEs). The spread of cross-border trade and investment accelerated in the latter half of the 20th century. Today, some start-ups are born as "international new ventures" and establish an overseas presence from the very outset (Oviatt and McDougall 1994).

"Explanations" for the multinational form of organization abound, particularly from an economic perspective. These include theories of interest rate arbitrage (Aliber 1970), multiplant economies (Caves 1980), the extension of monopoly power (Hymer 1976), and the minimization of transaction costs (Hennart 1982). Researchers in the field of ▶ international business (IB) developed theories that were somewhat better grounded, such as internalization to offset possible market failures (Buckley and Casson 1976). The ▶ "eclectic paradigm" (Dunning 1988) combined internalization with Hymer-esque ownership advantages and location-specific factors to yield a richer and more robust model of the determinants of ▶ foreign direct investment (FDI) and multinational business activity.

© Macmillan Publishers Ltd., part of Springer Nature 2018
M. Augier, D. J. Teece (eds.), *The Palgrave Encyclopedia of Strategic Management*,
https://doi.org/10.1057/978-1-137-00772-8

Other theories of FDI have come to emphasize knowledge transfer (Teece 1976; Kogut and Zander 1993) or, more generally, capability development and transfer (Teece 1986, 2014). Capability-based theories have the potential to "explain" not only the existence of multinationals and their investment patterns but also how they exploit their internal and external networks to differentiate themselves from rivals in the effort to build ▶ sustainable competitive advantage (SCA).

Non-capabilities Approaches

Many of the non-capabilities theories about the MNE and FDI have their own entries in this encyclopedia. The two leading theories, which will be briefly described here, are known as "internalization" and "the eclectic paradigm."

The internalization perspective has dominated much of the literature on the MNE over the past 30 years (Dunning and Lundan 2008). It is primarily associated with the claim that FDI is a response to transaction cost issues that cause market-based contracting to be less attractive. A less discussed variant claims that internalization occurs because of the relative transaction efficiency of internal resource transfers and learning (Teece 2014).

The transaction cost approach to the MNE, which was advanced by Buckley and Casson (1976), Rugman (1981), Teece (1981), and others, focused on potential problems arising from asset specificity and renegotiation risk that cause possible market failure. This potential "failure" of market contracting can be overcome by the internalization of the affected cross-border activity. The MNE form of organization is also seen as minimizing the transaction costs resulting from the public goods aspects of certain intermediate, mostly intangible, assets by adopting managerial control of these assets rather than attempting to contract with others for their use.

The learning and technology transfer approach to the MNE has been less explored within the internalization school of research than approaches based on market failures and inefficiencies. The technology transfer argument emphasizes the value of a unifying organizational culture and the ease of coordination inside the firm relative to the market. Integration of activities within a firm opens pathways to learning and to sharing know-how and expertise through cross-border transfers within the MNE that might be blocked between legally separate entities concerned about leaking valuable information. The introduction of learning as a consideration moved internalization closer to a capabilities-based approach, in which learning is an important source of revitalization (Teece 1981).

The other leading non-capability approach to the MNE is the eclectic paradigm, which was developed by John Dunning (1995) and mentioned earlier in this article. The eclectic paradigm combines internalization with "ownership advantages" and host-country location factors to provide a more complete picture of MNE choice sets. The ownership factors are those that give a specific firm the competitive advantage that allows it to incur the expense of investing abroad and still earn a profit. Such advantages would certainly include organizational capabilities, and Dunning eventually incorporated the theory of dynamic capabilities into the eclectic paradigm (Dunning and Lundan 2010). However, the eclectic paradigm is generally used to explain only the geographic distribution of FDI and the cross-border investment decisions of firms, which are the concerns of the field of IB. It does not yield an understanding of the creation of ownership advantages or of their astute management to create firm-level advantage, which is seen as the province of international management studies.

Internalization and Dunning's ownership and location extensions in the eclectic paradigm have a number of shortcomings with respect to understanding the range of activities of the MNE and the advantages of specific firms. The most important of these is the limited attention to the role of entrepreneurs and entrepreneurial managers (Pitelis and Teece 2010; Teece 2014; Jones and Pitelis 2015). The activities of entrepreneurial managers, which, as will be discussed, make up an important part of the dynamic capabilities of the MNE, include the identification of opportunities, innovation with respect to the design of business models, and the creation and co-creation of markets. In short, internalization-based

approaches leave out most of what makes global firms viable over the long term: the creation and maintenance of a unique competitive advantage. Virtually the only managerial activity encompassed by internalization-based approaches to the MNE is the coordination of technology and other resources across borders.

Rather than squeezing capabilities concepts into the eclectic paradigm, as Dunning and Lundan (2010) have done, it can be argued that internalization is more properly seen as a subset of the development and deployment of capabilities (Teece 2014). As capabilities have become more globally dispersed and cross-border coordination less onerous, global supply chains seem less dependent than ever on internalization (i.e., on owning offshore factories).

Consider Apple, which became one of the largest computer and communications hardware firms in the world. It no longer owns its own manufacturing plants. Instead, it tightly coordinates supply relations with many companies, especially Foxconn, which is headquartered in Taiwan with its largest factories in China. Apple provides financing to some of its suppliers and may obtain exclusive purchase arrangements from them for periods up to several years. These non-internalized yet not-quite-arm's-length contractual arrangements appear to suffice for Apple to achieve the necessary coordination to leverage its considerable marketing and design capabilities while retaining flexibility to respond to demand changes.

Apple is just the most prominent example of an electronics industry in which large firms such as Hewlett-Packard have steadily sold off their offshore factories in favor of alliances with contractors, leaving the brand name firms responsible primarily for marketing and network coordination. It was the forerunner of an outsourcing trend that has spread in varying degrees to other sectors such as autos and from manufacturing to service activities.

Internalization perspectives do not provide a complete understanding of how contemporary outsourcing helps create firm-level advantages. Internalization needs to be combined with, and perhaps embedded into, a capability-based paradigm of the firm.

The Capabilities Approach

The capabilities approach to the MNE was built, in part, on the resource-based view of the firm, as well as extending the often-ignored learning and technology transfer branch of the internalization approach. Its leading expression is the dynamic capabilities framework, which was developed in the field of strategic management (Teece et al. 1997; Teece 2014).

An organizational capability allows firms to marshal resources to produce a desirable outcome with some degree of predictability. Most capabilities are somewhat fungible and can be used to support any of a variety of activities. Large organizations have many such capabilities and at any point in time some of them will be underutilized. Capabilities arise in part from learning, from combinations of organizational assets, and from acquisitions.

The ordinary capabilities that the firm needs to carry out a given programme of production can often be replicated – and imitated – with relative ease. Much know-how, which used to be proprietary by default as much as by intention, is now effectively in the public domain – available from consultants, schools of engineering, and the public literature. That's not to say that getting to the level of best practice is easy. It is not, but there is usually a fairly clear path for getting there if the cost of doing so is deemed worthwhile.

Dynamic capabilities, on the other hand, allow the organization and its management to integrate, build, and reconfigure internal and external competencies (including its ordinary capabilities) to address changing business environments (Teece et al. 1997). At a practical level, this involves sensing and evaluating threats and opportunities, designing structures and business models to respond to them, and adjusting and renewing the organization and its resources as needed. Strong dynamic capabilities will enable an MNE to constantly create new technologies, differentiated processes, and better business models to stay ahead of the competition, stay in tune with the market, and even, at times, shape the market.

A capacity to transfer technology and know-how (embedded in routines and resources,

including the minds of employees) across distances and borders is a key capability of the multinational firm (Teece 1976). Successful transfers of even ordinary capabilities to host-country subsidiaries can provide the basis for advantage in the host country. MNEs investing abroad "appear to adopt good management practices in almost every country in which they operate," and foreign subsidiaries are generally better managed than similar host-country firms (Bloom et al. 2012: 14). In other words, strong ordinary capabilities developed at home can (at least temporarily) be distinctive abroad. However, as competition increases in the host country, the strength or weakness of the MNE subsidiary's dynamic capabilities to respond to the changes becomes paramount.

As this suggests, subsidiaries must have their own capacity to anticipate and respond to changes in the local business environment and be given encouragement and autonomy to do so. KFC Japan, for example, became a success in the 1970s under the leadership of Loy Weston and Shin Ohkawara, who developed the local branch more as a fashion business than as fast food (Bartlett and Rangan 1986). Headquarters management must vet key resource commitments, but, in order to preserve agility, information must flow and decisions must be taken rapidly.

Headquarters can enhance local capabilities further by allowing and facilitating knowledge transfers amongst regional divisions and by encouraging and supporting the exploitation of cross-border complementarities. This approach requires top management to treat the organization more like an interconnected network than a rigidly vertical M-form hierarchy (Bartlett and Ghoshal 1989). Top management's fundamental roles are global ▶ asset orchestration (allocating financial and other resources to develop the most promising activities wherever they may appear) and the provision of company-wide strategic direction.

In this decentralized M-form model, subsidiaries generate know-how, capabilities, and products from their own history, circumstances, and innovation activities that can potentially be transferred to other business units at home or abroad (Birkinshaw and Hood 1998). Local knowledge creation and discovery of opportunities is encouraged and coordinated by the orchestration activities of headquarters management.

Comparison of Capabilities with Other Theories of the MNE

Countless studies have applied an internalization approach to the MNE to research on the "entry mode" of firms in overseas markets – that is, whether the MNE chooses to operate in a host country through an alliance or via some level of equity investment. However, from a strategic management perspective, the more urgent questions involve which markets to enter and when, not just how. Answering these requires a recognition and assessment of opportunities in multiple markets, the ability to transfer knowledge, an analysis of any capabilities gaps the firm may have in terms of the requirements to carry out its strategy, and so on. Any capability gaps, in turn, must be evaluated as make/buy/ally problems based on the strategic value of the capability, including its availability from other firms and the time required to "make" it internally versus the timing of the opportunity to be exploited.

A particular deficiency of internalization theories is their focus on the cost of entry while ignoring the opportunities (and risks) that entry affords. Consider, for example, a case where assets owned by the MNE are strongly complementary to (i.e., co-specialized with) assets in the host country. Such cases are quite common in a world where value is determined largely by ownership and control of ▶ intangible assets. The question facing the MNE is not so much whether a mode of operation in the host country will cost less when owned or contracted but whether integration of the cross-border activity within the firm will allow it to better capture value by exercising more control over the design of a value appropriation architecture (Pitelis and Teece 2010).

In other words, internalization analysis can reveal only part of the story behind the choice of the firm's boundary when entering overseas markets. A capabilities perspective allows consideration of other critical factors, including the feasibility of transferring the necessary

knowledge either internally or externally, the host country's treatment of intellectual property, and the attractiveness of potential host-country partners.

Internalization approaches that are based primarily on transaction cost or contractual analysis offer even less insight into the building of firm-level advantage. In such theories, if a firm has market power, it is taken as given and assumed to persist. The role of managers is reduced to determining the (global) boundaries of the firm by outsourcing until the cost of outsourcing the marginal activity is equal to the cost of performing it internally. There is little recognition of the importance of opportunity discovery, learning, strategy adjustment, and other forms of capability development and maintenance.

The dynamic capabilities framework, by contrast, models managers as exercising entrepreneurial and leadership functions that are vital to building and maintaining firm-level advantages in home and host-country markets. Global asset orchestration, business model design, and entrepreneurial cross-border market creation and co-creation are at the core of a capabilities-based approach to the MNE.

See Also

- Asset Orchestration
- Co-specialization
- Eclectic Paradigm
- Evolutionary Theory of the Multinational Corporation
- Federative Multinational Enterprise (MNE)
- Foreign Direct Investment (FDI)
- Hymer, Stephen Herbert (1934–1974): The MNE and International Business
- Intangible Assets
- Internalization Theory
- International Business
- Make-or-Buy Decisions: Applications to Strategy Research
- Market Entry Strategies
- Market Failures and MNEs
- M-Form Firms
- Multinational Corporations
- Multi-plant Economies
- Strategy and Structure of the Multinational Enterprise (MNE)
- Sustainable Competitive Advantage

References

Aliber, R.Z. 1970. A theory of direct foreign investment. In *The international corporation*, ed. C.P. Kindleberger. Cambridge, MA: The MIT Press.

Bartlett, C.A., and S. Ghoshal. 1989. *Managing across borders: The transnational solution*. Boston: Harvard Business School Press.

Bartlett, C.A., and U.S. Rangan. 1986. *Kentucky Fried Chicken (Japan) Limited (case no. 9-387-043)*. Boston: Harvard Business School.

Birkinshaw, J., and N. Hood. 1998. Multinational subsidiary evolution: Capability and charter change in foreign-owned subsidiary companies. *Academy of Management Review* 23: 773–795.

Bloom, N., C. Genakos, R. Sadun, and J. Van Reenen. 2012. Management practices across firms and countries. *Academy of Management Perspectives* 20: 12–33.

Buckley, P.J., and M.C. Casson. 1976. *The future of multinational enterprise*. London: Macmillan.

Caves, R.E. 1980. Investment and location policies of multinational companies. *Swiss Journal of Economics and Statistics* 116: 321–338.

Dunning, J.H. 1988. The eclectic paradigm of international production: A restatement and some possible extensions. *Journal of International Business Studies* 19: 1–31.

Dunning, J.H. 1995. Reappraising the eclectic paradigm in the age of alliance capitalism. *Journal of International Business Studies* 26: 461–493.

Dunning, J.H., and S.M. Lundan. 2008. *Multinational enterprises and the global economy*. Cheltenham: Edward Elgar.

Dunning, J.H., and S.M. Lundan. 2010. The institutional origins of dynamic capabilities in multinational enterprise. *Industrial and Corporate Change* 19: 1225–1246.

Hennart, J.-F. 1982. *A theory of multinational enterprise*. Ann Arbor: University of Michigan Press.

Hymer, S. 1976. *The international operations of national firms*. Boston: The MIT Press.

Jones, G., and C.N. Pitelis. 2015. Entrepreneurial imagination and a demand and supply-side perspective on the MNE and cross-border organisation. *Journal of International Management* 21: 309–321.

Kogut, B., and U. Zander. 1993. Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies* 24: 625–645.

Oviatt, B.M., and P.P. McDougall. 1994. Toward a theory of international new ventures. *Journal of International Business Studies* 25: 45–64.

Pitelis, C.N., and D.J. Teece. 2010. Cross-border market co-creation, dynamic capabilities and the

entrepreneurial theory of the multinational enterprise. *Industrial and Corporate Change* 19: 1247–1270.
Rugman, A.M. 1981. *Inside the multinationals: The economics of internal markets*. New York: Columbia University Press.
Teece, D.J. 1976. *The multinational corporation and the resource cost of international technology transfer*. Cambridge, MA: Ballinger Publishing.
Teece, D.J. 1981. The multinational enterprise: Market failure and market power considerations. *Sloan Management Review* 22: 3–17.
Teece, D.J. 1986. Transaction cost economics and the multinational enterprise: An assessment. *Journal of Economic Behavior & Organization* 7: 21–45.
Teece, D.J. 2014. A dynamic capabilities-based entrepreneurial theory of the multinational enterprise. *Journal of International Business Studies* 45: 8–37.
Teece, D.J., G. Pisano, and A. Shuen. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* 18: 509–533.

Capability Development

David J. Teece
Berkeley Research Group, LLC, Emeryville, CA, USA
Haas School of Business, University of California, Berkeley, Berkeley, CA, USA

Abstract

Capability development refers to creating a new capability or enhancing an existing one. Capability evolution, capability growth, capability expansion and capability maturation are often used interchangeably for this idea. A number of theoretical lenses – such as the knowledge-based approach, resource based view and evolutionary theory – offer insights into how firms develop organizational capabilities. Different authors highlight different factors as antecedents of capability that influence the development of capabilities, such as knowledge within the firm, experience, organizational learning, routines and non-routine actions (e.g., semi-continuous asset orchestration or redesigning routines).

Definition The term *capability development* refers to creating a new capability or enhancing an existing one. It is often also described interchangeably as capability evolution, capability growth, capability expansion or capability maturation.

The term *capability development* refers to creating a new capability or enhancing an existing one. It is often also described interchangeably as capability evolution, capability growth, capability expansion or capability maturation. Studies in strategic management often focus on the organizational capabilities of for-profit firms, distinct from a collection of individual skills.

Managing knowledge and resources at the organizational level involves the accumulation of tangible and ▶ intangible assets and the acquisition of skills, which provides opportunities for learning. As Pisano (2000: 129) concludes, 'Without learning, it is difficult to imagine from where a firm's unique skills and competences would come.' In this learning process, cognition that helps organizations identify market opportunities is central to dynamic capability development (Gavetti 2005), especially in capability development's early stages (Helfat and Peteraf 2003). Furthermore, capabilities can be acquired through learning from the experience of others via both alliances (Inkpen and Dinur 1998) and acquisitions (Karim and Mitchell 2000).

Deliberate learning that involves time and cognitive efforts also facilitates the process of capability development. Drawing on behaviour and cognitive traditions in ▶ organizational learning, Zollo and Winter (2002) identify different capability-building mechanisms by the level of cognition effort and learning investment – semi-automatic learning (e.g., learning by doing) and deliberate learning (e.g., knowledge articulation and codification) – and argue that ▶ dynamic capabilities emerge from the co-evolution of past experience, knowledge articulation and knowledge codification processes. In addition to building on existing resources (exploitative learning), organizations must integrate new resources and change ways of doing things, which require explorative learning.

Levinthal and March (1993) note the negative effects of exploitive learning processes – in particular, developing competence in particular areas

prevents organizations from exploring other competencies. This is echoed in Kogut and Kulatilaka (2001: 755): 'The pitfall is that learning increases the rigidity of the firm.' Leadership is required in order to create new capabilities regarding new opportunities, especially in a dynamic and uncertain environment (Sirmon et al. 2007). Many scholars agree that organizations must be 'ambidextrous' (O'Reilly and Tushman 2004), and improve their abilities to exploit existing competencies while simultaneously exploring new competencies. Some scholars (Teece 2007) would add that they not only need to sense and seize but to transform as well. Not all organizational units need to be doing all three all the time, but they do need to have latent capacities for all three.

Organizational learning is intimately tied to evolutionary theory, in which knowledge generated by collective learning in organizations resides in organizational routines (Nelson and Winter 1982). According to Nelson and Winter, an organizational capability involves collective action and a series of patterned routines resident in a set of actors within the organization. Therefore, it is said that 'routines are the building blocks of capabilities' (Dosi et al. 2000: 4). The accumulation of experience via learning mechanisms such as search routines leads to a capability (Winter 2000), and the organization's dynamic routines of continual improvement and experimentation can foster differentiated technological capabilities (Dosi 1988).

While recognizing the importance of routines, the dynamic capabilities framework also highlights the roles of non-routine actions – such as semi-continuous asset orchestration and creating/redesigning routines – in building dynamic capabilities whereby the firm is able to renew and adapt its current capabilities to rapidly changing environments (Teece et al. 1997; Teece 2012). Teece (2007) provides a portfolio of microfoundations for dynamic capabilities: change routines (e.g., product development) and analytical methodologies (e.g., investment choices). Eisenhardt and Martin (2000) identify specific organizational processes and routines as antecedents of dynamic capabilities: quality control routines, technology transfer routines and certain performance measurement systems. With hindsight, some of these might more appropriately be thought of as ordinary (rather than dynamic) capabilities.

A number of other theoretical lenses offer further insights into how firms develop organizational capabilities. These include the knowledge-based approach (e.g., Nonaka 1991; Kogut and Zander 1992; Grant 1996), the ▶ resource based view (Wernerfelt 1984; Barney 1991), organizational learning (Levitt and March 1988) and evolutionary theory (Nelson and Winter 1982). Also, knowledge-based perspective views implicit knowledge as the key source of competitive advantage. Individually (or collectively) held knowledge within the firm is seen as a basis for creating organizational capabilities (Grant 1996). Similarly, Kogut and Zander emphasize that different internal social relations (differences in the knowledge bases of individuals and groups) shape the 'combinative capabilities' of a firm – recombining and transforming old capabilities into new ones. According to this view, organizational capability develops to the extent to which knowledge among different parts of a firm is shared, recombined and integrated (e.g., Kogut and Zander 1992; Grant 1996).

The resource-based view acknowledges the importance of coordination but also embraces managerial strategies for developing capabilities (Wernerfelt 1984). Simply possessing valuable resources does not guarantee the development of capabilities and competitive advantage (Barney 1991); in addition, effective resource management – such as structuring resource portfolios and bundling resources – is considered just as important for building capabilities (Sirmon et al. 2007).

See Also

- ▶ Dynamic Capabilities
- ▶ Exploration and Exploitation
- ▶ Intangible Assets
- ▶ Learning and Adaptation
- ▶ Organizational Learning
- ▶ Resource-Based View

References

Barney, J.B. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17: 99–120.

Dosi, G. 1988. Sources, procedures, and microeconomic effects of innovation. *Journal of Economic Literature* 26: 1120–1171.

Dosi, G., R.R. Nelson, and S.G. Winter. 2000. *The nature and dynamics of organizational capabilities*. Oxford: Oxford University Press.

Eisenhardt, K.M., and J. Martin. 2000. Dynamic capabilities: What are they? *Strategic Management Journal* 21: 1105–1121.

Gavetti, G. 2005. Cognition and hierarchy: Rethinking the microfoundations of capabilities' development. *Organization Science* 16: 599–617.

Grant, R.M. 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal* 17: 109–122.

Helfat, C.E., and M.A. Peteraf. 2003. The dynamic resource-based view: Capability lifecycles dynamic capabilities deconstructed. *Strategic Management Journal* 24: 997–1010.

Inkpen, A.C., and A. Dinur. 1998. Knowledge management processes and international joint ventures. *Organization Science* 9: 454–468.

Karim, S., and W. Mitchell. 2000. Path-dependent and path-breaking change: Reconfiguring business resources following acquisitions in the U.S. medical sector, 1978–1995. *Strategic Management Journal* 21: 1061–1081.

Kogut, B., and N. Kulatilaka. 2001. Capabilities as real options. *Organization Science* 12: 744–758.

Kogut, B., and U. Zander. 1992. Knowledge of the firm, combinative capabilities and the replication of technology. *Organization Science* 3: 383–397.

Levinthal, D.A., and J.G. March. 1993. The myopia of learning. *Strategic Management Journal* 14: 95–112.

Levitt, B., and J.G. March. 1988. Organizational learning. *Annual Review of Sociology* 14: 319–340.

Nelson, R., and S. Winter. 1982. *An evolutionary theory of economic change*. Cambridge, MA: Belknap Press.

Nonaka, I. 1991. The knowledge-creating company. *Harvard Business Review* 69: 96–104.

O'Reilly, C.A., and M.L. Tushman. 2004. The ambidextrous organization. *Harvard Business Review* 82: 74–83.

Pisano, G.P. 2000. In search of dynamic capabilities: The origins of R&D competence in biopharmaceuticals. In *The nature and dynamics of organizational capabilities*, ed. G. Dosi, R. Nelson, and S. Winter. Oxford: Oxford University Press.

Sirmon, D.G., M.A. Hitt, and R.D. Ireland. 2007. Managing firm resources in dynamic environments to create value: Looking inside the black box. *Academy of Management Review* 32: 273–292.

Teece, D.J. 2007. Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal* 28: 1319–1350.

Teece, D.J. 2012. Dynamic capabilities: Routines versus entrepreneurial action. *Journal of Management Studies* 49: 1395–1401.

Teece, D.J., G. Pisano, and A. Shuen. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* 18: 509–533.

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

Winter, S.G. 2000. The satisficing principle in capability learning. *Strategic Management Journal* 21: 981–996.

Zollo, M., and S.G. Winter. 2002. Deliberate learning and the evolution of dynamic capabilities. *Organization Science* 13: 339–351.

Capability Lifecycle

Tiona Zuzul[1] and Constance E. Helfat[2]
[1]Harvard University, Business School, Cambridge, MA, USA
[2]Dartmouth College, Dartmouth, NH, USA

Abstract

The *capability lifecycle* provides a framework that articulates patterns in the evolution of a capability from birth through maturity, and subsequent branching into additional stages of development (Helfat and Peteraf, *Strategic Management Journal* 24: 997–1010, 2003). A capability's lifecycle begins with its founding, when a group of individuals organizes around an objective. Subsequently, in the development stage, a capability is built, refined and improved. Finally, a capability enters the maturity stage, where it ceases development and is maintained through exercise. Before or after reaching maturity, a capability can also transform through the '6 R's' of capability branching: retirement, retrenchment, renewal, replication, redeployment and recombination.

Definition The capability lifecycle is a framework that extends the resource-based view by articulating patterns and paths in the evolution of organizational capabilities. The lifecycle describes the development of a capability from founding to maturity, as well as subsequent branching into additional stages, and explains

the sources of heterogeneity in organizational capabilities at each stage.

The *capability lifecycle* provides a framework that articulates patterns in the evolution of a capability from birth through maturity, and subsequent branching into additional stages of development (Helfat and Peteraf 2003).

Much like products and technologies, the capabilities of an organization – which draw on resources to perform a coordinated set of tasks towards a particular end result – evolve over time. The capability lifecycle describes the stages of the evolution of organizational capabilities, and provides an overview of the main features of each stage (Helfat and Peteraf 2003). The intellectual roots of the capability lifecycle lie in evolutionary economics (Nelson and Winter 1982) and the ▶ Resource-Based View (RBV), or the idea that organizations differ in their resources and capabilities in important, durable ways that affect competitive performance (Penrose 1959; Rumelt 1984; Wernerfelt 1984; Peteraf 1993). The lifecycle view of capabilities adds dynamism to the RBV by noting that capabilities come about and shift over time, and that this evolution leads to heterogeneity in performance over time, both within and across firms. The lifecycle perspective thus provides an explanation for the heterogeneity of capabilities, and a framework for thinking about organizational development in a patterned, systematic way. In addition, the capability lifecycle includes an important potential role for ▶ dynamic capabilities in bringing about capability change.

Stages in the Capability Lifecycle

Founding

The lifecycle of a capability begins with its founding, when a group of individuals organizes around an objective. As an example, consider a company founded by a team of individuals in order to provide technological solutions that improve the environmental efficiency of buildings and built environments, as documented by Zuzul and Edmondson (2011). The accomplishment of such an objective depends on a team's capabilities: its ability to obtain and use resources – the tangible or intangible assets or inputs it owns or has access to – in coordinated, controlled ways. The team must therefore organize to build capabilities that enable it to transform its inputs in pursuit of its objective.

Each team has a set of initial endowments (Levinthal and Myatt 1994; Helfat and Lieberman 2002) – members' human, social and cognitive capital (Adner and Helfat 2003), and a set of team characteristics such as leadership and history of interaction – that shape the capabilities that it can build. In the technological solutions company mentioned above, the founding team members came from a variety of backgrounds, and some, including the team's leader, had experience in the software industry. These initial endowments led team members to organize to build an ▶ alliance capability (Anand and Khanna 2000) – including the ability to find and recruit partners, a significant element in the business models of many software companies – because they determined that an alliance capability would be important in achieving their objective. More generally, team members' endowments are an initial source of heterogeneity in ▶ capability development; teams with different individual capital and patterns of interrelationships are likely to organize to build different capabilities, even towards the accomplishment of identical aims.

Development

After a team has defined a central objective and organized to build the capabilities it needs, it begins to develop each capability. In the development stage, a capability is built, refined and improved. Teams can develop a capability either from scratch or by imitating capabilities that exist in another organization: members might look for guidance in the capability development of their competitors, for instance, or in the organizations to which they had previously belonged.

Once founded, a capability tends to improve over time through both cumulative experience (learning-by-doing) and deliberate ▶ organizational learning (Winter 2000; Edmondson et al. 2001; Nelson and Winter 2002). In the

technology solutions company, the team developed its alliance capability through the accumulation of experience, as the new company learned from its experience with a growing number of partners. The team also learned more deliberately through attempts at process improvement; the company, for instance, adopted a standardized model of partner acquisition it termed the 'partner pipeline'. Furthermore, it learned through deliberate experimentation, by testing alternatives and reflecting on outcomes through cycles of feedback; employees, for instance, shared their experiences with partner acquisition in weekly operations meetings. As in this example, each learning process leads to improvements in a capability, which is gradually honed until it ceases developing and reaches maturity.

The development stage presents multiple possibilities for heterogeneity: teams can differ in the success and speed with which they develop capabilities, and the point at which they cease to improve capabilities (Helfat and Peteraf 2003). As a result, as Helfat and Peteraf (2003) write, 'some versions of a capability are better than others' (p. 999).

Maturity

Finally, a capability enters the maturity stage, where it ceases development and is maintained through exercise. This may happen because the capability has an efficiency limit – an inherent bound given available technologies and inputs – or because members satisfice by determining that a particular level of skilfulness is sufficient to accomplish their objectives (Winter 2000).

As we will describe, not all capabilities reach maturity: capabilities that no longer match internal and external environmental conditions can be selected out of the organization (Helfat and Peteraf 2003). The capabilities that remain can become deeply entrenched in organizational memory as tacit routines. Organizational forgetting can lead to declines in a capability's efficiency; how well the capability is maintained therefore depends on how consistently it is exercised. Exercised sufficiently, the ability to identify, attract and retain potential partners might, for example, become an implicit routine in our technology company.

Capability Branching

A capability's progress from founding to maturity does not always follow the same developmental path. As a capability evolves from founding to maturity, its path can be shaped through changes in the demands of its internal and external environment. Internal factors like new managerial decisions, and external factors like changes in supply and demand, present both threats and opportunities for a capability's development. Helfat and Peteraf (2003) posit that a capability can respond to these threats and opportunities in patterned ways, before or after reaching maturity. More specifically, a capability can transform through the '6 R's' of capability branching: retirement, retrenchment, renewal, replication, redeployment and recombination. As a capability evolves, it may pass through several branching stages, and finish far from its origins.

Threats in an organization's internal and external environment can render a capability obsolete; a capability once necessary for the attainment of organizational aims may no longer be relevant. New antitrust regulations might, for example, make the pursuit of further alliances impossible for our technology company. In response to this threat, the company may retire the capability – cease developing and using it – or it may retrench the capability – gradually limit its development and use until it declines entirely.

An organization can also respond to threats and opportunities by redesigning its capability set and honing and adapting its existing capabilities. An organization might renew a capability: a crisis or an opportunity may deepen management's perceived need for a capability, raising aspirations (Winter 2000) and leading them to improve or modify the capability. The emergence of a salient competitor might lead our technology company, for example, to further invest in its alliance capability to raise its efficiency and counter this threat.

An organization can also transfer an existing capability into new markets. It can replicate a capability by reproducing it in another geographic market (Winter and Szulanski 2001), or redeploy it by transferring it to a related product market (Helfat and Raubitschek 2000). The technology

solutions company, for example, could replicate its alliance capability in a new country, or redeploy this capability to develop a new product. Finally, an organization can combine a capability with other capabilities in order to transfer it into a new market: our technology company could combine its existing alliance capability with a newly acquired lobbying capability to begin encouraging foreign governments to tighten their environmental regulations, thereby increasing demand for its products abroad. In each case, transferring or altering an existing capability into new markets is often a more efficient response than the creation of a new capability from scratch.

Capability branching is a final source of potential organizational heterogeneity. The same internal or external conditions will not affect capabilities in all organizations in the same way. Any threat or opportunity can be met with a number of different reactions: not all organizations will develop their alliance capability in the same way as our technology solutions company, although many may meet similar threats and opportunities (Adner and Helfat 2003).

Building on the Capability Lifecycle

The aim of the capability lifecycle is to provide a framework for thinking about capability development, and the impact that differential development can have on organizational heterogeneity and performance over time (Helfat and Peteraf 2003). Theoretical and quantitative work has built on this framework to explore particular stages in a capability's lifecycle. Theories of strategic performance, for example, have emphasized that managers' differential ability to found, develop and transfer capabilities makes some entrepreneurial firms better able to discover and exploit opportunities (Zahra et al. 2006) and create value (Sirmon et al. 2007). With respect to the alliance capability of the technology solutions company discussed above, research on alliances suggests that the capability to ally successfully can be developed over time, and needs to be exercised and maintained (Anand and Khanna 2000; Rothaermel and Deeds 2006; Kale and Singh 2007).

The lifecycle framework could further benefit from in-depth qualitative studies that explore when and how particular organizations found, develop and branch their capabilities – and how this affects firm performance. A recent study of capability development and decline at Smith Corona, formerly one of the world's largest typewriter companies, illustrates this sort of qualitative approach (Danneels 2011). In an overview of 21 years of company history, Danneels (2011) demonstrates how capabilities can become irrelevant through internal and external changes, and how Smith Corona's inability to develop new capabilities or transform existing ones resulted in organizational failure. As the capability lifecycle perspective suggests, organizational performance depends on responses to threats and opportunities – and on the ability to develop appropriate capabilities. Further qualitative work can continue to refine and illustrate this theoretical framework. Qualitative and quantitative longitudinal studies could benefit from the lifecycle framework, which provides a lens with which to systematically parse and analyse data on firm heterogeneity, performance and response to change.

See Also

- ▶ Alliance Capability
- ▶ Capability Development
- ▶ Dynamic Capabilities
- ▶ Firm Resources
- ▶ Organizational Learning
- ▶ Resource-Based Theories
- ▶ Resource-Based View

References

Adner, R., and C. Helfat. 2003. Corporate effects and dynamic managerial capabilities. *Strategic Management Journal* 24: 1011–1025.

Anand, B., and T. Khanna. 2000. Do firms learn to create value? The case of alliances. *Strategic Management Journal* 21: 295–316.

Danneels, E. 2011. Trying to become a different type of company: Dynamic capability at Smith Corona. *Strategic Management Journal* 32: 1–31.

Edmondson, A.C., R.M. Bohmer, and G.P. Pisano. 2001. Disrupted routines: Team learning and new technology adaptation. *Administrative Science Quarterly* 46: 685–716.
Helfat, C.E., and M.B. Lieberman. 2002. The birth of capabilities: Market entry and the importance of pre-history. *Industrial and Corporate Change* 11: 725–760.
Helfat, C.E., and M.A. Peteraf. 2003. The dynamic resource-based view: Capability lifecycles. *Strategic Management Journal* 24: 997–1010.
Helfat, C.E., and R.S. Raubitschek. 2000. Product sequencing: Co-evolution of knowledge, capabilities and products. *Strategic Management Journal* 21: 961–980.
Kale, P., and H. Singh. 2007. Building firm capabilities through learning: The role of the alliance learning process in alliance capability and firm-level alliance success. *Strategic Management Journal* 28: 981–1000.
Levinthal, D., and J. Myatt. 1994. Co-evolution of capabilities and industry: The evolution of mutual fund processing. *Strategic Management Journal* 15: 45–62.
Nelson, R.R., and S.G. Winter. 1982. *An evolutionary theory of economic change*. Cambridge, MA: Belknap Press.
Nelson, R.R., and S.G. Winter. 2002. Evolutionary theorizing in economics. *Journal of Economic Perspectives* 16: 23–46.
Penrose, E. 1959. *The theory of the growth of the firm*. New York: Wiley.
Peteraf, M.A. 1993. The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal* 14: 179–192.
Rothaermel, F., and D. Deeds. 2006. Alliance type, alliance experience and alliance management capability in high-technology ventures. *Journal of Business Venturing* 21: 429–460.
Rumelt, R.P. 1984. Towards a strategic theory of the firm. In *Competitive strategic management*, ed. R.B. Lamb. Englewood Cliffs: Prentice Hall.
Sirmon, D.G., M.A. Hitt, R.D. Ireland, and A. Texas. 2007. Managing firm resources in dynamic environments to create value: Looking inside the black box. *Academy of Management Review* 32: 273–292.
Wernerfelt, B. 1984. A resource-based view of the firm. *Strategic Management Journal* 5: 171–180.
Winter, S.G. 2000. The satisficing principle in capability learning. *Strategic Management Journal* 21: 981–996.
Winter, S.G., and G. Szulanski. 2001. Replication as strategy. *Organization Science* 12: 730–743.
Zahra, S.A., H.J. Sapienza, and P. Davidsson. 2006. Entrepreneurship and dynamic capabilities: A review, model and research agenda. *Journal of Management Studies* 43: 917–955.
Zuzul, T.Z., and A.C. Edmondson. 2011. Ambiguity squared: Growing a new business in a nascent industry. Working paper, Harvard Business School.

Capital Asset Pricing Model (CAPM)

Rajeev Bhattacharya
Berkeley Research Group, Boston, MA, USA

Abstract

The capital asset pricing model (CAPM) for a security is a linear relationship between the expected excess return of the security and the expected excess return of the market. It was developed by William Sharpe, John Lintner and Jan Mossin. It is a useful framework to discuss idiosyncratic and systematic risk. The security market line is a powerful graphical construct of the CAPM. While the CAPM has strong underlying assumptions, recent research has relaxed many of these assumptions. It is commonly used to calculate cost of capital and required rate of return.

Definition The capital asset pricing model (CAPM) for a security is a linear statistical relationship between the expected excess return of the security and the expected excess return of the market, where expected excess return of a security (market) is defined as the expected return of the security (market) minus the return of a risk-free asset.

Let

- $R_i =$ return of security i
- $R_M =$ return of the market portfolio
- $R_f =$ return of the risk-free asset.

The capital asset pricing model (CAPM) for security i is:

$$E(R_i) - R_f = \beta_i\big(E(R_M) - R_f\big);$$

that is,

$$\Gamma_i = \beta_i \Gamma_M,$$

where $\Gamma_M = E(R_M) - R_f$ is the market risk premium, and $\Gamma_i = E(R_i) - R_f$ is the risk premium for security i.

The CAPM can also be written as:

$$E(R_i) = R_f + \beta_i\big(E(R_M) - R_f\big);$$

that is,

$$E(R_i) = R_f + \beta_i \Gamma_M.$$

History

Bodie et al. (2008: 293) explained that '[t]he capital asset pricing model is a set of predictions concerning equilibrium expected return on risky assets. Harry Markowitz laid down the foundations of modern portfolio management in 1952. The CAPM was developed 12 years later in articles by William Sharpe, John Lintner and Jan Mossin.' See also Brealey and Myers (2003), Damodaran (2002), Markowitz (1999), Miller (1999), and Sharpe (1964).

Idiosyncratic Risk

Idiosyncratic risk of security i is defined as:

$$\begin{aligned} IR_i &= SD\big(\big(R_i - R_f\big) - \beta_i\big(R_M - R_f\big)\big) \\ &= SD\big(R_i - \beta_i R_M - (1 - \beta_i)R_f\big). \end{aligned}$$

Consider a stock i that is underpriced (overpriced) according to the information available to an arbitrageur. In order to exploit this profitable opportunity, the arbitrageur will construct the following arbitrage portfolio, if the arbitrageur were constrained to using only the market index and the risk-free asset – see Bhattacharya and O'Brien (2015) for a discussion of the possibilities when a wider set of securities is available to the arbitrageur:

- Arbitrage numerator: go long (short) on the mispriced stock – let's say by \$1, which is purely a normalization.
- Other arbitrage legs:
  - Go short (long) α_M on the market index; and
  - Go short (long) α_f on the risk-free asset.

The total amount on these legs has to add up to \$1 short (long); and how much to go short (long) on each leg is called the corresponding *hedge ratio*. The ratios α_M and α_f are the relevant hedge ratios, and the *zero-net-investment condition* will require $1 - \alpha_M - \alpha_f = 0$. The return on period i of this hedge portfolio is $R_{i,\,t} - \alpha_M R_{M,\,t} - \alpha_f R_{f,\,t} = \eta_{i,\,t}$, which combined with the zero-net-investment condition $1 - \alpha_M - \alpha_f = 0$ yields $(R_{i,\,t} - R_{f,\,t}) = \alpha_M (R_{M,\,t} - R_{f,\,t}) + \eta_{i,\,t}$.

The ordinary least squares (OLS) estimates $\widehat{\alpha}_M$ and $\widehat{\alpha}_f = 1 - \widehat{\alpha}_M$ give us the hedge ratios that optimize the assumed objective of minimum standard deviation, among zero-net-investment portfolios consisting of the market index and the risk-free asset (this simplification is due to Wurgler and Zhuravskaya (2002)). $SD\big(\widehat{\eta}_{i,\,t}\big)$ is the idiosyncratic risk of the stock i (i.e., idiosyncratic risk is the ex post holding cost for an arbitrageur who is constrained to using only the market index and the risk-free asset as legs of the arbitrage portfolio). See Pontiff (2006) for a detailed discussion on idiosyncratic risk as a holding cost of arbitrage.

Systematic Risk

Systematic risk (or non-diversifiable risk or market risk) of security i is defined as:

$$SR_i = \beta_i = \frac{\mathrm{Cov}(R_i, R_M)}{\mathrm{Var}(R_M)}.$$

Although this market beta is widely used as a measure of the 'riskiness' of a security, it is actually a measure of how the security's return varies with the market, not of the riskiness of the security per se. However, for an investor holding the market portfolio, β_i measures the marginal effect on risk of an increment in the holding of security i, everything else remaining the same.

Special Cases

Risk-Free Instrument

When the security i is a risk-free instrument (e.g., proxied by a t-bill), $\beta_i = 0$.

Market Portfolio

When the security i is the market portfolio (e.g., proxied by a market index), $\beta_i = 1$.

It is often argued that a beta above one signifies an asset of above-average riskiness, whereas a beta below one signifies an asset of below-average riskiness. Also, a riskier security will have a higher beta and will be discounted at a higher rate. CAPM is consistent with the risk-averse investor's demanding a higher expected return for a riskier asset.

Security Market Line

The horizontal axis represents beta, and the vertical axis represents expected return. When the CAPM is plotted along these coordinates, the resulting graph is called the security market line. The vertical intercept of the *security market line* is the nominal risk-free rate of return, and its slope is the market risk premium. The value of the security market line corresponding to $\beta = 1$ is the expected return of the market portfolio.

The security market line is a useful tool to determine whether an asset offers a reasonable expected (or 'fair') return. If the plot of a security is above the security market line, the security is undervalued; similarly, if the plot is below the security market line, the security is overvalued. The expected return of a security minus the expected return on the security market line corresponding to its beta is referred to as the alpha (α) of the securityã therefore, $\alpha_i > 0$ for an undervalued security i, and $\alpha_i < 0$ for an overvalued security i.

Under CAPM, in equilibrium, expected $\alpha_i = 0$ for each security i. However, we find that, on average, low-beta securities have positive alphas and high-beta securities have negative alphas.

It can be argued that security analysis is about identifying securities with non-zero alphas–an investor (or a fund manager) would increase the weights of securities with positive alphas and reduce the weights of securities with negative alphas. Such behavior would increase the price of securities with positive alphas and reduce the price of securities with negative alphas, which would exert pressure in the direction of equilibrium with zero alphas.

It is sometimes important to compare the CAPM against independent estimates of the returns of the security – such independent estimates include comparables analysis. As with any other technique, CAPM would be ex post correct if the estimated price equaled the discounted sum of cash flows accruing to the security.

Suppose there are N sources of extra-market risk (e.g., industry, inflation) for which there are N associated hedge portfolios with returns $R_1,\ldots R_N$. Then, the multi-index form of the CAPM is

$$E(R_i) - R_f = \beta_{iM}\big(E(R_M) - R_f\big) + \sum_{n=1}^{N} \beta_{in}\big(E(R_n) - R_f\big).$$

Fischer Black derived a more general version of the CAPM in 1972 – in this version, the expected return of an asset in excess of the zero-beta return, is linearly related to its beta. The zero-beta portfolio is the portfolio with the minimum variance of all portfolios uncorrelated with the market portfolio. For the Black version, returns are generally stated on an inflation-adjusted basis.

Assumptions and Limitations

The following assumptions are made under CAPM:

- Investors are rational and risk-averse mean-variance optimizers. Investors prefer higher-mean and lower-risk investments. Standard deviation or variance is assumed to be an adequate measure of risk. This is true if normality of returns holds, but may not reflect more general measures of risk and attitudes toward risk.
- Investors are myopic (i.e., they only plan for one holding period). This is a strong assumption, but it can be relaxed for inter-temporal decision making.
- Investors hold diversified investments. In particular, each investor holds a combination of

the market portfolio, where the proportion of each asset equals the market capitalization of the asset divided by the market capitalization of all assets, and the risk-free asset. The market portfolio will be on the efficient frontier. However, the amount invested in the market portfolio by an investor will depend on the investor's wealth and attitude toward risk. A market index is an incomplete proxy for the market portfolio.

- All assets are publicly traded and perfectly divisible. This assumption rules out investments in non-traded assets, such as human capital and private enterprises.
- There are many investors, all of whom are price-takers (i.e., they take prices as given). In other words, each investor's wealth is not significant enough to enable the investor to influence prices by her/his actions. This is the assumption underlying perfect competition in microeconomics.
- Investors can borrow and lend at the risk-free rate – this assumption can be relaxed without affecting the qualitative nature of the arguments.
- Homogeneous ▶ expectations hold (i.e., each investor views her/his investment opportunities in an identical manner). In other words, investors are identical except for potentially different wealth and potentially different attitudes toward risk. This is a strong assumption.
- There are no transaction costs and no tax implications. This is also a strong assumption, because trades involve transaction costs such as commissions and fees that depend on frequency and size of trades. Taxes depend on whether the income is from interest, dividends, or capital gains, and investors are in different tax brackets. This assumption can also be somewhat relaxed.

Like any other model, CAPM has its drawbacks, especially as a predictor of actual investment behavior by firms, but its applicability and simplicity make it a useful and popular model of risk and return. An implication of CAPM is that there would be no trades in equilibrium, since investors are assumed to be homogeneous. CAPM is not consistent with size and value effects captured by the Fama and French (1992) three-factor model:

Investors in the CAPM world care only about the systematic or undiversifiable risk of a company and its investments, and not about the portion of the variance of a firm's returns that does not covary with the market. Since the market can diversify away all diversifiable risk, the theory implies that firms should have no internal diversification needs . . . Diversified firms, however, not only exist but also are large and numerous. The CAPM does not help to explain their presence. Nor does the model always provide guidance to the diversified firm that seeks to evaluate an investment project whose risk differs from that of the firm as a whole. Presumably, the firm could estimate the systematic risk of an individual project by using the beta of a single-product firm that undertakes investments similar to the one contemplated by the diversified firm . . . [t]his is more easily said than done. Consider the extreme case where the firm is deciding whether to invest in research and development for a completely new product. . . First, there is no comparable single-product firm beta available. Second, the beta for the entire existing firm may be the incorrect one to use for the new product (Helfat 1988: 7–8).

[N]ot only might the objective function of the firm differ from that implied by the CAPM but also information problems might cause firms to have difficulty using the CAPM to evaluate potential investments . . . [I]mperfect information may well influence the way in which managers evaluate project-specific risks . . . To obtain the capital market risk-adjusted required rate of return for a project, the firm must know the systematic risk of an individual project; this requires the firm to know the covariance between the project's return and the market return (Helfat 1988: 12).

The portfolio selection model would be expected to produce different results, both in form and content, than . . . the CAPM. The form differs primarily in the following two aspects. First, the portfolio selection model yields shares of the investment portfolio allocated to different investments; the CAPM provides information about which projects to undertake but not in

what proportions. Second, the comparative spending predictions of the portfolio model . . . indicate changes in the shares of the portfolio devoted to different investments, rather than absolute spending changes. The CAPM may provide information on the direction of absolute spending changes between periods if projects change from acceptable to unacceptable (or vice versa) but does not indicate changes in investment expenditure shares of the firm's total budget

> . . . Most important, the portfolio selection model and the CAPM emphasize different types of risk. The portfolio selection model focuses on covariance risk between firm-level investments, whereas the CAPM focuses on covariance risk between the firm's investments and the market. (Helfat 1988: 30)

CAPM can provide the required rate of return for a firm's projects – this provides the 'internal rate of return' or the minimum 'hurdle rate' that a project has to yield in order for the project to be acceptable to investors, given the beta of the firm.

CAPM can also be used to set prices for regulated utilities. Given the beta of a regulated utility, CAPM can provide the fair rate of return that investors should get. The rate-setting body can set prices at levels that would generate that level of return for investors.

The CAPM is widely used to estimate a firm's cost of capital. Public estimates of beta – the covariance between the returns on the firm's stock and the returns on a market index such as the S&P 500 – are readily available.

See Also

- ▶ Arbitrage and Its Limits
- ▶ Capital Structure
- ▶ Expectations
- ▶ Hedging Strategies
- ▶ Risk and Uncertainty

References

Bhattacharya, R., and O'Brien, S. 2015. Arbitrage risk and market efficiency: Applications to securities class actions. *Santa Clara Law Review.*

Bodie, Z., A. Kane, and A.J. Marcus. 2008. *Investments*. New York: McGraw-Hill.

Brealey, R., and S. Myers. 2003. *Principles of corporate finance*. New York: McGraw-Hill/Irwin.

Damodaran, A. 2002. *Investment valuation*. New York: Wiley.

Fama, E., and K. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47: 427–465.

Helfat, C. 1988. *Investment choices in industry.* Cambridge, MA: MIT Press.

Markowitz, H. 1999. The early history of portfolio theory: 1600–1960. *Financial Analysts Journal* 55: 5–16.

Miller, M. 1999. The history of finance. *Journal of Portfolio Management* 25: 95–101.

Pontiff, J. 2006. Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting & Economics* 42: 35–52.

Sharpe, W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.

Wurgler, J., and E. Zhuravskaya. 2002. Does arbitrage flatten demand curves for stocks? *Journal of Business* 75: 583–608.

Capital Budgeting

Catherine A. Maritan
Syracuse University, Whitman School of Management, Syracuse, NY, USA

Definition Capital budgeting is a planning and decision-making process used to identify and evaluate opportunities to invest in long-term physical assets, and to allocate capital funds to selected investments.

Capital budgeting decisions are among the most critical managerial decision made in a firm because 'almost everything about a firm – its physical assets, and how they are used; its people, reputation, and skills; its products and services; its customers, channels of distribution, and brands; its financial performance – can be traced back to particular [investment decisions made] years or even decades ago' (Barwise et al. 1987: 2).

Development of the techniques used to evaluate capital investment opportunities has been largely the domain of financial economists. In

the 1950s, the pioneering work of Dean (1951) and others marked the beginning of modern capital budgeting based on the use of discounted cash flow (DCF) approaches to capture the expected economic performance of an investment. DCF methods characterize an investment project as a series of future expected cash flows that have been discounted by taking into account the time value of money in order to calculate their present value (Brealey and Myers 2003). A project that is expected to generate a return in excess of the cost of capital used to fund it is deemed to be worth undertaking. The most commonly used DCF methods are net present value (NPV) and internal rate of return (IRR), and their adoption has been widespread (Graham and Harvey 2001). Finance researchers have developed various refinements of these methods aimed at providing more accurate estimates of future cash flows, incorporating appropriate risk measures into discount rates, and delineating methods for comparing projects with different underlying characteristics. A notable relatively recent addition to DCF approaches is the application of financial option pricing models to real assets, known as ▶ real options valuation. Incorporating real options calculations into investment project evaluation allows the value of operating flexibility, or the ability of managers to make or revise decisions at a future time, to be captured by the financial models (Trigeorgis 1996).

While finance scholars have focused on valuation of capital investments, management scholars have treated financial evaluation as one component of a larger, complex organizational planning and ▶ decision-making. Beginning with the seminal work of Bower (1970), field studies of capital investments in firms have provided rich depictions of investment decision-making that recognize the effects of behavioural, political and other organizational factors on how opportunities are identified, how projects are defined, and which projects receive managerial support and approval (e.g., Bromiley 1986; Butler et al. 1993; Carter 1971). Contextual elements such as organization structure, the distribution of information in a firm, performance measurement and reward systems, cognitive limitations and cognitive biases of managers, interpersonal relationships and power structures have been found to shape investment decision processes and outcomes. This body of process research has also linked investment decision-making to acquiring competitive capabilities (Baldwin and Clark 1992; Maritan 2001) and to strategy-making more broadly (Bower and Gilbert 2005), illustrating the important role of capital budgeting in strategic management.

See Also

- ▶ Decision-Making
- ▶ Real Options
- ▶ Resource Allocation Theory

References

Baldwin, C.Y., and K.B. Clark. 1992. Capabilities and capital investment: New perspectives on capital budgeting. *Journal of Applied Corporate Finance* 5: 67–82.

Barwise, P., P. Marsh, and R. Wensley. 1987. Strategic investment decisions. *Research in Marketing* 9: 1–57.

Bower, J.L. 1970. *Managing the resource allocation process*. Boston: Harvard University Graduate School of Business Administration.

Bower, J.L., and C.G. Gilbert. 2005. *From resource allocation to strategy*. Oxford: Oxford University Press.

Brealey, R.A., and S.C. Myers. 2003. *Principles of corporate finance*, 7th ed. New York: McGraw-Hill.

Bromiley, P. 1986. *Corporate capital investment: A behavioral approach*. New York: Cambridge University Press.

Butler, R., L. Davies, R. Pike, and J. Sharp. 1993. *Strategic investment decisions*. London: Routledge.

Carter, E.E. 1971. The behavioral theory of the firm and top-level corporate decisions. *Administrative Science Quarterly* 16: 413–428.

Dean, J. 1951. *Capital budgeting: Top management policy on plant equipment and product development*. New York: Columbia University Press.

Graham, J.R., and C.R. Harvey. 2001. The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics* 60: 187–243.

Maritan, C.A. 2001. Capital investment as investing in organizational capabilities. *Academy of Management Journal* 44: 513–531.

Trigeorgis, L. 1996. *Real options: Managerial flexibility and strategy in resource allocation*. Cambridge, MA: MIT Press.

Capital Structure

C. Paul Wazzan
Berkeley Research Group, LLC, Los Angeles, CA, USA

Abstract

Franco Modigliani and Merton Miller proposed that, in 'perfect' markets, a firm's capital structure is irrelevant to its value. Subsequent research introduced real-world frictions such as taxes, bankruptcy costs, agency costs and informational asymmetry into the analysis, resulting in additional theories of optimal capital structure such as the trade-off, pecking order, free cash flow, asset substitution and debt overhang theories. Ultimately, one would expect firms to strategically pursue a capital structure that maximizes the value of the firm.

Definition Capital structure refers to the way a firm finances its operations through a combination of equity and debt. A firm that sells $1 in equity (e.g., through a stock offering) and $9 in debt (e.g., through a bond issuance) would have a capital structure of 10% equity and 90% debt.

Capital Structure in a Perfect Market

Modigliani and Miller ('MM') (1958) define a perfect market as one with no taxes; with fixed investment decisions (i.e., raising money through debt will not lead the firm to assume risky projects); with no transaction costs combined with active shareholders (i.e., investors can undo actions of the firm, such as leverage); with no bankruptcy or agency costs; and with symmetric information.

MM's first proposition is that, under these conditions, the value of a company is independent of its capital structure. The value (V) of the firm is independent of whether the firm is financed completely by debt, equity or a combination of these.

$$\text{MM Proposition 1}: \text{Value}_{\text{Levered}} = \text{Value}_{\text{Unlevered}}$$

The second MM proposition is that the cost of equity for a leveraged firm is equal to the cost of equity for an unleveraged firm, plus an added premium for financial risk.

$$\text{MM Proposition 2}: k_e = k_o + \frac{D}{E}(k_o - k_d)$$

Where: k_e is the firm's cost of equity; k_o is the firm's unlevered cost of capital; k_d is the firm's cost of debt; and $\frac{D}{E}$ is the debt/equity ratio. As leverage increases, investment risk is shifted from equity holders to debt holders, but total risk is conserved, and hence no extra value is created. While the MM theories are well understood and generally accepted, markets are not in fact perfect: this gives rise to the following additional theories of capital structure.

Trade-Off Theory

Modigliani and Miller (1963) extend their initial analysis to include debt. In most countries with developed capital markets, corporate profits are taxed, and interest paid on debt is generally treated by firms as a deductible expense. The basic insight of MM theory is that the tax deductibility of interest makes debt financing valuable because the cost of capital decreases as the proportion of debt in the capital structure increases.

Example 1

A firm generates $100 in profits with certainty in every period in perpetuity; the risk-free rate (Rf) is 10% and the corporate tax rate (TC) is 40%. If the firm is 100% equity financed, shareholders receive $(1 - 0.40)^{\star}\$100 = \$60$ in each period. The value (V) of the firm is equal to the value of equity (E): $\$60/0.10 = \$600$. Now assume that the firm takes on $500 in 10% debt and pays current equity holders a $500 dividend. Shareholders now receive $(1–0.40)^{\star}$ ($100–$50) = $30 in each period, which implies that (E) = $30/

0.10 = $300 and V = $300 + $500 = $800. Equity holders are now worth $800 (the received dividend payment of $500, plus the $300 that they still hold in equity).

At the limit, one might assume that the firm would continue to issue debt until the firm was completely debt financed. Unfortunately, this would be incorrect, as one must also consider the direct costs (e.g., management time spent dealing with creditors; legal expenses; bankruptcy) and indirect costs (e.g., financial distress that creates a ▶ moral hazard wherein equity holders take actions that adversely affect bondholders) of having too much debt. As the amount of debt increases, the probability of financial distress and, ultimately, bankruptcy increases. This leads to the trade-off theory, whereby a firm will increase debt until the marginal cost of bankruptcy equals the marginal increase in tax savings.

Pecking Order Theory

▶ Asymmetric information involves a situation in which one party has better or more complete information about the state of the world (or probabilistic outcomes) than another party (i.e., the seller knows more than the buyer about the true value of a firm). This leads to two main types of problems: ▶ Adverse selection and moral hazard. The firm's actions (i.e., raising money through debt or equity) 'signal' information to investors about managers' beliefs in the value of the firm.

Myers and Majluf (1984) argue that equity is a less preferred means to raise capital because when managers, who are assumed to be better informed than investors about the true condition of the firm (e.g., are privy to non-public information), issue new equity, investors believe that managers think that the firm is overvalued and that managers take advantage of this informational asymmetry. As a result, investors will place a lower value on the new equity issuance. Moreover, since debt obligations are paid first, while equity holders are the residual claimant, investors in debt are less exposed to errors in firm valuation. Consequently, issuing debt is less subject to this informational asymmetry problem, and as long as debt is fairly valued, this would be preferable to equity. Equity would only be issued under conditions where debt was costly (e.g., the firm already has a very high debt/equity ratio and the costs of financial distress or bankruptcy are great).

This leads to pecking order theory: companies prioritize, or order, their sources of financing from least expensive (internal) to most expensive (external). Internal equity (e.g., free cash flow) is used first, followed by debt issuance (since owners maintain full ownership), followed by equity issuance (which would dilute ownership).

Free Cash Flow Theory

Jensen and Meckling (1976) indicate that agency costs are inevitable: corporate managers will act in their own self-interest (e.g., invest in wasteful projects; initiate takeovers or mergers to expand personal power; take projects that make them indispensable to the firm and increase their bargaining power vis-à-vis shareholders). Shareholders attempt to align corporate managers' interests with their own interests by a number of methods including board of director oversight or compensation packages. Jensen (1986) hypothesized a free cash flow theory that seeks a solution to the problem of managers unwisely spending the firm's free cash flow (i.e., by growing the firm to increase their own power rather than maximizing shareholder value). Jensen's theory states that debt may have the beneficial effect of forcing a firm to pay out cash that would otherwise be misused. The free cash flow theory can be summarized thus: unless free cash flow is given back to investors, management has an incentive to destroy firm value through empire building and perquisites. Increasing leverage imposes financial discipline on management. Timely payment of debts and dividends can also be monitored by outsiders and taken as a signal of financial health.

Asset Substitution Theory

Jensen and Meckling (1976) define asset substitution as stockholders' choosing risky projects that expropriate value from debt holders. As the debt/equity ratio increases, stockholders have an increased incentive to undertake risky (and in some cases negative net present value (NPV)) projects. If the project is successful, shareholders get all the upside (i.e., because debt holders receive a fixed-interest payment regardless of how large the profits are); if it is unsuccessful, debt holders get much of the downside (i.e., the firm goes bankrupt).

Example 2

Assume a firm is faced with selecting between two projects with the following payoffs (good and bad outcomes are equally likely).

	Bad outcome	Good outcome	Expected value
Project 1	$50	$100	$75
Project 2	$25	$115	$70

Shareholders will opt for project 1 because the expected value is higher. Now assume that investors must borrow $40 in order to finance these same two projects. The payoffs are now as follows.

	Bad outcome	Good outcome	Expected value
Project 1	$50 − $40 = $10	$100 − $40 = $60	$35
Project 2	$25 − $40 = $0	$115 − $40 = $75	$37.50

Shareholders will now opt for project 2 because the expected value of that project is higher. Note that bondholders would prefer project 1 because they would receive full payment of $40 in either outcome. In the event of a bad outcome, project 2 returns only $25 to bondholders and zero to shareholders.

Underinvestment or Debt Overhang Theory

Myers (1977) defines the underinvestment or debt overhang problem as what occurs when the preexistence of debt causes gains from future positive NPV projects to accrue to debt holders rather than shareholders. Thus, management has an incentive to reject positive NPV projects, even though they have the potential to increase firm value.

Example 3

Assume a firm has $90 in senior debt and is already invested in a project with equal probability of good and bad outcomes with the following payoffs (Table 1).

Now assume that the firm has the opportunity to invest $5 (which it must raise through junior debt) in a new project that returns $6 in both the good and bad states. This is a positive NPV project (i.e., NPV of $1 in all states) and would normally be taken by the firm. The problem is that the gains from making this investment accrue to the senior debt holders, as shown below (Table 2).

*Junior debt holders must receive $10 in the good state in order to agree to make the

Capital Structure, Table 1

Bad outcome	Good outcome	Senior debt expected value	Shareholder expected value
$0	$100	0.5($0) + 0.5($90) = $45	0.5($0) + 0.5($10) = $5

Capital Structure, Table 2

Bad outcome	Good outcome	Senior debt expected value	Junior debt expected value*	Shareholder expected value
$0 + $6 = $6	$100 + $6 = $106	0.5($6) + 0.5 ($90) = $48	0.5($0) + 0.5($10)	0.5($0) + 0.5 ($6) = $3

loan, because they know that in the bad state they receive nothing.

Note that shareholders, by investing in a positive NPV project, are actually worse off – an expected value of $3 versus $5 without taking on the new project and junior debt.

Conclusions

This article presents the fundamental Modigliani and Miller capital structure theorems and then introduces real-world frictions, which generate a number of additional theories that impact capital structure. Myers (2001) correctly states that there is no single universal capital structure theory and notes that, in practice, firms across industries and even within industries have dissimilar capital structures. Myers concludes that capital structure theories should be considered as conditional – there are real-world examples of all of the theories that impact on firm behaviour.

See Also

- ▶ Adverse Selection
- ▶ Agency Problems
- ▶ Agency Theory
- ▶ Asymmetric Information
- ▶ Capital Asset Pricing Model (CAPM)
- ▶ Moral Hazard
- ▶ Principal Agent

References

Jensen, M.C. 1986. Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review* 76: 323–329.

Jensen, M.C., and W.H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Management* 3: 305–360.

Modigliani, F., and M.H. Miller. 1958. The cost of capital, corporate finance, and the theory of investment. *American Economic Review* 48: 261–297.

Modigliani, F., and M.H. Miller. 1963. Corporate income taxes and the cost of capital: A correction. *American Economic Review* 53: 443–453.

Myers, S.C. 1977. Determinants of corporate borrowing. *Journal of Financial Economics* 5: 147–175.

Myers, S.C. 2001. Capital structure. *Journal of Economic Perspectives* 15: 81–102.

Myers, S.C., and N.S. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13: 187–221.

C

Capitalism

Hamid H. Kazeroony
Walden University, School of Management, Minneapolis, MN, USA

Abstract

Capitalism is a dialectical system of production explaining how the exchange of money for labor transfers wealth from the laborer, producing products/services to controllers of the processes, namely, the owners of capital. Capitalism through facilitating the conversion of raw material to finished products and services, using the unpaid portion of the labor, the surplus value, used in the production, accumulates wealth. Capitalism as a mode of production is one historical epoch, transforming local feudalist systems into global industrial production system. Capitalism as a system of production includes social, economic, and political forces and dynamics, each giving rise to its dialectical movement.

Definition Capitalism is a dialectical system of interactions and production, rooted in Marxist philosophy, in which the surplus value of labor is transferred to the owners of capital (the capitalists) as profit, and social relationships are subjected to a complex set of values emanating from the individual and aggregate interactions, exchanging labor, commodities, and money.

Capitalism Overview

Capitalism is a dialectical process connecting historical epochs through constant reconciliation of

theses and antitheses, creating new socioeconomic-political syntheses, moving the production forces without clear direction. Capitalism was articulated by Marx and Engels (1981) who outlined the basic arguments for the dialectical transformation of feudalism into capitalism. They argued that social structure became polarized between peasants and landlords, creating a pathway to capitalism where economic value of labor was transferred to the owners of capital through the bourgeoisie's efforts in facilitating social and economic interactions between the two forces. Capitalism, as outlined in *The Communist Manifesto*, was fully explained from economic, social, and political perspectives in various Marx's writings.

Dialectical Materialism Process

Through correspondence with Proudhon, Marx (1992), in an effort to clarify the concept of capitalism, began developing the notion of historical determinism and its effects on mode of production. In his critique of German philosophy, Marx (1993), by challenging Hegelian dialectical idealism, based on the influences of Feuerbach (Tucker 1972), developed the dialectical materialism to better explain dynamics of capitalism. Dialectical Materialism is significant in defining capitalism because it (a) "is the process of material production, determining all other aspects of social life" (Afanasyev 1987, p. 36), (b) is a nonlinear motion continually reconciling the opposites (theses and syntheses), creating new syntheses, and (c) interchangeably transforms social, political, and economic structures and relationships. Dialectical materialism explains how the mode of production changes the social relationship which in turn manifests itself in new economic relationships, giving rise to new classes, creating new political systems supporting the emerging ecosystem. For example, feudalism witnessed the rise of industrial advances such as steam engines, textile machines, and the rise of factory production, competing for the labor working on land as peasants. Although the technology was initially used to facilitate the transformation of agricultural products such as cotton to cloth and make higher-quality products to expand on the riches of landowners, the method of production required disengagement of some laborers from farming, moving them to production lines. The thesis (the need for peasants paying high rents for the land) and antithesis (the need for cheap labor in the factories) led to a new synthesis (a class of workers depending on industrialists for their livelihood). Each economic synthesis creation transformed social, economic, and political structures dialectically to what we now know as capitalism.

The dialectical materialism process operates in an open system (Bertalanffy 1973; Checkland 1999) where any new synthesis created in economic, social, and political structure can lead to changes in other components of capitalism.

Capitalism as Explained by Marx

Marx applied dialectical materialism to history to explain the nature of socioeconomic and political change through time. Marx argued that dialectical materialism is manifested by the antagonism of production forces in each epochal stage of history, creating changes, moving the socioeconomic and political structure in different directions which is nonlinear, although its progression, at some point, may appear as such (Kazeroony 2005). In essence, Marx's idea of dialectical materialism rests on the concept of complexity where random patterns can dynamically affect each other's behavior in different ways and directions (Kazeroony 2005). Within this context, Marx observed the following relationship between production forces:

$$\mathrm{RW} + L + \mathrm{PR} + C1 = \mathrm{Pr} + F + C2$$

where RW is the cost of raw material input into production/services; L is the cost of price; PR is the cost of processes such as moving the raw material from mine to factory, marketing, etc.; $C1$ is the required capital which is either borrowed at the cost (interest) or the owner of capital seeks to compensate for the opportunity cost of its capital by determining an interest; Pr. is the price of finished product/service; F is the profit for the

owner of capital (manufacturer, service provider, industrialist); and *C*2 is the retained capital for perpetuating the production of goods/services.

Rise of Global Capitalism

Through the nineteenth century as industrialization took hold and expanded in the twentieth and the twenty-first centuries, the European and later American industrialists, to feed the emerging capitalism as a dominant mode of production, began aggressive international trade to open markets for their excess production at home and obtain raw material at cheap prices from other countries. The European and American economic expansionism led to the globalization of economic activities, creating structural changes and causing fundamental shifts in economic and political disparities, creating mass disfranchisement for various ethnicities and countries, particularly, African and Asian countries, creating the capitalist world order (Wallerstein 1979).

Global Capitalism and Its Future

In a highly globalized capitalism, "transnational capitalism builds a system of institutions that dominates the structure of nation-state, exceed their functions, facilitate network of supranational integration" (Hernandez 2014, p. 23). Technological innovation such as Cloud systems for information and application storage and development across borders, computerized financial transactions among private entities across several borders instantaneously, conducting hedging activities, insurance transactions, banking, currency conversion, and moving profits produced by owners of capital between various countries, leaves capitalists at an advantage vis-à-vis governmental entities, empowering capitalism while reducing the power of officials who are designated to serve the interests of disfranchised and the worker who owns the labor and has produced the products/services and responsible for creation of surplus value.

Global capitalism's achievements are measured by factors such as productivity, inflation, and gross domestic product, just to name a few, all of which are subject to arbitrary definition. However, global capitalism perpetuation rests on the behavior of those consumers who continue to accumulate products/services despite any individual propensity to use them continuously, leading to the waste of natural resources consumed in the process, inevitably making raw material more expensive and depleting natural resources. On the other hand, there are those consumers who, despite their propensity to use basic products/services continuously and effectively, do not have sufficient financial means to afford buying them and, therefore, being disfranchised and continuously marginalized in the global capitalist system. Over time the dislocation of resources and needs has spread among countries and within each country. In the second half of the twentieth century and the first decade of the twenty-first century, from oil crisis of 1970s to financial crisis of 2008, disparities between rich and poor among individuals and nations have become more pronounced questioning the future direction of capitalism.

Additional Reading and References

Afanasyev, V.G. 1987. *Dialectical materialism*. Rev. ed. New York: International.

Bertalanffy, L.V. 1973. *General system theory: Foundations, development, applications*. Rev. ed. New York: G. Braziller.

Checkland, P. 1999. *Soft systems methodology: A 30-year retrospective*. Chichester: John Wiley.

Getting Started With Trading. n.d. Retrieved October 26, 2014, from CME Group website: http://www.cmegroup.com/education/getting-started.html.

Hernandez, J.V. 2014. *Capitalism and the social relationship: An organizational perspective*, ed. H. Kazeroony and A. Stachowicz-Stanusch, pp. 16–35. Basingstoke: Palgrave Macmillan.

IMF eLibrary Data. n.d. Retrieved October 26, 2014, from International Monetary Fund website: http://www.imf.org/external/data.htm#guide.

Kazeroony, H.H. 2005. *Leaders' perception of organizational change*. ProQuest Dissertations and Theses, 219. Retrieved from http://search.proquest.com/docview/305348856?accountid=35812. (305348856).

Kazeroony, H., and Stachowicz-Stanusch, A. 2014. Capitalism and the social relationship: An organizational perspective. (Ed.). Basingstoke: Palgrave-Macmillan, UK.

Learning Centre. n.d. Retrieved October 26, 2014, from London Stock Exchange website: http://www.londonstockexchange.com/news/learning-centre/learning-centre.htm.
Marx, K. 1976. *Capital: A critique of political economy*. Vol. I, Trans. B. Fowkes. New York: Penguin Books.
Marx, K. 1992. *The poverty of philosophy*. New ed. New York: International.
Marx, K. 1993. Grundrisse: Foundations of the critique of political economy (rough draft). Trans. M. Nicolaus. London: Penguin Books in association with New Left Review.
Marx, K., and Engels, F. 1981. *The communist manifesto*. Trans. S. Moore. Harmondsworth: Penguin.
Tucker, R.C. 1972. *Philosophy and Myth in Karl Marx*. 2nd ed. Cambridge: Cambridge University Press.
Wallerstein, I.M. 1979. *The capitalist world-economy: Essays*. Cambridge: Cambridge University Press.

Capturing Value from Advantages

Christos N. Pitelis
Brunel University London, Uxbridge, UK

Abstract

This article gives an overview of the literature on the issues of capturing ▶ value from advantage. It discusses the foundational work of ▶ Stephen Hymer and his recognition of the importance of foreign direct investment (FDI). The article then explores the contributions of ▶ David Teece, who considers the strategy to be adopted by an innovator seeking to maximize the ▶ profit from innovations and highlights the importance of complementarities. The work of later scholars, such as George Richardson and ▶ Edith Penrose, is then considered. Finally, there is some discussion of real-world examples in which the principles of the theory can be seen to provide considerable illumination. The article concludes with some points about the current gaps in the literature and possibilities for future research.

Definition Capturing value from advantages is the process whereby a firm acts in such a way as to leverage (and thereby profit from) the advantages it has over other firms. This may involve arrangements such as licensing and franchising its goods or services and cooperating with other firms.

'Profiting (or capturing ▶ value) from advantages' is a time-honoured theme in industrial organization (IO) and international strategic management. The originator of the theme is arguably ▶ Stephen Hymer (1970), whose 1960 Ph.D. thesis at MIT served to establish him as the father figure of the theory of the multinational enterprise (MNE), foreign direct investment (FDI) and international business (IB) (see Dunning and Pitelis 2008). Hymer claimed that the need of firms to exploit (and therefore ▶ profit from) their advantages is a crucial factor in their decision to undertake FDI rather than alternative modalities of foreign operations, such as ▶ licensing, franchising and cooperation.

Hymer focused mainly on 'monopolistic advantages' by firms and the efforts by firms to capture monopolistic rents from these advantages, rather than on the process of their derivation, that is, the creation of advantages. This 'profiting from advantages' theme was inspired by the earlier work of Jo Bain (1956). Bain had attributed the ability of firms to charge prices in excess of the (perfectly) competitive ones, therefore to capture oligopoly rents, as barriers to entry, such as absolute cost advantages, economies of scale and product differentiation (or preference barrier). Bain paid particular attention to the underlying advantages afforded to firms through such barriers to competition. Hymer (1976) explicitly drew on Bain's analysis of advantages and extended it to the area of MNEs. Bain and Hymer paid limited attention to the role of innovation as an advantage (Dunning and Pitelis 2008). This would partly explain their focus on 'profiting from advantages' (or capturing supernormal profits) rather than on the value-creating properties of advantages.

▶ David Teece (1986) helped to address these gaps. In this article, entitled 'Profiting from technological innovation', he addressed the issue of how an innovator can enhance the chances of capturing as high a share as possible from the value created by an innovation. In particular, Teece observed that it is common for innovators to lose out to competitors who possess

complementary assets and capabilities. In this context, business strategy can be crucial, mainly in the form of attempts by the innovator to establish a base of complementary assets and capabilities. Assuming absent or failing markets for ideas, the choice of the mode of acquisition of such assets and capabilities will, in turn, depend on factors such as the 'appropriability regime' (how easy it is to protect an innovation, through, for example, patents or secrecy) and the existence or otherwise of a 'dominant design' (the emergence of which tends to shift competition from design to price). In brief, when the 'appropriability regime' is 'weak' and there exists a 'dominant design', innovators are well advised to acquire complementary assets and capabilities such as distribution through, for example, contracts, integration or collaboration. The choice of modality will depend crucially on the nature of complementary assets and the competitive position of the innovator vis-à-vis its rivals.

Despite dealing with the archetypal concern of transaction costs theory (the choice between contracts, integration, collaboration), and in contrast to Hymer (1968), who synthesized transaction costs and oligopolistic rivalry, Teece took a route more akin to the resource-based view (RBV thereafter), and the (dynamic) capabilities perspective, which he subsequently co-founded. His resource and capabilities lens allowed him to go beyond Hymer's theme and to include innovation as an advantage which creates value (but it is not by itself sufficient to capture this value).

The Bain–Hymer–Teece tradition can be usefully complemented by other seminal contributions, notably those of ▶ Edith Penrose (1995) and George Richardson (1972). In his 1972 article, Richardson built on an earlier insight from Penrose (1995) to explore the choice between market, integration and cooperation. He used two main categories: similarity of activities and complementarity of activities. Similar activities were said to be those whose efficient implementation required the same underlying capabilities. Activities were complementary when efficient dispensation, requisites and underlying capabilities were used jointly. In this context, Richardson observed that production-related efficiency dictates that firms considered integrating when their (planned) activities were both similar and complementary to those of other firms (with which they might intend to integrate). Cooperation was best when there existed complementary but dissimilar activities. Weakly complementary, dissimilar activities were best left to markets. Teece's focus on complementary capabilities and assets is in line with Richardson. Teece took the analysis further in that complementarity of activities and assets is seen as both a reason why innovators may fail to profit from their innovation and as a reason for the choice of modality – that is, integration vis-à-vis market and cooperation. The role of complementary assets as a prerequisite for ▶ profiting from innovation is a significant standalone contribution. For the choice of modality, Richardson's concept of similarity of activities is likewise a useful addition to Teece's analysis. In the presence of dissimilar activities, market or cooperation arguably dominates integration.

The problem of capturing value/profiting from innovation and other firm-specific advantages in a dynamic environment characterized by Schumpeterian competition is also an important aspect of the work of Penrose (1995). She claimed that in order to achieve sustainable long-term performance firms needed to build 'technological' or 'relatively impregnable bases'. For Penrose (1995: 137):

> In the long run the profitability, survival, and growth of a firm does not depend so much on the efficiency with which it is able to organize the production of even a widely diversified range of products as it does on the ability of the firm to establish one or more wide and relatively impregnable 'bases' from which it can adapt and extend its operations in an uncertain, changing, and competitive world.

'Relatively impregnable bases' were seen by Penrose as technological and know-how-based bundles of tacit knowledge which are hard for rivals to initiate. The concept is akin to, and supportive of, Teece's arguments, which focused on the need to acquire assets and capabilities that allow innovators to capture value. Both pre-date the subsequent focus of the RBV on VRIN-type resources, namely those which are Valuable, Rare,

Inimitable and Non-Substitutable (Barney 1991). An important part of the Penrosean story, however, is that innovation can itself be an element of a 'relatively impregnable base', and that the last mentioned may involve more than the acquisition of complementary assets and capabilities. An implication is that the acquisition of complementary assets and capabilities may not suffice to allow firms to profit from an innovation. Building 'relatively impregnable bases' can be expensive and time-consuming and is more likely to be achieved by established firms. Richardson and Penrose did not deal with transaction costs. These are of essence when dealing with the choice of modality, as we now know from Coase (1937), Hymer (1990), Williamson (1975) and subsequent literature (see Williamson 2005, for an overview). Teece's (1986) reference to transaction costs was not extensive, despite his own earlier contributions (e.g., Teece 1982). His choice of transaction costs analysis focuses on intangible assets-related problems similar to those analysed by Buckley and Casson (1976) for the case of the MNE. It could be suggested that additional considerations such as, for example, the importance of post-contract 'bilateral oligopoly'-type problems (Williamson 1975; Hymer 1990) could be taken into account. For instance, firms could select integration to reduce transaction costs arising from post-contract hold-ups involving small numbers (bilateral monopoly).

If we take as an example EMI's failure to capture value sustainably from its invention of the CT scanner, the following insights from the aforementioned contributions may be useful. As noted by Teece, EMI possessed neither similar nor complementary capabilities for the production and exploitation of the CT scanner. In this context its best option might be to sell or license the idea/technology. However, the market for ideas/technologies is notoriously imperfect, restricting this option for EMI. In addition to the well-known problems discussed in the literature (such as 'Arrow's paradox', due to the 'public goods' nature of knowledge, see, e.g., Buckley and Casson 1976), and/or the presence of opportunistic buyers (as in Williamson 1975; Hymer 1990), a genuine (non-opportunism-related) problem in this case is the prediction of demand. In EMI's case the early predictions proved to be very pessimistic (Bartlett 2005). In such a context, even an honest buyer might be unwilling to buy and, in any case, unwilling to pay a price above that consistent with demand projections at the time. Such a potential buyer might have been Siemens, or other companies in the medical equipment sector. Given problems with selling (for then CEO Powell, this would be 'selling our birth-right') (Bartlett 2005: 194), an alternative possibility examined by Teece was for EMI to collaborate with a company that possessed the capabilities required for production and distribution of CT scanners. However, in this case the problem of valuing the technology is still present, leading to high pre-contract transaction costs, but also high post-contract transaction costs due to conditions of 'bilateral oligopoly' (Williamson 1975; Hymer 1990). In addition, EMI would carry a risk of its technology being expropriated by other firms, which possessed the requisite production and complementary capabilities for the exploitation of the innovation, were it to choose to partner with them.

The moral is that when there exists stronger and better positioned competitors, one faces a world of very imperfect choices. In the absence of a strong appropriability regime, for example through a very strong patent and/or maintaining industrial secrecy (see Bartlett 2005), EMI could only hope to either acquire or to gradually build complementary assets and 'impregnable bases' in order to achieve comparable competitive strength to its rivals, and then aim to profit fully from its innovation.

Given its stronghold in its existing sector, it was a realistic option for EMI to acquire an existing player(s) in the US, its target expansion market. Teece did not explore this possibility; it was not pursued by EMI either (Bartlett 2005). This is almost paradoxical, as going alone involved many risks, aired at the time within the company, which, like strong competitors, was characterized by an absence of manufacturing capabilities and a lack of knowledge of the US market (Bartlett 2005). Had EMI chosen the route of diversification through acquisition (brownfield investment), like, for example, Santander and

CEMEX, its fortunes might have been different – we will never know. The choice of 'greenfield' foreign direct investment may explain, at least in part, its failure to profit significantly from its innovations.

Importantly, the possibility of acquiring rivals is normally unavailable to small firms and/or individual innovators. Considering the very substantial resource and (transaction) costs required (to acquire) or build complementary assets and level the playing field, it is hardly surprising that the best some start-ups can often hope for is to be taken over by a larger firm. This raises the important question of how a small firm and/or individual innovator can capture value from their innovation/advantages, and what role business and public policy can play in this context.

To conclude, Richardson's analysis complements Teece's suggestion that EMI's decision to 'make' through greenfield investment was theoretically ill-advised to start with. Transaction costs arguments, on the other hand, point to the limitations of both cooperation and market-based strategies. Once greenfield investment had been chosen from the three imperfect alternatives, the building of complementary assets was, according to Teece, the only choice. This is also difficult, however, as proved to be the case with EMI. Gans et al. (2001) and Gans and Stern (2003) confirmed the view that the choice of the type of competition by small and medium-sized enterprises (SMEs) will depend on the degree of imperfection of the market for ideas. For example, in the biotechnology sector, where patterns are relatively effective and there exists a 'market for ideas', cooperation with larger players is more common than in electronics, where the absence of such conditions obliges SMEs to attempt to compete head-on with existing competencies. This raises important questions for both business policy and public policy, which are beyond our scope in this entry.

Despite its contribution, there are important underlying issues, not explicitly discussed, in the 'advantages' literature, notably the issue of the nature and determinants of value creation and value capture, the relationship between value capture and value creation and (their impact on) the sustainability of the value-creation process. In addition, the focus on value capture normally takes the stance of the firm and/or the home country. This raises the question of the relationship between a firm's (or a nation's) value-capture strategies and the overall sustainability of the value-creation process. More recent extensions of the 'advantage' perspective by Pitelis (2009) and Mahoney et al. (2009) address some of those limitations.

See Also

- Comparative Advantage
- Hymer, Stephen Herbert (1934–1974): The MNE and International Business
- Licensing
- Multinational Corporations
- Penrose, Edith T. (1914–1996)
- Profit
- Profiting from Innovation
- Teece, David J. (Born 1948)
- Value

References

Bain, J.S. 1956. *Barriers to new competition: Their character and consequences for manufacturing industries*. Boston: Harvard University Press.

Barney, J.B. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17: 99–120.

Bartlett, A.C. 2005. EMI and CT scanner [A] and [B]. In *Cases to accompany contemporary strategy analysis*, ed. M.R. Grant. Oxford: Blackwell.

Buckley, P.J., and M.C. Casson (eds.). 1976. *The future of multinational enterprise*. London: Macmillan.

Coase, R.H. 1937. The nature of the firm. *Economica* 4: 386–405.

Dunning, J.H., and C.N. Pitelis. 2008. Stephen Hymer's contribution to international business scholarship: An assessment and extension. *Journal of International Business Studies* 39: 167–176.

Gans, J.S., and S. Stern. 2003. The product market for 'ideas': Commercialization strategies for technology entrepreneurs. *Research Policy* 32: 333–350.

Gans, J.S., D.H. Hsu, and S. Stern. 2001. When does start-up innovation spur the gale of creative destruction? *RAND Journal of Economics* 33: 571–586.

Hymer, S.H. 1976. *The international operations of national firms: Study of direct foreign investment*. Cambridge, MA: The MIT Press.

Hymer, S.H. 1990. The large multinational corporation. In *Multinational corporations*, ed. M. Casson. Hants: Edward Elgar.

Mahoney, T.J., A.M. McGahan, and C.N. Pitelis. 2009. The interdependence of private and public interests. *Organization Science* 20: 1034–1052.
Penrose, E.T. 1995. *The theory of the growth of the firm*, 3rd ed. Oxford: Oxford University Press.
Pitelis, C. 2009. The co-evolution of organizational value capture, value creation and sustainable advantage. *Organization Science* 30: 1115–1139.
Richardson, G.B. 1972. The organisation of industry. *Economic Journal* 82: 883–896.
Teece, D.J. 1982. Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior & Organization* 3: 39–63.
Teece, D.J. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy* 15: 285–305.
Williamson, O.E. 1975. *Markets and hierarchies: Analysis and antitrust implications: A study in the economics of internal organization*. New York: Free Press.
Williamson, O.E. 2005. The economics of governance. *American Economic Review* 95: 1–18.

Case Method, the

Joseph Bower
Harvard University, Business School, Boston, MA, USA

Abstract

The case method was developed concurrently with the emergence of business schools as a way of teaching future executives evidence-based problem solving in the classroom. Harvard Business School faculty led in developing the method. A particular challenge in the writing of cases is finding the balance between enough complexity so that the problem posed reflects reality and supports alternative approaches to resolution, and too much complexity, which makes it impossible for the student to prepare. A great virtue of the method is that it replicates the managerial work involved in solving a problem within a group.

Definition The case method is the use of a case study as a basis for classroom discussion whose purpose is the discovery, in class, of generally useful concepts, or the means of using concepts to resolve specific problems.

The use of the 'case method' in the teaching of business skills has its origins in the development of the Harvard Business School (HBS). But the case method itself has older origins, in the training of lawyers in the Anglo-Saxon Common Law tradition. In turn, the clinical training of physicians is reflected in the adaptations of the ancient apprenticeship system to the modern medical setting. The origins of the method are important to an understanding of the uses to which case method teaching have been put in management education.

Common Law is different from its continental or Chinese cousins in that its dictates are the consequence of cumulative precedent, the resolutions of cases tried before judges and juries applying legislated law, rather than the judicial application of a centrally derived code. The cases used in the training of lawyers in the Anglo-Saxon system are the opinions written by judges to resolve controversies that have the form of an argument between a plaintiff and a defendant, such as *Brown* v. *Board of Education of Topeka Kansas*. Students studying that case examine the reasoning used to change the law regarding segregation in the provision of education in the USA. The facts they consider are those summarized by the judge and the reasoning will be that of the judge dealing with the facts and arguments of the lawyers in question. The students will not discuss what happened in the courtroom, the tactics of the lawyers, the nature of the appeals process, or other forces that might have affected the specific outcome. Their subject matter is the law, not 'lawyering'. At least since the late nineteenth century in the US, young lawyers studied law at university, but they then went on to learn the practice of law through serving an apprenticeship – as had been the case for centuries. Even today in the US, graduating from law school does not make one a lawyer. One is still required to pass a state bar exam.

In medical schools, 'cases' perform almost the opposite function. Medicine is taught as a science. Topics such as anatomy and infectious disease are studied in courses using books and articles that compile contemporary understanding. But the application of that knowledge in the curing of

patients is taught in the hospital or clinic by observing doctors working with patients. Later on, this involves learning by doing, in effect, practising the new-found knowledge on the patient. In other words, the students learn medicine in the classroom but they learn 'doctoring' in the clinical setting. The chief difference between the training of doctors today and in the past lies in the science that they are applying and in the range of tools and drugs that are available to them.

At the Harvard School of Business Administration, a different approach was adopted from the outset. The first dean, Edwin F. Gay, wrote that

> Unlike the older professions, with their well-established University instruction and tried methods, Business as a department of University training, has still, to a large extent, to invent its appropriate means of instruction and to form its own traditions. From the mass of accumulating business experience, a science must be quarried. Not only must the fundamental principles guiding conservative business be elucidated, but the art of applying those principles in various fields of business enterprise must be taught in a scientific spirit. What for lack of a better term, may be called the 'laboratory method' of instruction must be introduced, wherever possible, if the School is to fulfill efficiently the intention of its founders. (Copeland 1957: 27)

In a later statement, Gay framed the objective more sharply: 'In the courses on Commercial Law, the case system will be used. In the other courses an analogous method, emphasizing classroom discussion in connection with lectures and frequent reports on assigned topics – what may be called the "problem method" will be introduced as far as practicable' (Copeland 1957). This process took a considerable amount of time – the first cases, as we would recognize them today, being introduced only in the 1920s. Yet in 1911–1912, A. W. Shaw had started a course in Business Policy (today often taught as separate courses in Strategy, Strategic Management or Organization Management) which featured living cases. A local business would visit the class to describe a certain problem. Students would write a report on the problem and then their ideas would be reviewed by the business and by Shaw.

Very shortly after the systematic collection of certain industry data commenced in the Bureau of Business Research, Melvin T. Copeland was encouraged by the School's next dean, Wallace Donham, to publish a book of case studies in his field of commercial organization – what later became known as marketing. Soon after the 1920 publication and use of the casebook, the Research Bureau was asked to begin the collection and writing up of business cases, work that continues to the present day under what is now called the Division of Research and Course Development.

The question 'Why the push for cases?' gets to the heart of case method teaching. An emphasis on cases was encouraged because they were seen to provide a foundation for discussion-based learning. Today it is well established that problem-based learning is the most powerful way of engaging students in acquiring knowledge and skills for subsequent use in solving problems. Psychologists have written and lectured at considerable length about this proposition. What Dean Gay and his colleagues understood from the start was that the habit of disciplined data-based problem-solving required training in the use of analysis to make a choice. They could try to do this with live cases or by posing problems to the class based on current events or materials to which they had access. But to approach the development of a body of knowledge around a topic of importance, say consumer marketing or corporate finance, meant studying problems in the field in a comparative fashion that provided the basis for generalization, turning those studies into cases that were concise enough for students to prepare so that *they in turn* could 'quarry the ideas' for themselves and, by making them their own, make them useful.

In other words, the case method is a way of helping students to learn how problems may be solved and decisions made in a social setting – the work of management. In the process, through induction from a carefully conceived sequence of cases, students find the generalizations that lie at the basis of particular fields. However, it must be stressed that these are only generalizations because the essential idea underlying case method teaching is that there is no single right answer to a good case. There are acceptable answers

associated with plans of action that make sense. There are arguments that make sense and others that do not, given a set of facts. But if the question relates to commercial or managerial action – as opposed to a piece of quantitative analysis or the appropriate accounting for a transaction – then there will be alternative approaches in all situations. Indeed, finding a new alternative when none has been offered may be much the best answer, even if it is not well worked out.

To make this approach successful, cases must present a situation, describing its substance and context with sufficient completeness. This must be supplemented by adequate exhibits outlining aspects such as financial performance, markets and competition, organization and historical background. Writing a 'good' case is challenging for most faculty members, since a teaching case is not the same thing as a case for the classroom. A case that is collected for purposes of research tends to be as complete and detailed a description as possible. But for a case to provide the basis for discussion, it must be short and clear enough so that students are able to prepare it (which implies several readings as well as careful analysis) in the time available for that process – usually not more than 2 h. They must be able to argue their perspective based on the data so that artfully led discussion permits the discovery of new ideas and concepts.

Obviously, this will not be the approach taken by a faculty member seeking only to illustrate the use of a tool. A case may work for that purpose, but then the subject matter is a kind of applied engineering – using a tool when that tool provides the answer. There may be other ways of designing illustrations of tools. More interesting is when the case is used to explore whether the tool is a useful way of dealing with the problem. Then the students will learn both how the tool may be used, and when it may not be appropriate and why.

The case method class can be opened in many ways, but classically it begins with the problem facing the case protagonist: what should Smith do about Johnson's proposition? The goal is set: any discussion will have merit because it will help resolve Smith's problem. The early discussion is used to frame the alternatives available and set the agenda for what needs to be discussed. The instructor will have spent a great deal of time before class pondering how to raise questions so that key aspects of the case are considered if they are not raised by the students. But in good case discussions, debate among the students may go beyond what has been considered by the instructor.

The path of a class depends upon the teaching objective, but a skilful case method instructor will have the class inducing key ideas from the case, from comparison with other cases, and from experience with using tools and concepts provided in readings or 'technical notes'. In a strategy course, for example, three parallel cases on participants in a single industry may provide the basis for developing important ideas about successful approaches to competition in an oligopolistic setting. In a marketing course, a case in which a manufacturer has a frustrating time getting its retail customers to present a coherent pattern of pricing to the final consumer may introduce students to a range of trade practices companies might adopt, their legality and their effectiveness.

During a class there are many opportunities for an instructor to pause to introduce ideas, or to lift the unit of analysis in discussion from an individual firm to a class of firms. The end of class provides the opportunity to highlight points made by the students in a way that drives home *their* findings. The instructor needs to be careful, however, for the more the summary becomes a lecture that was prepared before class, the less the members of the class perceive themselves as having taken part in a self-managed learning experience. The case method is not an indirect approach to lecturing.

In fact, the process of a case method class is the real subject of the case method. In the class, a leader facilitates a discussion among a group of well-prepared people who have studied pertinent data and concepts in order to make a decision with consensus support that flows from the intelligent analysis of a problem facing a manager. It is this use of the scientific method by a group that can focus because of the availability of well-prepared case leadership and management that the case method is all about. So, what we teach is how we teach.

See Also

► Organizational Learning

References

Copeland, M.T. 1957. *And mark an era: The story of the Harvard business school*. Boston: Little Brown.

Causal Ambiguity

Richard Reed[1] and Robert J. Defillippi[2]
[1]Cleveland State University, Cleveland, OH, USA
[2]Suffolk University, Boston, MA, USA

Abstract

Inter-firm causal ambiguity arises from the tacitness within skills and routines underlying firm competencies, complexity among combinations of competencies and the specificity of assets within competencies (Reed and DeFillippi Acad Manag Rev 15: 88–102, 1990). It can thus help sustain a competitive advantage because it is difficult for competitors to understand the relationship between a firm's inputs and its outputs. More recently, causal ambiguity has also been viewed as an intra-firm phenomenon that can limit the extension of a competitive advantage because of the difficulty of replication of competencies to other parts of the firm.

Definition Causal ambiguity is a limiting factor to imitability. It is embedded in the relationship between a firm's business inputs and outputs, it helps protect any competitive advantage that the firm may have from imitation, and it thus helps protect superior firm performance.

Causal ambiguity was identified by Lippman and Rumelt (1982) as an imitation-limiting condition that helped explain performance differences among firms. They described it as the 'basic ambiguity concerning the nature of the causal connections between actions and results [among factors of production]' (p. 418). Their model recognizes not only the existence of economic rents that attract imitation and new entrants into the industry, but also how competition lowers profits. Their arguments included both intraand inter-firm ambiguity, but they did not identify their sources.

The concept of causal ambiguity has been incorporated into the emerging ► resource-based view by Reed and DeFillippi (1990). Reed and DeFillippi held intra-firm ambiguity constant at or near zero, and explained inter-firm causal ambiguity in terms of the tacitness arising from the skills and routines embedded in competencies, complexity among combinations of competencies and the specificity of assets within competencies. They described how, either individually or in combination, those characteristics can create causal ambiguity in the relationship between inputs (competencies) and outputs (firm performance). Assuming a competency-based ► competitive advantage, that ambiguity becomes a mechanism for sustaining the advantage and consequent superior performance. It is important to realize that both Lippman and Rumelt (1982) and Reed and DeFillippi (1990) saw causal ambiguity flowing from a lack of understanding by observers of the relationship between business inputs and outputs. Neither pair of authors saw it as a source of firm performance. Rather, it is a mechanism that helps a firm maintain an existing superior performance differential by making imitation difficult. Some authors, either explicitly or implicitly, have portrayed causal ambiguity as an actual source of firm performance rather than a mechanism protecting superior performance. For example, Powell et al. (2006: 176–181) stated that 'studies have not affirmed the empirical linkage between ambiguity and performance … we explain why the ambiguity performance connection is weaker than previously supposed'.

Since the seminal works of Lippman and Rumelt (1982) and Reed and DeFillippi (1990), numerous variations of the definitions have appeared. For example, in his widely cited work that helped establish the resource-based view (RBV), Barney (1991: 109) noted that the idea of causal ambiguity had been around for some

time and he explained it as the links between inputs and outputs being 'understood only very imperfectly'. There also have been several attempts at construct refinement. For example, as discussed in more detail below, Mosakowski (1997) provided a hierarchy of types, King and Zeithaml (2001) separated the construct into characteristic ambiguity (competency tacitness and complexity) and linkage ambiguity (the links between competencies and competitive advantage), and Ryall (2009) separated it into intrinsic and subjective ambiguity within the context of knowledge that involves both causality and ambiguity. All of that said, there is a consistency at the core of these developments that causal ambiguity makes imitation difficult. For example, McEvily and Chakravarthy (2002) examined the effects of causal ambiguity on innovation and found that complexity and tacitness in technological knowledge helped protect major product improvements from competitor imitation, but not minor improvements; technological specificity (design specificity for specific customers) helped protect minor improvements. McEvily et al. (2000) examined how causal ambiguity affects the duration of competitive advantage and determined that when it is wholly effective – Barney's (1991) 'strong form' of sustainability of advantage – then the firm also needs to build barriers to resource substitution to protect superior performance.

In reviewing the literature on causal ambiguity, it rapidly becomes clear that most of the earlier work on causal ambiguity was at the macro level, and linked it with competency-based advantage, imitation and sustained firm performance. However, as work has progressed it has also been embraced at the micro level, with attention being devoted to decision makers, knowledge transfer and replication of actions within the firm. That split logically results in research that has addressed the inter- and intra-firm effects of causal ambiguity. Following her earlier work on causal ambiguity (King and Zeithaml 2001), King (2007) took upon herself the task of more clearly separating inter- and intra-firm causal ambiguity and addressing the effects of both on sustainable competitive advantage. She provided an extensive summary of the literature, noting that it has been conceptualized both as a strategic and cognitive construct. She argued that where inter-firm causal ambiguity would help sustain a competitive advantage from imitation, intra-firm ambiguity would make reinvestment in sources of ambiguity difficult and would thus limit extension of any competitive advantage because of the difficulty of replication of competencies to other parts of the firm.

At the level of intra-organizational imitation, Szulanski (1996) examined 'internal stickiness' in the transfer of best practices within the firm. Among other factors, such as absorptive capacity, he addressed how causal ambiguity impeded the transfer of that knowledge. Mosakowski's (1997: 416) hierarchy of types of causal ambiguity was concerned with how it affects the decision maker. She defined ambiguity in terms of performance distributions generated by a 'unique, stochastic causal structure and a unique specification of inputs', and then went on to model it as effects that arise from future (external) uncertainty, uncertainty from a lack of knowledge of the decisions made by others in the organization, uncertainty about the value of inputs and uncertainty about future causal structures. Simonin (1999) delved into the transferability of knowledge – processes and technology – in alliances. He found that knowledge ambiguity mediated tacitness, complexity, partner prior-experience, cultural distance and organizational distance on knowledge transfer, which, in turn, were moderated by collaborative know-how, learning capacity and alliance duration. King and Zeithaml (2001) addressed the 'causal ambiguity paradox' – if rivals cannot understand the input–output relationship, how can managers in the focal firm understand it – which not only acts as a block on imitation but also as a block to factor mobility (Lippman and Rumelt 1982). As mentioned earlier, King and Zeithaml examined causal ambiguity according to characteristic and linkage ambiguity, within the way that senior and middle managers viewed organizational competencies. What they found is that linkage ambiguity partially mediates the relationship between characteristic ambiguity and sustained performance. Blyler and Coff (2003) explored how causal ambiguity affects rent appropriation and deduced that it

creates a power vacuum into which employees can insert themselves and appropriate firm rents. Szulanski et al. (2004) examined the question of how the intra-organizational transfer of knowledge associated with superior performance is affected by causal ambiguity. They found that 'causal ambiguity drives a wedge between perception and the reality of organizational practices', and that trust-worthiness of the source providing the information on the relationship between inputs and outputs first weakens and then ends up negative as causal ambiguity increases (Szulanski et al. 2004: 608). Inkpen (2008) extended thinking on causal ambiguity to include the context in which knowledge is created. Using the GM and Toyota joint venture as a case example, he argued that communication and social interaction are necessary for overcoming contextual barriers to knowledge transfer between firms. Finally, Coff and Kryscynski (2011) explored human capital-based competitive advantage and, while their focus was clearly within the firm, they did note that when causal ambiguity is derived from tacit knowledge it creates problems of imitation for both people within the firm and for competitors.

Over the 30-plus years since Lippman and Rumelt (1982) brought causal ambiguity to our attention, it has remained firmly entrenched in the domain of strategy. As Lockett (2001) noted, although some RBV concepts from strategy have made the transition to economics, causal ambiguity is not one of them, perhaps because of the difficulties associated with measuring it. It is normal for a construct to morph over time into something quite different from its original conceptualization as scholars improve upon the original thinking, or as the needs of their particular research questions drive a need for modification. That has not happened with causal ambiguity, perhaps because it has been embedded within strategic management research as a cornerstone of the RBV. The only degradation of the construct has been an occasional attempt to portray it as a source of firm performance rather than a protector of performance differentials. The central thesis of it being a limiting factor to imitability has remained intact, along with tacitness, complexity and specificity in competencies.

See Also

- ▶ Competitive Advantage
- ▶ Firm Resources
- ▶ Resource-Based View

Acknowledgements We wish to acknowledge Susan Storrud-Barnes, Cleveland State University, for her comments and suggestions. We have examined causal ambiguity through the lens of influential research (i.e., work that is widely cited), so we also wish to acknowledge Vivien Jancenelle, graduate assistant, for his help in identifying those articles.

References

Barney, J.B. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17: 99–120.

Blyler, M., and R.W. Coff. 2003. Dynamic capabilities, social capital, and rent appropriation: Ties that split pies. *Strategic Management Journal* 24: 677–686.

Coff, R., and D. Kryscynski. 2011. Invited editorial: Drilling for micro-foundations of human capital-based competitive advantages. *Journal of Management* 37: 1429–1443.

Inkpen, A.C. 2008. Knowledge transfer an international joint ventures: The case of NUMMI and General Motors. *Strategic Management Journal* 29: 447–453.

King, A.W. 2007. Disentangling interfirm and intrafirm causal ambiguity: A conceptual model of causal ambiguity and competitive advantage. *Academy of Management Review* 32: 156–178.

King, A.W., and C.P. Zeithaml. 2001. Competencies and firm performance: Examining the causal ambiguity paradox. *Strategic Management Journal* 22: 75–99.

Lippman, S., and R.P. Rumelt. 1982. Uncertain imitability: An analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics* 13: 418–438.

Lockett, A. 2001. The resource-based view and economics. *Journal of Management* 27: 723–754.

McEvily, S.K., and B. Chakravarthy. 2002. The persistence of knowledge-based advantage: An empirical test for product performance and technological knowledge. *Strategic Management Journal* 23: 285–305.

McEvily, S.K., S. Das, and K. McCabe. 2000. Avoiding competence substitution through knowledge sharing. *Academy of Management Review* 25: 294–311.

Mosakowski, E. 1997. Strategy making under causal ambiguity: Conceptual issues and empirical evidence. *Organization Science* 8: 414–442.

Powell, T.C., D. Lovallo, and C. Caringal. 2006. Causal ambiguity, management perception, and firm performance. *Academy of Management Review* 31: 175–196.

Reed, R., and R.J. DeFillippi. 1990. Causal ambiguity, barriers to imitation, and sustainable competitive advantage. *Academy of Management Review* 15: 88–102.

C

Ryall, M.D. 2009. Causal ambiguity, complexity, and capability-based advantage. *Management Science* 55: 389–403.
Simonin, B.L. 1999. Ambiguity and the process of knowledge transfer in strategic alliances. *Strategic Management Journal* 20: 595–624.
Szulanski, G. 1996. Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal* 17: 27–43.
Szulanski, G., R. Cappetta, and R.J. Jensen. 2004. When and how trustworthiness matters: Knowledge transfer and the moderating effect of causal ambiguity. *Organization Science* 15: 600–613.

CEO Compensation

Cynthia E. Devers[1] and W. Gerard Sanders[2]
[1]Texas A&M University, College Station, TX, USA
[2]Rice University, Houston, TX, USA

Abstract

Theory suggests that agency conflicts can be partially mitigated through effective and well-designed CEO compensation. In the modern era, equity-based compensation has been the primary means used by firms to achieve incentive alignment. However, findings from empirical research are somewhat at odds with agency theory-inspired conventional wisdom and suggest the need to focus future research on areas that promise to advance our understanding of how executives perceive and respond to compensation: pay and pay disparities within and among levels of analysis, finer-grained distinctions among individual pay elements, and causes of unintended consequences of ▶ incentives.

Definition CEO compensation refers to the way in which firms pay the chief executive officer (and by extension, the top management team); it includes the level, form, dispersion and effects of senior executive compensation.

The assumption that agency problems arise when ownership and managerial control are separated (e.g., Berle and Means 1932) is deeply rooted in executive compensation theory and practice. Given that monitoring executives' behaviours is difficult and costly, compensation scholars often advocate mitigating agency problems via incentive alignment. Incentive alignment involves including incentive-based pay forms in executives' compensation packages to tie their compensation to firm performance or other outcomes important to owners (Jensen and Meckling 1976). Agency theory suggests that incentive-alignment practices motivate CEOs to increase shareholder wealth (Eisenhardt 1989) and may be an effective and efficient means of circumventing the difficulties and costs of directly supervising managerial behaviour (Finkelstein et al. 2009).

The relationship between firm performance and CEO pay has been extensively researched. Early efforts focused on the premise that the covariance of firm performance and executive pay would be an indication of incentive alignment. Owing to small observed correlations, some scholars have questioned whether tying pay to performance is an effective incentive alignment mechanism (Tosi et al. 2000; Dalton et al. 2007). With respect to the use of incentive-based forms of pay, research has also raised serious questions about their efficacy (see Devers et al. 2007). This work also underscores the importance of expanding conventional executive compensation research to important but lesser-developed areas. In this article, we focus on three emerging areas that we believe offer much promise to advance our understanding of how executives perceive and respond to compensation: pay and pay disparities within and among levels of analysis, finer-grained distinctions among individual pay elements, and the unintended consequences of CEO pay.

Pay and Pay Disparities Within and Among Levels of Analysis

The majority of executive compensation research has focused heavily on CEO pay. Recent research suggests that expanding this scope to examine the pay of the CEO's ▶ Top Management Teams

(TMT) may enhance our knowledge of how compensation mitigates agency problems and how senior executives individually and collectively perceive and respond to their pay. For instance, Carpenter and Sanders (2002) argued that TMT incentive alignment may be just as important as CEO incentive alignment for reducing agency problems. In support, they found that TMT incentive alignment fully mediated the positive effects of CEO incentive alignment on firm performance. Firms that aligned the ▶ incentives of the CEO's top team achieved higher performance levels, while those that aligned only CEOs' incentives did not. Similarly, in a multinational corporation setting, Carpenter and Sanders (2004) also found that, although CEO pay alone did not significantly affect firm performance, both TMT pay and TMT long-term incentive pay positively influenced subsequent firm performance.

Other research suggests that considering pay disparities among hierarchical levels may also be relevant to executive compensation research. For example, the upper echelons (UE) perspective emphasizes that top managers often make decisions interdependently. Although some evidence suggests that pay disparities can increase executives' motivation and performance, via competition (e.g., Lazear and Rosen 1981; Main et al. 1993), UE scholars have argued that pay disparities can negatively influence team functioning and individuals' contributions to firm goals (Henderson and Fredrickson 2001). Recent work indicates that pay disparities are often perceived, 'as unfair and disruptive of social relationships' (Pfeffer 2007: 121). Other scholars have found that pay disparities can negatively correlate with individual and team productivity and performance (e.g., Bloom 1999; Carpenter and Sanders 2004; Siegel and Hambrick 2005; Wade et al. 2006), particularly in contexts that require interdependence (Henderson and Fredrickson 2001; Shaw et al. 2002; Siegel and Hambrick 2005; Fredrickson et al. 2010).

This research shows that pay disparities can influence individuals' behaviours in material ways. Thus, we suggest that scholars continue broadening their focus from the CEO to the TMT and beyond, to develop a deeper and more nuanced understanding of how inter- and intra-level pay differences influence executives' perceptions and actions, and firm performance.

Elements of Pay

Compensation researchers have generally aggregated incentives into a single pay measure. However, recently, scholars have begun focusing on the distinct pay elements, revealing intriguing patterns that were heretofore not generally contemplated. Most interesting is evidence showing that distinct forms of incentive-based pay appear to influence executive behaviour in asymmetric ways. For example, Sanders (2001) demonstrated that CEO stock ownership decreased acquisitions and divestitures while CEO stock options motivated repeated acquisitions and divestitures.

Extending Sanders' (2001) work, Certo and colleagues (2003) found that initial public offering (IPO) investors responded more favourably to stock options (increased IPO valuations) when IPO executives also held firm stock. Devers and colleagues (2008) dissected the compensation package in an even finer-grained way. Like Sanders (2001), they found that the value of CEOs' stock reduced firm ▶ risk taking and that CEOs' *unexercisable* option spread values (i.e., the difference between current share price and option exercise price) positively influenced firm risk-taking. However, their results further showed that CEOs who held *exercisable* options with high spread values and/or restricted stock reduced their risk-taking, presumably to mitigate downside compensation risk. Similarly, using survey data from IPO firm CEOs, Larraza-Kintana and colleagues (2007) found that the value of those CEOs' in-the-money stock options was negatively associated with firm risk-taking. On the other hand, the variability in CEOs' cash-based pay positively influenced risk-taking.

Although work in this area is fairly new, we argue that continuing this line of research is particularly germane to advancing our knowledge of how the various elements of compensation (individually and collectively) influence executive action and firm performance.

Unintended Consequences

The incentive alignment argument implicitly assumes that incentives mitigate executive opportunism and ▶ risk aversion and, by doing so, motivate actions designed to enhance long-term firm value (Devers et al. 2007). Nevertheless, recent research suggests that executives often respond to incentive compensation in ways that benefit themselves at the firm's (shareholders') expense. For example, Sanders and Hambrick (2007) argued that the asymmetric risk properties of stock options would cause increases in risk-taking along three dimensions: the size of risky investments, greater investment in projects with high variance of potential outcomes, and greater investment in projects with likelihood of extreme gains/losses. They reported that CEO stock options engendered high levels of investment outlays and brought about extreme corporate performance (big gains and big losses), suggesting that stock options prompt CEOs to make high-variance bets, not simply larger bets. Perhaps more importantly, in a finding that stands in stark contrast to incentive alignment arguments, they found that option-loaded CEOs delivered more big losses than big gains, suggesting that high levels of stock option pay leads to excessive focus on the possible large gains associated with risky investments without corresponding consideration of downside risk.

Scholars and practitioners have also argued for awarding incentives to motivate executives to focus on long-term, as opposed to short-term performance (e.g., Jensen and Murphy 1990). Nevertheless, Guidry et al. (1999) found that incentive-based pay appeared to encourage executives to exploit their short-term bonuses by increasing short-term firm value, at the expense of long-term performance. Further, Souder and Bromiley (forthcoming) found a negative relationship between stock-based compensation (stock and stock options) and the durability of new capital expenditures. The authors noted: '[d]irectly contrary to arguments that stock-based compensation should induce managers to take a longer view, we find evidence that stock-based compensation induces shorter duration investments' (p. 27).

Other scholars have similarly argued that high incentive levels motivate executives to opportunistically acquire other firms for personal benefits, often irrespective of performance (Grinstein and Hribar 2004). For example, directors often award new incentives and other compensation to top managers following acquisitions, without regard for how the acquisition performed (Harford and Li 2007). Thus, although acquisitions often result in stock price reductions that also decrease acquiring firm executives' incentive values, those decreases are often offset by additional compensation awards owing to firm growth and discretionary bonus pay (Bliss and Rosen 2001).

In a similar vein, scholars have found evidence that incentive pay motivates CEOs to opportunistically schedule awards prior to anticipated stock price increases (Yermack 1997), release information around scheduled awards (Aboody and Kasznik 2000) or partake in stock option backdating schemes (Lie 2005). Still other research appears to show that CEO stock options are positively associated with increased open market share repurchases that increase the value of those options at the expense of dividends (Sanders and Carpenter 2003).

Finally, more recent research suggests a pernicious relationship between financial legerdemain and executive compensation. For example, Bergstresser and Philippon (2006) found that incentives appeared to motivate executives to manipulate earnings and, in turn, to cash in equity-based pay when their firms' values were artificially inflated. Similarly, Burns and Kedia (2006) found that earnings misreporting was strongly influenced by the sensitivity of CEOs' stock option portfolios to firm share price. Of even more concern, Harris and Bromiley (2007) reported evidence that accounting misstatements and fraud were associated with high levels of stock option pay.

In sum, this growing body of research documents some troubling unintended consequences of incentive-based pay that significantly challenge arguments for incentive alignment.

Conclusion

The goal for our review is to expand conventional executive compensation research to three emerging areas that we believe hold potential in order to address important gaps in our understanding of executive compensation. We hope that, by focusing on these three developing areas, we have provided a foundation that scholars can further build on to advance our understanding of how executives perceive and respond to executive compensation.

See Also

- ▶ Incentive Design
- ▶ Incentives
- ▶ Risk Aversion
- ▶ Risk-Taking
- ▶ Top Management Teams

References

Aboody, D., and R. Kasznik. 2000. CEO stock option awards and the timing of corporate voluntary disclosures. *Journal of Accounting and Economics* 29: 73–100.

Bergstresser, D., and T. Philippon. 2006. CEO incentives and earnings management. *Journal of Financial Economics* 80: 511–529.

Berle, A.A., and G.C. Means. 1932. *The modern corporation and private property*. New York: Harcourt, Brace & World.

Bliss, R.T., and R.J. Rosen. 2001. CEO compensation and bank mergers. *Journal of Financial Economics* 61: 107–138.

Bloom, M. 1999. The performance effects of pay dispersion on individuals and organizations. *Academy of Management Journal* 42: 25–40.

Burns, N., and S. Kedia. 2006. The impact of performance-based compensation on misreporting. *Journal of Financial Economics* 79: 35–67.

Carpenter, M.A., and W.G. Sanders. 2002. Top management team compensation: The missing link between CEO pay and firm performance? *Strategic Management Journal* 23: 367–375.

Carpenter, M.A., and W.G. Sanders. 2004. The effects of top management team pay and firm internationalization on MNC performance. *Journal of Management* 30: 509–528.

Certo, S.T., C.M. Daily, A.A. Cannella Jr., and D.R. Dalton. 2003. Giving money to get money: How CEO stock options and CEO equity enhance IPO valuations. *Academy of Management Journal* 46: 643–653.

Dalton, D.R., M.A. Hitt, S.T. Certo, and C.M. Dalton. 2007. The fundamental agency problem and its mitigation. In *Academy of management annals*, ed. J.F. Walsh and A.P. Brief. New York: Erlbaum.

Devers, C.E., A.A. Cannella, G.P. Reilly, and M.E. Yoder. 2007. Executive compensation: A multidisciplinary review of recent developments. *Journal of Management* 33: 1016–1072.

Devers, C.E., G. McNamara, R.M. Wiseman, and M. Arrfelt. 2008. Moving closer to the action: Examining compensation design effects on firm risk. *Organization Science* 19: 548–566.

Eisenhardt, K.M. 1989. Agency theory: An assessment and review. *Academy of Management Review* 14: 57–74.

Finkelstein, S., D.C. Hambrick, and A.A. Cannella Jr. 2009. *Strategic leadership: Theory and research on executives, top management teams, and boards*. Oxford: Oxford University Press.

Fredrickson, J.W., A. Davis-Blake, and W.G. Sanders. 2010. Sharing the wealth: Social comparisons and pay dispersion in the CEO's top team. *Strategic Management Journal* 31: 1031–1053.

Grinstein, Y., and P. Hribar. 2004. CEO compensation and incentives: Evidence from M&A bonuses. *Journal of Financial Economics* 73: 119–143.

Guidry, F., A.J. Leone, and S. Rock. 1999. Earnings-based bonus plans and earnings management by business-unit managers. *Journal of Accounting and Economics* 26: 113–142.

Harford, J., and K. Li. 2007. Decoupling CEO wealth and firm performance: The case of acquiring CEOs. *Journal of Finance* 62: 917–949.

Harris, J., and P. Bromiley. 2007. Incentives to cheat: The influence of executive compensation and performance on financial misrepresentation. *Organization Science* 18: 350–367.

Henderson, A.D., and J.W. Fredrickson. 2001. Top management team coordination needs and the CEO pay gap: A competitive test of economic and behavioral views. *Academy of Management Journal* 44: 96–117.

Jensen, M., and W. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics* 3: 305–360.

Jensen, M.C., and K.J. Murphy. 1990. Performance pay and top-management incentives. *Journal of Political Economy* 98: 225–264.

Larraza-Kintana, M., R.M. Wiseman, L.R. Gomez-Mejia, and T.M. Welbourne. 2007. Disentangling compensation and employment risks using the behavioral agency model. *Strategic Management Journal* 28: 1001–1019.

Lazear, E.P., and S. Rosen. 1981. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy* 89: 841–864.

Lie, E. 2005. On the timing of CEO stock option awards. *Management Science* 51: 802–812.

Main, B., C.A. O'Reilly, and J. Wade. 1993. Top executive pay: Tournament or teamwork? *Journal of Labor Economics* 11: 607–628.

Pfeffer, J. 2007. Human resources from an organizational behavior perspective: Some paradoxes explained. *Journal of Economic Perspectives* 21: 115–134.
Sanders, W.G. 2001. Behavioral responses of CEOs to stock ownership and stock option pay. *Academy of Management Journal* 44: 477–492.
Sanders, W.G., and M.A. Carpenter. 2003. Strategic satisficing? A behavioral-agency theory perspective on stock repurchase program announcements. *Academy of Management Journal* 46: 160–178.
Sanders, W.G., and D.C. Hambrick. 2007. Swinging for the fences: The effects of CEO stock options on company risk taking and performance. *Academy of Management Journal* 50: 1055–1078.
Shaw, J.D., N. Gupta, and J.E. Delery. 2002. Pay dispersion and workforce performance: Moderating effects of incentives and interdependence. *Strategic Management Journal* 23: 491.
Siegel, P.A., and D.C. Hambrick. 2005. Pay disparities within top management groups: Evidence of harmful effects on performance of high-technology firms. *Organization Science* 16: 259–274.
Souder, D., and P. Bromiley. Forthcoming. Explaining temporal orientation: Evidence from the durability of firms' capital investments. *Strategic Management Journal*.
Tosi, H.L., S. Werner, J.P. Katz, and L.R. Gomez-Mejia. 2000. How much does performance matter? A metaanalysis of CEO pay studies. *Journal of Management* 26: 301–339.
Wade, J.B., C.A. O'Reilly, and T.G. Pollock. 2006. Overpaid CEOs and underpaid managers: Fairness and executive compensation. *Organization Science* 17: 527–544.
Yermack, D. 1997. Good timing: CEO stock option awards and company news announcements. *Journal of Finance* 52: 449–476.

Certainty Equivalence

Daniel E. Ingberman
University of California Berkeley, Walter A. Hass School of Business, Berkeley, CA, USA

Abstract

Strategies in dynamic games can be derived as the solutions of stochastic dynamic optimization problems. General solutions to such problems can be elusive, even with modern techniques. Simon (1955, 1957) showed that an important class of 'linear-quadratic' problems obey on 'certainty equivalence' – such stochastic dynamic optimization problems can be solved as if there is no uncertainty, by substituting expected values for all uncertain state variables. This insight became the basis for Simon's 'bounded rationality' as well as rational expectations in economics.

Definition A stochastic dynamic optimization problem displays 'certainty equivalence' when the problem's solution can be found by solving a similar non-stochastic problem in which the uncertain state variable at every point in time is assumed to be known and equal to its expected value.

Most decisions are made under uncertainty and have dynamic consequences. For example, the decision to buy and maintain a house rests, at least in part, on expectations as to the future course of housing prices, and is subject to wealth constraints that may depend on past decisions as well as exogenous conditions. The purchase decision may also rest on expectations of how home ownership will affect future incentives in alternative possible future situations. If housing prices fall significantly below the purchase price, there may be an incentive to renege on a mortgage commitment, which may entail a set of additional consequences (such as denial of credit or bankruptcy). Because these possibilities can interact with the incentives to buy, sell and maintain a property, uncertainty about the future can fundamentally change housing decisions and strategies.

Such decision problems can be analysed using *dynamic programming* (otherwise known as s*tochastic optimal control*), a form of mathematical optimization theory. However, even in non-strategic environments under conditions of certainty, dynamic programming can be analytically demanding, and explicit closed-form solutions are known for only certain limited classes of problems (Bertsekas 1976; Hansen and Sargent 2007). While other problems may be analysed with numerical methods, these solutions can be difficult to generalize. If one additionally postulates that the decision maker is uncertain about the future environment that will determine the outcomes of their choices, optimal dynamic decisions

may depend, crucially, on the nature of that uncertainty, and, consequently, characterizing such strategies can become quite complex.

Around the same time that he was developing his path-breaking insights regarding bounded rationality (Simon 1955, 1957), Herbert Simon demonstrated that a particular class of dynamic programming problems obeys *certainty equivalence*, which enables the dynamic component to be separated from the uncertainties. Simon's method was extended to the multivariate case by Theil (1957). An influential application of certainty equivalent programming methods to production planning can be found in the work of Holt and colleagues (1960). More recently, it has been shown that certainty equivalence properties emerge even when there is uncertainty about model specification (see, e.g., Hansen and Sargent 2007: 21–37). Specifically, in a dynamic programming problem, at every moment in time the decision maker selects a value of the *control variable* to maximize the discounted present value of the utility (or other objective function), subject to the evolution of the *state variables*, which describe the underlying economic environment as a function of past control variables as well as exogenous variables and random shocks (normally assumed to be the source of the uncertainty in the problem). Simon showed that if the single-period utility is a quadratic function of the state and the control, and the current state evolves as a linear function of the past states and the random shock's value, then the optimal control law is found by assuming that future state variables will take their expected values with certainty. In other words, one can solve such a 'linear-quadratic' problem as if there is no uncertainty, using the expected values of all uncertain variables in place of probability density functions, without worrying about the possible interactions between the uncertain environment and optimal dynamic decisions. The existence of risk or uncertainty reduces the maximized value of the dynamic utility function, but does not change the optimal choices of the control variables.

Certainty equivalence therefore separates dynamic ▶ decision-making problems into two separate parts – 'optimization' and 'forecasting' – which highlights the formation of expectations and beliefs, and how they are translated into behaviour.

This insight has spawned two related and very substantial but disparate literatures. Because complicated optimization problems can be approximated by related problems for which certainty equivalence holds – what Simon characterized in his 1978 Nobel address as 'an approximating, satisficing simplification' – then, as Simon suggests, one might plausibly imagine a bounded-rational decision maker acting as if they are solving a dynamic optimization problem that obeys certainty equivalence (Simon 1992: 359). He states:

> Hence the assumption of quadratic costs reduces the original problem to one that is readily solved. Of course the solution, though it provides optimal decisions for the simplified world of our assumptions, provides, at best, satisfactory solutions for the real-world decision problem that the quadratic function approximates. In principle, unattainable optimization is sacrificed for in practice, attainable satisfaction. If human decision makers are as rational as their limited computational capabilities and their incomplete information permit them to be, then there will be a close relation between normative and descriptive decision theory. (Simon 1992: 351)

But as a descriptive matter or as the basis for normative or policy analysis, indeterminacy still exists: how are expectations formed? For this reason, Simon argued that uncertainty requires an 'elaboration of the model of the decision process', which would 'incorporate the notions of bounded rationality: the need to search for decision alternatives, the replacement of optimization by targets and satisficing goals, and mechanisms of learning and adaptation' (Simon 1992: 360, 366). Alternatively, one can follow the logic of certainty equivalence to the ▶ rational expectations hypothesis pioneered by Simon's colleague and co-author at Carnegie Mellon, John Muth, and simply assume that 'expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory' (Muth 1961: 318). Certainty equivalence is widely acknowledged to be crucial to the development of the theory of rational expectations (see, e.g., Lucas and Sargent 1981: xiv–xvi).

See Also

- ▶ Decision-Making
- ▶ Rational Expectations

References

Bertsekas, D.P. 1976. *Dynamic programming and stochastic optimal control*. New York: Academic.
Hansen, L.P., and T.J. Sargent. 2007. *Robustness*. Princeton: Princeton University Press.
Holt, C.C., F. Modigliani, J.F. Muth, and H.A. Simon. 1960. *Planning production, inventories and work force*. Englewood Cliffs: Prentice Hall.
Lucas Jr., R.E., and T.J. Sargent. 1981. *Rational expectations and econometric practice*. Minneapolis: University of Minnesota Press.
Muth, J.F. 1961. Rational expectations and the theory of price movements. *Econometrica* 29: 315–353.
Simon, H.A. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* 69: 99–118.
Simon, H.A. 1957. *Models of man*. New York: Wiley.
Simon, H.A. 1992. Rational decision-making in business organizations: Nobel memorial lecture, 8 December 1987. In *Nobel lectures, economic sciences, 1969–1980*, ed. E. Lindbeck. Singapore: World Scientific Publishing.
Theil, H. 1957. A note on certainty equivalence and dynamic planning. *Econometrica* 25: 346–349.

Chaos Theory

Bernard Forgues and Raymond-Alain Thietart
ESSEC Business School, Cergy-Pontoise, Paris, France

Abstract

Firms can be described as non-linear dynamic systems in that they are home to counteracting forces simultaneously at play. As a consequence, firms or some of their processes can enter a chaotic state. In such a state prediction is impossible, especially at a global scale and in the long run. Closure is still possible as firms may be attracted towards specific configurations. Finally, because contextual conditions are never the same, no similar actions can yield the same result.

Definition Chaos theory originates from the mathematical analysis of non-linear dynamic systems. Applied to management, it allows the tackling of issues related to complex interactions and unpredictability.

A Quick Overview of Chaos Theory

Chaos theory, or deterministic chaos, may be traced back to mathematician Henri Poincaré, working at the end of the nineteenth century, and more recently to meteorologist Edward Lorenz. They observe that low dimensional systems' behaviour – systems which can be represented by a limited number of variables, for example less than six – starting from nearly identical states could diverge dramatically. The famous metaphor of 'butterfly effect', according to which 'the flap of a butterfly's wings in Brazil [can] set off a tornado in Texas' (Lorenz 1972), illustrates this phenomenon of divergence: a sensitivity dependence on initial conditions. A major consequence of this sensitivity is unpredictability, especially in the long run.

In the mid-1970s, Mitchell Feigenbaum observed bifurcation processes taking place in non-linear dynamic systems. He showed how apparent random behaviour can originate from very simple models. Under certain conditions and as a function of its parameters' values, the behaviour of a model evolves from stability to periodicity to apparent randomness. A system bifurcates and jumps from one state to the other as a reaction to small variations in its guiding rules.

Working from a different perspective, Benoît Mandelbrot studied the distribution of stock price changes. He observed recurring patterns at several scales: years, months, days, hours. He then searched for such scale invariance in a variety of apparently random data, such as Britain's coastline. He found similar results. All these systems displayed self-similarity (i.e., they had the same patterns over different scales); they are said to be fractals. Mandelbrot found many more examples of scale invariance in nature, such as snowflakes or Romanesco broccoli. Under some conditions, complex systems, natural or artificial, reveal the characteristic of scale invariance.

These complementary works opened up the possibility to go beyond apparent randomness in low dimensional dynamic systems and spurred interest in a wide range of research domains, which have since embraced the broader field of complexity science (Thietart and Forgues 2011).

Firms as Non-Linear Dynamic Systems

Firms are non-linear dynamic systems, in that many of their elements are linked by non-linear relationships and are interconnected with one another (Thietart and Forgues 1995). For instance, having resource slack allows adjusting for increases in demand. This can foster performance up to a certain point. Having too much slack comes at a cost, which lowers performance. Slack and performance are linked with a non-linear (inverted U-shaped) relationship (George 2005). Similarly, Semadeni and Cannella (2011) have shown that post-spin-off child firms benefit from having some links to the parent, but having too many links is negatively related to performance. All in all, non-linear relationships are more the norm than the exception in business. Further, elements within a firm are interconnected in countless ways. Revisiting empirical research on CEO performance, Blettner et al. (2012) conclude that complex interdependency is the driver of the CEO performance effect. In an overwhelming (yet simplified) 'partial CEO map', they show a web of relationships between variables having a direct or indirect impact on firm performances (Blettner et al. 2012: 994).

These recent empirical studies echo observations made long ago. Thompson (1967: 6) already noted that 'the complex organization is a set of interdependent parts which together make up a whole in that each contributes something and receives something from the whole, which in turn is interdependent with some larger environment'. Strategy-making has also long been seen as a combination of systematic approaches, carefully planned and implemented on the one hand, and of emerging responses, based on opportunities, chance, shifting coalitions and the like (Mintzberg and Waters 1985).

This does not mean that firms are like natural systems wherein laws are immutable. On the contrary, the very structure of the organizational system changes as managers act, people learn, the external environment evolves. But all those forces add instability and complexity. Further, their dynamics are impossible to assess with precision and comprehensiveness, let alone predict. The fact that a firm can be described as a non-linear dynamic system doesn't imply it is in a chaotic state. More precisely, deterministic chaos can happen in the presence of forces simultaneously acting in opposite directions. This often happens in firms, where some forces push towards stability and order whereas others push towards instability and disorder. Firms usually have a formal structure, monitor through control systems and engage in planning. Yet at the same time they launch innovations, welcome initiatives and leave room for experimentation. The simultaneous coupling of these counteracting forces can yield to chaos, as we have shown elsewhere (Thietart and Forgues 1997).

As a consequence, although chaos theory comes from the hard sciences, it can be usefully translated to inform strategic management. In particular, one can derive a number of qualitative properties of great importance to both theory and practice. Among these qualitative properties we focus below on bifurcation processes, sensitivity to initial conditions, strange attractors, scale invariance and time irreversibility.

Consequences for Strategic Management

Bifurcation Processes and Stepwise Change

Systems can experience abrupt changes even if changes in parameter values are only small and happen smoothly. Such an evolution is called a bifurcation, and it has been widely documented in firms. For instance, firms can undergo step-by-step changes or proceed through quantum jumps, for instance when facing a crisis. This has positive and negative consequences for strategy. On the one hand, whatever the state a firm is in, it can be disrupted. Firms can always evade competence traps or stuck-in-the-middle positions. On the

other hand, bifurcation processes contribute to making change highly unpredictable. Because even small and smooth changes can yield dramatic consequences, strategy implementation remains subject to bifurcations that can push the firm in unexpected situations. To address this issue, firms that compete on the edge (Brown and Eisenhardt 1998) implement a variety of small-scale and big-scale changes. Rather than shifting the firm abruptly towards new configurations in rare occurrences, they change through a number of small steps (like exploitation and benchmarking) and occasional bigger ones (like acquisitions). In such cases, managers implement a sort of trial and error approach with multiple options. Strategy is no longer one fixed plan to be implemented, with goals to be reached before formulation of a new strategy. It is rather continuously revised and evolves over time.

Sensitivity to Initial Conditions and the Impossibility to Predict

Sensitivity to initial conditions (popularized as Lorenz's 'butterfly effect') means that a small change in one variable has an unpredictably large impact on the system's evolution. Put differently, two situations that look similar even upon close examination can have two dramatically different trajectories. The speed at which such divergence occurs is measured with Lyapunov exponents. Examples in business situations abound. Fifty years ago, Stinchcombe (1965) suggested that firms are durably imprinted by the conditions they met upon founding. This intuition has been confirmed repeatedly in observations on organizational growth, survival rates and performance. However, another consequence of sensitivity to initial conditions is more problematic for managers. Since even one small change in one variable can yield dramatic consequences, 'forecasting is impossible, especially at a global scale and in the long term' (Thietart and Forgues 1995: 26). When formulating strategies, managers have to bet on the future or at least – and perhaps less worryingly – to draw possible scenarios. They can also proceed incrementally, in a step-by-step manner. This allows the observation of consequences of actions and analysis of the newly created situation that the firm is in. Unfortunately, if managing through a series of incremental, short-term changes improves predictability, it risks trading strategy for day-to-day operations. Another consequence of sensitivity to initial conditions has to do with benchmarking and me-too strategies. Many managers feel the urge to replicate the strategies that brought success to competitors. However, initial conditions are always different: those that prevail at a competitor's do not hold at the focal firm's. As a consequence, one cannot expect to reach the same results. Again, this has been confirmed, for instance by empirical observations of failing me-too strategies (Kim and Mauborgne 2004).

Strange Attractors and the Attraction Towards Configurations

Non-linear dynamic systems can be represented in a phase space. Over time, one variable evolves towards a region in this space that is called an attractor. Interestingly, the dynamic system behaviour remains on the attractor's limits. One famous attractor is Lorenz's 'strange attractor', which, in a three-dimension space, is shaped like an eight. Each iteration of the dynamic system will result in a point on this eight-shaped attractor. Note, however, that it is impossible to predict where on the attractor this point will be. As a consequence, although predictability is limited, the set of possibilities is nonetheless restricted. A metaphor proposed by Thietart and Forgues (1995: 26) to illustrate this is that of 'an island of stability [emerging] in a sea of chaos'. The time it takes for an attractor to evolve towards a strange attractor depends on how open it is to its environment, a condition physicists call dissipation. By analogy, firms exchanging resources with their environment tend to move towards organizational configurations. Given the number of dimensions needed to describe firms, a huge number of combinations are theoretically possible. However, only a small number of patterns are to be found. These viable configurations are the attractors towards which businesses evolve. As Mintzberg (1979: 300) puts it, '[t]here is order in the world (. . .) a sense of union or harmony that grows out of the natural clustering of elements, whether they be stars, ants,

or the characteristics of organizations'. Similarly, although the number of actions managers can engage in is infinite, Porter (1980) has convincingly argued that there are only three possible strategies at the business level. Those 'generic' strategies – differentiation, cost leadership and market segmentation – result from two dimensions. One is the decision to focus on a market segment or to aim at the entire market. The other is based on whether the firm's competitive advantage relies on benefiting from a perceived uniqueness in its product or from having the lowest cost of operation.

Fractal Forms and Invariance at Different Organizational Scales

Fractal forms, popularized by their beautiful representations, are self-similar patterns, meaning that they look the same at different scales. The property they exhibit is known as scale invariance. Similar patterns are thus expected to occur at different levels. In the business realm, levels include industries, strategic groups, firms, strategic business units, subunits, groups, teams and individuals. A number of business phenomena have been shown to exhibit similar patterns at different levels. This is the case, for instance, in learning, which allows the basing of predictions for organizational learning on results from individual learning. More generally, a variety of major topics are analysed at the inter-organizational, organizational and intra-organizational levels, as exemplified in Baum's (2002) handbook. In strategic management, scale invariance can be found in the firm's strategic structuring. For instance, companies are organized into smaller strategic units, thus allowing managers to have a very high business focus and precise performance metrics at a higher level. Some companies have also been successful by replicating at a global scale what they had been doing at a smaller, local, scale. Counter-examples abound, though, and the question of scale invariance remains largely undecided at this stage.

Time Irreversibility and the Non-Replicability of Past Situations

Time irreversibility can be thought of as bringing another dimension to sensitivity to initial conditions, in that the reasoning and consequences are of the same kind. In other words, it is the temporal non-replicability equivalent to the spatial non-predictability seen above. Theoretically, a non-linear dynamic system can go back to its initial state, but the probability of this happening is so small that one can reasonably assume that chaotic systems never lead twice to the same situation. This has huge managerial consequences, since it implies that replicating an action will never lead to the same result. Although managers could feel the urge to replicate the strategies that brought them success in the past, they cannot expect these to bring the same level of performance. Quite the contrary: Miller (1992) has shown that companies applying the very same recipes that once produced excellent results eventually met failure, a phenomenon he dubbed the 'Icarus paradox'. More precisely, he observed that growth-driven entrepreneurial builders turn into greedy imperialists by overtaxing their resources; quality-driven craftsmen turn into irrelevant, detail-obsessed tinkerers; flexible and innovative pioneers turn into utopian escapists squandering resources; and successful salesmen organizations into aimless, bureaucratic drifters with a disjointed line of me-too offerings. As observed by Brown and Eisenhardt (1998), whereas traditional strategy aims at building a sustainable competitive advantage, a 'competing on the edge' strategy is highly temporary and unpredictable. Today's successful strategy could turn into tomorrow's recipe for disaster.

See Also

- ▶ Path Dependence in Technologies and Organizations
- ▶ Simulation Modelling and Business Strategy Research

References

Baum, J.A.C. (ed.). 2002. *The Blackwell companion to organizations*. Oxford: Blackwell.

Blettner, D.P., F.R. Chaddad, and R.A. Bettis. 2012. The CEO performance effect: Statistical issues and a complex fit perspective. *Strategic Management Journal* 33: 986–999.

Brown, S.L., and K.M. Eisenhardt. 1998. *Competing on the edge: Strategy as structured chaos*. Boston: Harvard Business School Press.
George, G. 2005. Slack resources and the performance of privately held firms. *Academy of Management Journal* 48: 661–676.
Kim, W.C., and R. Mauborgne. 2004. Blue ocean strategy. *Harvard Business Review* 82: 76–84.
Lorenz, E.N. 1972. Predictability: Does the flap of a butterfly's wings in Brazil set off a tornado in Texas? 139th meeting of the American Association for the Advancement of Science, 29 Dec.
Miller, D. 1992. *The icarus paradox: How exceptional companies bring about their own downfall*. New York: HarperCollins.
Mintzberg, H. 1979. *The structuring of organizations: A synthesis of the research*. New York: Prentice Hall.
Mintzberg, H., and J.A. Waters. 1985. Of strategies, deliberate and emergent. *Strategic Management Journal* 6: 257–272.
Porter, M. 1980. *Competitive strategy: Techniques for analysing industries and competitors*. New York: Free Press.
Semadeni, M., and A.A. Cannella. 2011. Examining the performance effects of post spin-off links to parent firms: Should the apron strings be cut? *Strategic Management Journal* 32: 1083–1098.
Stinchcombe, A.L. 1965. Social structure and organizations. In *Handbook of organizations*, ed. J.G. March. Chicago: Rand McNally & Company.
Thietart, R.A., and B. Forgues. 1995. Chaos theory and organization. *Organization Science* 6: 19–31.
Thietart, R.A., and B. Forgues. 1997. Action, structure and chaos. *Organization Studies* 18: 119–143.
Thietart, R.A., and B. Forgues. 2011. Complexity science and organization. In *The SAGE handbook of complexity and management*, ed. P. Allen, S. Maguire, and W. McKelvey. London: Sage.
Thompson, J.D. 1967. *Organizations in action*. New York: McGraw-Hill.

Charismatic Leadership

Richard Hall[1] and David Grant[2]
[1]University of Sydney, Sydney, NSW, Australia
[2]Griffith University, Business School, Brisbane, QLD, Australia

Definition Charismatic leadership is a style of leadership distinguished by high levels of expressiveness, self-confidence, moral conviction and emotional resonance, resulting in strong follower identification with the leader, alignment with the vision of the leader and inspiration to perform for the leader.

In his study of charismatic leadership, Weber (1947) observed that during periods of crisis a leader with exceptional qualities often emerges. Subsequent theories of charismatic leadership have extended Weber's work by defining these exceptional qualities in terms of key traits, behaviours and motivation. These include high levels of emotional expressiveness, self-confidence, self-determination, and a degree of freedom from internal conflict, which in turn facilitate a strong moral conviction in the leader of the righteousness of their views and values (House 1977). Studies have also highlighted the insight displayed by charismatic leaders into the needs, values and aspirations of followers, the leaders' eloquence and communication skills, their energy, and their self-sacrificial disposition (Bass 2008). Still other studies have examined the ways in which charisma, through a range of social and psychological processes, leads to follower identification with a charismatic leader, thereby influencing the followers' behaviour and attitudes (e.g., Shamir et al. 1993).

The influence of charismatic leaders on followers is thought to be profound and powerful. In organizational settings, the 'routinization' or institutionalization of charisma is often attempted, and it can occur in different ways: through the transfer of charisma to a new leader anointed by the old leader; through the translation of the charismatic leader's vision, values and norms into administrative rules and principles; or through the embedding of those values in an ▶ organizational culture.

Charismatic leadership is often associated with *transformational leadership*, or the process of motivating followers by appealing to their higher-order needs for achievement and self-actualization and moving them beyond self-interest to a concern for the organization and its objectives. While some see charismatic and transformational leadership as synonymous, and while charisma is an element of some conceptualizations of transformational leadership, the former places greater emphasis on the attribution of charisma to, and the identification of followers with, an extraordinary leader (Yukl 2010).

Charismatic leaders can have positive and negative effects on organizations. On the one hand, they can inspire sustained high performance;

ensure that social values are widely understood, shared and consistently used as the basis for action; instill trust; encourage open communication; and engage and align followers with the organization's mission and vision. Charismatic leaders might therefore be especially effective in developing, promoting, implementing and sustaining organizational strategy, particularly when it is radical or transformational.

On the other hand, it has been argued that charismatic leadership has a 'dark side': that it can encourage an uncritical and unquestioning organizational culture, suppress alternative views, stifle creativity, create enemies and allies and leave the organization and its members dependent and vulnerable. The influence wielded by unethical charismatic leaders can also allow them to manipulate followers to act in ways that are in the interests of the leader rather than those of the organization and its stakeholders. Accordingly, charismatic leadership has been associated with narcissistic leadership (Maccoby 2000), or the tendency of a leader to become preoccupied and even obsessed with his or her own power and self-aggrandisement. It has also been linked to the rise of the 'hero leader' and the 'curse of the superstar CEO' (Khurana 2002) – instances where organizations have hired high-profile, charismatic and dynamic leaders to lead radical change, with sometimes disastrous consequences.

The recognition of problems associated with the dark side of charismatic leadership has led to theories of 'post-heroic leadership'. Rather than focusing on the extraordinary personality of the leader, these theories emphasize a form of leadership that is shared and distributed throughout the organization and focused on achieving 'learning and growth for the organization as well as the people involved' (Fletcher 2004: 649). While post-heroic leadership seems more relevant to contemporary organizations, charismatic leadership continues to be a focus of much contemporary research.

See Also

► Leadership
► Organizational Culture

References

Bass, B.M. 2008. *The bass handbook of leadership*, 4th ed. New York: Free Press.
Fletcher, J.K. 2004. The paradox of postheroic leadership: An essay on gender, power and transformational change. *Leadership Quarterly* 15: 647–661.
House, R.J. 1977. A 1976 theory of charismatic leadership. In *Leadership: The cutting edge*, ed. J.G. Hunt and L.L. Larson. Carbondale: Southern Illinois University Press.
Khurana, R. 2002. The curse of the superstar CEO. *Harvard Business Review* 80: 60–66.
Maccoby, M. 2000. Narcissistic leaders: The incredible pros, the inevitable cons. *Harvard Business Review* 82: 69–77.
Shamir, B., R.J. House, and M.B. Arthur. 1993. The motivational effects of charismatic leadership: A self-concept based theory. *Organization Science* 4: 1–17.
Weber, M. 1947. *The theory of social and economic organizations*. New York: Free Press.
Yukl, G. 2010. *Leadership in organizations*, 7th ed. Upper Saddle River: Pearson.

Chicago School

Ross B. Emmett
Michigan State University, James Madison College, East Lansing, MI, USA

Abstract

Postwar economists at the University of Chicago initiated an economic approach that focused on the analytical relevance of the basic principles of price theory to empirical problems arising from the actual operation of markets. The Chicago School provided key ideas in the development of applied economic policy in industrial organization, labour economics, financial economics, economic development, law and economics, public finance, monetary policy, education and social policy and international economics.

Definition The Chicago School was made up of a group of economists, many of whom were associated with the University of Chicago, dedicated to the analytical relevance of price theory to empirical problems in the operation of markets across the spectrum of subfields in economics.

After the Second World War, economists associated with the University of Chicago initiated an economic approach that differed from the emerging consensus of the Anglo-American disciplinary mainstream. Where others adopted Keynesian macroeconomics, Chicago turned to monetary theory and history for the conditions of stable real economic growth; where economists at both Cambridges created theories of imperfect (monopolistic) competition, Chicago plumbed the depths of Marshallian perfect competition theory for new insights; and where the mainstream formalized Walrasian general equilibrium theory, Chicago focused on developing a scientific toolkit conducive to applied policy. These differences not only placed the Chicago School at odds with the mainstream's macroeconomics and industrial organization theories, but led to its leadership in the development of labour economics, law and economics, public finance, monetary theory, international economics, economic history, applied econometrics, social economics and financial economics. By the mid-1970s, the Chicago School was recognizable enough for its most familiar figure – Milton Friedman – to be honoured with the Nobel Memorial Prize in Economics (1976). During the subsequent 30 years, nine other Chicago economists were so honoured, and two other economics laureates were chosen for work undertaken when they were at Chicago. The Chicago School also influenced policy developments in the Reagan and Thatcher administrations, as well as government ministries and central banks around the world. By the beginning of the twenty-first century, the Chicago School had permeated the discipline sufficiently to often be considered near the centre of many branches of the economics profession.

The Chicago Approach

The Chicago School combined three aspects of the work of earlier Chicago economists to transform economics into an applied public policy science. The first is the well-known Chicago emphasis on the socially efficacious regulation of human behaviour in free markets through the price mechanism (Reder 1982: 13). Frank H. Knight's appreciation of the price mechanism was introduced to many Chicago students through his book *The Economic Organization* (Knight 1951). An appreciation for competitive market efficiency was reinforced in the mandatory price theory course, taught in the pre-war years by either Knight or Jacob Viner, and in the postwar era by Milton Friedman, Arnold Harberger and eventually Gary Becker. In financial economics, Chicago's reliance on price mechanisms to regulate market behaviour was captured in the efficient market hypothesis (emh) (Fama 1965).

Secondly, the Chicago School emphasized the 'analytical relevance' of basic economic principles to policy analysis, rather than the attempt to build theoretical models with greater 'descriptive accuracy' in their underlying assumptions (Friedman 1953). Knight was again central in initiating this aspect of Chicago economics because, despite his belief that the ideal type of perfect competition could not be used directly as a basis for empirical prediction (Knight 1999), he insisted that its basic principles could provide insight into policy disputes. His students overcame his reluctance to predict policy outcomes, rallying behind Friedman's (1953) call to test the 'analytical relevance' of basic Marshallian price theory by empirical investigation of policy effectiveness.

To render Marshallian principles operational for an applied policy science, Chicago economists needed to focus the economist's attention on changes in the cost structure surrounding the individual's decision. Although the principle which accomplished this focus can be stated as the rule to look at changes in tastes or values only after exhausting consideration of changes in cost structures, it generally goes by the name that Stigler and Becker used as the title of their article on the topic: *de gustibus non est disputandum* (Stigler and Becker 1977). Becker went on to make the assumption of stable tastes and values over time and among people the centrepiece of his 'economic approach to human behavior' (Becker 1976; also Becker and Murphy 2000), which Chicago economists have used to explore almost every aspect of human life. Other social scientists

claimed that Chicago School economists were engaged in 'economic imperialism' – a term Chicago-oriented economists have since embraced (Lazear 2000).

The third aspect of the Chicago School – the development of the tools for empirical analysis that assisted the work of the applied policy economist – has often been overlooked. Here, too, the postwar Chicago School followed in an earlier Chicago tradition, best represented by Wesley Mitchell, Henry Schultz and Paul Douglas. While Schultz and Douglas taught several of the members of the postwar Chicago School, it was perhaps Mitchell, who had been one of Chicago's first doctoral students in economics, who had the greatest impact through the National Bureau of Economic Research (NBER), which he founded at Columbia University in the 1920s. A strong connection developed between Chicago and the NBER in the early 1940s. Friedman completed his doctoral studies at Columbia and then worked with Simon Kuznets and later Anna Schwartz. The Friedman/Kuznets study of professional incomes (1945) formed the foundation for Friedman's permanent income hypothesis (1957). And Friedman and Schwartz worked together for over 20 years on their studies of US monetary history, which provided empirical support for monetarism (Friedman and Schwartz 1963, 1982).

Equally important to Chicago empirical side was the arrival in the mid-1940s of T. W. Schultz and D. Gale Johnson. Schultz became chair of the department in 1945 and steered its development until the 1970s. Between his leadership, the infusion of a strong applied policy focus in agricultural economics, public finance, Latin American economic development and labour economics, and the presence of Johnson (Antle and Sumner 1996), Zvi Griliches (1957), Gregg Lewis (1963) and Al Harberger (1954, 1962), the Chicago School found an empirical method that complemented its price theoretic approach. The Chicago tradition in econometric policy analysis has been furthered by the work of Nobel laureate James Heckman, whose econometric tools have transformed labour, education and social policy evaluation.

Empirical Analysis of Market Operation

While the Chicago School is often criticized for its use of simplistic competitive models, it became the leader in the empirical analysis of the outcomes of actual market operation. George Stigler's interest in how knowledge acquisition affects an economic agent's market behaviour led him through the application of statistical decision theory within a competitive price model to information search theory (Stigler 1961), a theme which Chicago economists, following Stigler's lead, made central to labour economics (Stigler 1962) and industrial organization (Stigler 1964), among other applications. Stigler also re-examined regulation, arguing that the existence of demanders as well as suppliers created a market-like regulatory environment. But in a political context without competitive pressures, demanders will capture the suppliers, rendering regulation socially inefficient (Stigler 1971). Stigler's empirical work on regulation also led him to be more sanguine regarding the benefits of antitrust action (Stigler 1982).

A similar development occurred with the work of ▶ Ronald H. Coase, who arrived at the University of Chicago's Law School in 1964, and took over editorship of the *Journal of Law & Economics*. Coase's two major articles, on the firm (Coase 1937) and on private versus social costs (Coase 1960), both continue the Chicago notion of using the basic insights of price theory to examine the operation of real markets. The first article introduced the costs of creating and sustaining markets, initiating the modern ▶ theory of the firm and the development of ▶ transaction cost economics, which reaches beyond industrial organization to law and economics, economic history and the emergence of new institutional economics. The second played an important part in the development of law and economics, as well as informing almost every area of economic policy analysis. Within the Chicago School, the use of Coase (1960) was largely shaped by Stigler's interpretation, labelled the 'Coase Theorem': in perfect competition, private and social costs would be equal (Stigler 1966: 113).

In the postwar period, the Chicago School developed an approach to economics that focused on the application of price theory to interesting empirical problems in the actual operation of markets. Although the Chicago approach differed significantly from its contemporaries in the 1950s and 1960s, the singularity of its vision placed the School at the vanguard of the emergence of the discipline's microeconomic theory and policy orientation in the latter third of the twentieth century.

See Also

- ▶ Coase, Ronald (Born 1910)
- ▶ Regulation/Deregulation
- ▶ Theory of the Firm
- ▶ Transaction Cost Economics

References

Antle, J.M., and D.A. Sumner (eds.). 1996. *Selected papers of D. Gale Johnson, vol 1: The economics of agriculture*. Chicago: University of Chicago Press.

Becker, G.S. 1976. *The economic approach to human behavior*. Chicago: University of Chicago Press.

Becker, G.S., and K.M. Murphy. 2000. *Social economics: Market behavior in a social environment*. Cambridge, MA: Belknap Press.

Coase, R.H. 1937. The nature of the firm. *Economica* 4(new series): 386–405.

Coase, R.H. 1960. The problem of social cost. *Journal of Law and Economics* 3: 1–44.

Fama, E. 1965. The behavior of stock-market prices. *Journal of Business* 38: 34–105.

Friedman, M. 1953. The methodology of positive economics. In *Essays in positive economics*. Chicago: University of Chicago Press.

Friedman, M. 1957. *A study of the consumption function*. Princeton: Princeton University Press.

Friedman, M., and S. Kuznets. 1945. *Income from independent professional practice*. New York: NBER.

Friedman, M., and A.J. Schwartz. 1963. *A monetary history of the United States, 1867–1960*. Princeton: Princeton University Press.

Friedman, M., and A.J. Schwartz. 1982. *Monetary trends in the United States and the United Kingdom, their relation to income, prices, and interest rates, 1867–1975*. Chicago: University of Chicago Press.

Griliches, Z. 1957. Hybrid corn: An exploration in the economics of technological change. *Econometrica* 25: 501–522.

Harberger, A. 1954. Monopoly and resource allocation. *American Economic Review* 44: 77–87.

Harberger, A. 1962. The incidence of the corporate income tax. *Journal of Political Economy* 70: 215–240.

Knight, F.H. 1951. *The economic organization*. New York: Augustus M. Kelley.

Knight, F.H. [1930] 1999. Statics and dynamics: Some queries regarding the mechanical analogy in economics. In *Selected Essays by Frank H. Knight*, vol. 1: *'What is Truth' in economics?*, ed. R.B. Emmett. Chicago: University of Chicago Press.

Lazear, E. 2000. Economic imperialism. *Quarterly Journal of Economics* 115: 99–146.

Lewis, H.G. 1963. *Unionism and relatives wages in the United States: An empirical inquiry*. Chicago: University of Chicago Press.

Reder, M.W. 1982. Chicago economics: Permanence and change. *Journal of Economic Literature* 20: 1–38.

Stigler, G.J. 1961. The economics of information. *Journal of Political Economy* 69: 213–225.

Stigler, G.J. 1962. Information in the labor market. *Journal of Political Economy* 70: 94–105.

Stigler, G.J. 1964. A theory of oligopoly. *Journal of Political Economy* 72: 44–61.

Stigler, G.J. 1966. *The theory of price*, 3rd ed. New York: Macmillan.

Stigler, G.J. 1971. The theory of economic regulation. *Bell Journal of Economics and Management Science* 2: 3–21.

Stigler, G.J. 1982. The economists and the problem of monopoly. *American Economic Review* 72: 1–11.

Stigler, G.J., and G.S. Becker. 1977. De gustibus non est disputandum. *American Economic Review* 67: 76–90.

China: Strategy in a Growth-Driven Economy

Marshall W. Meyer
University of Pennsylvania, The Wharton School, Philadelphia, PA, USA

Abstract

This article explores how firms strategize in China, where business can be more growth- than profit-driven – and the issue of whether or not the country needs to strategize at all, given that China is not a simple command economy where strategy emanates from the centre, nor is it entirely capitalistic and profit-driven. Rather, it is a decentralized economy where local units compete for growth, with unanticipated and sometimes untoward consequences for firms and markets. Firms engage in

diverse strategic responses to this, departing significantly from Western practice. It is possible that China will look to fundamental managerial innovation rather than to strategies drawn from Western experience to build firms of global scale and scope.

How do firms strategize in China where business can be more growth- than profit-driven? Or need they strategize at all? The latter is a significant question, because in China's growth-driven economy (a) the government, through the planning process and agencies like the National Development and Reform Commission, makes key economic decisions, (b) capital is abundant for sectors favoured by government policy but scarce for other sectors, (c) profits and shareholder returns may be less important than production and employment, and (d) preserving assets may be more important than deploying assets efficiently. At the asymptote, one could argue that strategy given these conditions is either (a) unimportant since firms, including privately owned firms, must ultimately adhere to government policy, or (b) exceedingly difficult to implement and execute consistently since reversals of government policy are frequent and unpredictable.

The picture becomes more complicated because in China (a) the economy is decentralized and local governments compete for growth targets, (b) an element of competition is local protectionism, (c) protectionism gives rise to market fragmentation, and (d) fragmentation reduces or reverses the advantages of domestic enterprises in comparison with foreign enterprises. These factors together render Chinese markets hypercompetitive for firms not enjoying the protection of the state. The strategic responses of firms include (a) buffering hypercompetition by seeking government aid, (b) intensifying competition by initiating price wars, (c) pursuing opportunistic revenues by entering into seemingly profitable lines of business whether related or unrelated to existing businesses, and (d) radically decentralizing the firm in order to best accommodate demand in highly fragmented and rapidly changing markets. Interestingly, the first and second responses may be combined, where (a) the firm, consistent with the preference for growth, is a revenue- rather than a profit-maximizer, (b) revenue maximization is achieved by cutting costs and pursuing government subsidies, (c) an opportunity to sideline rivals with less government backing and hence less staying power is perceived, and (d) if successful the strategy yields a triple win for the firm, the government and government officials: the firm eliminates competitors, the government realizes higher regional domestic product (the regional counterpart of gross domestic product or GDP) and tax revenues, and officials are advanced to higher posts based on successful performance.

This article begins by establishing that China is a growth-driven economy. A recent paper (Meyer 2011) is summarized, where I coin the term institutionalized GDP growth, and I identify mechanisms through which institutionalized growth penetrates to the firm level. Next I consider how firm strategies have been shaped by (a) DECENTRALIZATION, regional competition and market fragmentation, (b) foreign competition and (c) relentlessly increasing costs. The ideas developed here are preliminary and intended as suggestive of the subtlety and variety of strategies Chinese firms have evolved. The third section shows that a decentralized growth-driven economy is especially conducive to price wars, and then compares the price-war strategy with an emerging counter-strategy. Interestingly, the counter-strategy is supported, among other things, by a radical decentralization of the firm paralleling and leveraging the deep decentralization of the Chinese economy. There nevertheless remains a critical difference between the price-war strategy and the counter-strategy: whereas the price wars are easily triggered and imitated, the counter-strategy is neither easily implemented nor imitated.

China as a Growth-Driven Economy

China has a growth-driven economy. GDP targets have been hardwired into successive 5-year plans,

progress towards these targets is measured at four levels of government (national, provincial, prefectural/municipal and county/district) and reported semi-annually, and these targets have been fulfilled and over-fulfilled mainly owing to government policies promoting investment in fixed assets and low-cost manufacturing for export. From 2003 to 2011, Chinese GDP grew at 9% or more annually; only in 2012 did growth fall to 7.8%. Throughout, fixed-asset investment grew more rapidly than GDP, even though net exports moderated from 2009 onwards; investment-driven growth, as a consequence, shrank household consumption in proportion to GDP, inverting the increased contribution of household consumption to GDP that is normal for countries at China's stage of development. Currently, China's 12th Five-Year Economic Plan calls for a modest reduction of the official GDP growth target, from 7.5% to 7% coupled with a boost in consumer spending. Observers remain sceptical, however: 'The last Five-Year plan (2005–2010) opened with similarly fervent pledges to rebalance, but exactly the opposite happened: investment now plays an even bigger role in China's GDP formation, while the share of household consumption as a percentage of GDP continues to shrink' (The Telegraph 2011).

It is less clear whether Chinese *firms*, as distinct from China, are more growth- than profit-driven; in other words, whether preternatural or institutionalized growth operates at the level of firms. The Chinese central government has clearly favoured large firms, evident in the policy of 'grasping the large while letting go of the small' (*zhua da fang xiao*) adopted by the 15th Communist Party Congress in 1997. Still, profit-seeking behaviour abounds as evident by the fact that (a) some of the largest state owned enterprises remain highly profitable and seek to retain their earnings while the State-owned Assets Supervision and Administration Commission (SASAC) has been trying to extract a portion of these earnings, as dividends, from these enterprises, and (b) according to the latest Hurun rich list (Hurun Report 2013), 251 individual Chinese amassed fortunes in excess of US$1 billion in 2012, the wealthiest concentrated in real estate. However, profit-seeking may be attenuated by (a) weak corporate governance, as evidenced by the presence of a controlling shareholder in more than 99% of Shanghai-listed firms (Amit et al. 2010: Table 1), (b) SASAC and China Securities Regulatory Commission (CSRC) regulations, dating from 2006, severely limiting stock options and grants of restricted stock to managers, and (c) various government policies forcing large enterprises to swallow increased commodity prices rather than passing them on to customers. Whether and how growth targets are communicated directly to Chinese managers, and with what effect, is for the most part unexplored. However, interaction between the government officials and senior management can be intense, as indicated by the following: (a) unscheduled meetings with the government may consume half or more of senior managers' time, and (b) the proportion of former government officials, communist party officials and former state-owned enterprise managers among private business owners has risen dramatically, from one in three in 2004 to two in three in 2007 (Asia Times 2007). A reasonable hypothesis is that the closer and denser the ties between managers and the government, the greater the salience and the stronger the incentives to meet growth targets, and hence the more responsive the firm to these targets. Thus, SOEs will be more sensitive to growth targets than private firms, and firms closer to Beijing or located in inland regions anticipating especially rapid development will be more sensitive to growth targets than firms in the south of China.

Strategies in a Growth-Driven Economy

Strategies in Response to Decentralization, Regional Competition and Fragmentation

If China were a Soviet-style centrally planned economy with targets assigned annually to sectors and firms, then growth would be attenuated owing to the well-known ratchet effect (Weitzman 1980). However, China has (a) pursued administrative decentralization where state ownership, defined as ownership by the whole people, is retained but responsibility for enterprise operations and

results in all but the largest firms is delegated to local officials and enterprise managers (Wu 2005), and (b) pitted local governments against each other (Xu 2011) by proffering political advancement to officials achieving the highest rates of GRP (gross regional product) and employment growth (Li and Zhou 2005; see, however, Shih et al. (2012), who argue that network ties are more important to advancement than growth). The combination of administrative decentralization and competition for regional growth has led to regional preferences if not outright protectionism (though the strongest evidence for regionalism comes from the late 1990s – see Young 2000; Poncet 2003, 2005; Bai et al. 2004); local preferences and protectionism, in turn, have rendered many industries fragmented and markets hypercompetitive (Section 7 of the 2008 Anti-Monopoly Law bans administrative or local-government monopoly, but its enforcement remains uneven). As a consequence, building firms of national scope and consolidating markets has been a strategic priority for the Chinese central government and many managers, yet remains elusive for most firms.

Fragmentation remains greatest in legacy industries like steel. At the behest of the central and Shanghai governments, one of China's leading steel firms, Shanghai-based Baosteel Group, merged with two local steel producers, Shanghai Steel and Meishan Steel, in 1988, combining a highly efficient producer with two inefficient producers. A decade later, Baosteel was encouraged to absorb the Guangdong Steel Group, also an inefficient producer. The integration of Baosteel with Guangdong Steel remains incomplete because the interests of Baosteel, a State Council company, are at odds with the Guangdong and Guangzhou governments: whereas Baosteel has sought both to reduce the capacity of Guangzhou Steel and to realize gains from labour-saving technology, the Guangzhou government has sought to protect the workforce as well as its investment in Guangzhou Steel.

Consolidation of non-legacy industries is more or less daunting depending on (a) the number of jobs at stake, (b) the threat, if any, to a local monopoly, and (c) the impact on local government revenues. China International Marine Container (CIMC), a leading manufacturer of transportation equipment, successfully integrated the Chinese shipping container industry in the early 1990s, owing to a compelling value proposition *and* support of the Shenzhen government: (a) integration would yield substantial scale and scope efficiencies; (b) few jobs were at stake since the Chinese shipping container industry was undeveloped; (c) there were no local monopolies since shipping containers were built to global standards and sold to global customers; and (d) local government shareholders of CIMC subsidiaries would earn dividends based on production rather than profitability. So compelling was this proposition that, despite staunch opposition from China Ocean Shipping Company, COSCO, one of two major shareholders and a State Council company, CIMC implemented the strategy, consolidated the Chinese shipping container industry and, ultimately, dominated the global shipping container market (Meyer and Lu 2005). CIMC subsequently diversified into semi-trailers. However, replicating the container strategy in the semi-trailer business proved challenging because standards and specifications for domestic semi-trailers were controlled by local governments intent on protecting local producers. CIMC chose not to challenge local standards and specifications but, rather, to manufacture semi-trailer modules assembled to multiple standards and specifications; in the long run, of course, CIMC aims to reduce the panoply of local standards currently in place.

Local protectionism and industry fragmentation increase costs on several dimensions, including (a) capacity costs (owing to the penchant for fixed-asset investment), (b) operating costs (owing in part to inland logistics costs, higher in China than the US or India), and (c) transaction and administrative costs (owing to the labyrinth of regulations and government entities). Short-run capacity costs are deferred by rolling over bank loans financing expansion. Operating costs are offset by subcontracting, adding to the risk of the well-known 'bullwhip effect' in supply chains and placing quality at risk (Lyles et al. 2008). It is expected that massive infrastructure development will curb logistics costs directly, by reducing

the cost of transport, and indirectly, by encouraging industry consolidation. Transaction and administrative costs are managed by cultivating officials whose action or inaction can affect these costs dramatically, making the relationship with government a strategic priority for many firms.

Strategies in Response to Foreign Competition

With reform and opening beginning in 1978, China adopted policies designed to attract foreign investment. These policies included special economic zones and free trade zones, preferential tax policies and 'one-stop shopping', facilitating licensing and hiring of qualified workers by foreign-invested and wholly foreign-owned enterprises. While these policies drew manufacturing to China and created much-needed employment, they left domestic firms disadvantaged in several respects, including (a) key capabilities, including product design, remaining offshore; (b) branding of domestic firms; and (c) weak domestic institutions such that the environment most conducive to financing and operating large business firms, including a predictable legal process, also remaining offshore. In response, Chinese firms took several tacks, among them (a) paralleling the earlier experience of the 'tiger' nations of South East Asia, upgrading from OEM to ODM (original design manufacturing) capability; (b) acquiring Western brands, most notably Lenovo's acquisition of the ThinkPad brand and Geeley's acquisition of Volvo; and (c) 'round-tripping', that is, investing in overseas (mainly Hong Kong) holding companies whose operating units are, in turn, located in China (Lenovo is again illustrative). In markets for complicated machinery, for example, elevators or electrical generating equipment, dominated by foreign firms, domestic firms have also pursued a niche or 'sandwich' strategies of providing replacement parts or maintenance, since they can operate on a smaller scale than global firms.

The ODM, brand acquisition, round-tripping and sandwich strategies are largely piecemeal, and the Chinese government has encouraged firms to pursue more comprehensive strategies including: (a) innovation and (b) development of rural distribution. Concerning innovation, China has issued a 'National Patent Development Strategy (2011–2020)', calling for two million patents to be filed annually by 2020 (compared to 244,000 US patents in 2010). The patent development strategy may be accompanied by a dramatic increase of national investment in R&D: the Battelle Institute (2013) projects that 2013 Chinese R&D expenditures will be 1.65% of GDP, compared with 1.55% in 2011. (By comparison, 2013 US R&D expenditures are projected at 2.66% of GDP compared with 2.7% in 2011). Less visible but potentially quite significant is the development of rural distribution in China. Markets in the countryside, still half of China's population, are pursued by (a) word-of-mouth marketing campaigns in rural villages, (b) Internet-based information kiosks in locales too small to support retail outlets, (c) unique countryside logistics systems where delivery vans circle among multiple towns and villages rather than travelling to and from a central distribution point, and (d) products uniquely suited for rural customers, for example, inexpensive single-use sachets of personal care and cleaning products, and clothes washers that are rat-proof and can double as vegetable washers.

Strategies in Response to Inflation

China has experienced periodic bouts of high inflation as recently as 2007–2008 and 2010–2011. Quite apart from inflation driven by currency exchange policies, an inflow of 'hot money' seeking RMB appreciation, and the massive 2008–2010 fiscal stimulus, the economy appears to have passed the 'Lewis turning point', where a labour surplus turns into a shortage, reflected initially in rural wages and subsequently in urban wages. New evidence suggests that rural wages began rising dramatically from 2003 and urban wages from 2006 (Zhang et al. 2011), even though wage inflation paused in 2008–2009. The strategic implications are profound given China's dependence on low-cost manufacturing for export. Among other possibilities, firms may seek to: (a) substitute capital in the form of machinery, process efficiency and so on for labour (e.g., China's largest auto producer, SAIC, is

automating rapidly), (b) relocate inland where labour costs are lower (e.g., Foxconn's relocation of assembly operations from Shenzhen to Chengdu), or (c) migrate from low-value to high value products (e.g., the Wenzhou button industry has moved from ordinary to laser-etched buttons commanding premium prices). For pillar Chinese industries, however, the choices are more daunting. Steel again is illustrative. China imports iron ore. As wages and commodity prices increase, largely owing to Chinese demand, the iron ore exporters, principally Australia and Brazil, are attempting vertical integration of the industry by initially acquiring shipping capacity and later developing domestic production capacity. (Australia and Brazil have already acquired shipping capacity, glutting the market and driving bulk shipping rates towards historic lows.) The Chinese are also attempting vertical integration by investing in overseas iron ore producers to stabilize prices and preclude their vertical integration. (Chinalco tried unsuccessfully to acquire a substantial interest in Australia's Rio Tinto.) Whether China's experience in steel will be repeated in other key industries, coal for example, is less clear. Regardless, the joining of wage and commodity inflation with China's dependence on imported iron ore and, increasingly, coal, reduces the likelihood that the global steel industry will remain centred in China.

Strategies in Response to Price Wars

Price wars, as mentioned, are commonplace in China, hardly surprising since one would expect price competition to be more frequent where growth takes precedence over profitability, markets are fragmented, and accountability to shareholders remains weak in comparison with the US and the EU. A crucial strategic question for certain Chinese firms, a small number to be sure, is not how and when to initiate price wars but, rather, how to compete effectively while avoiding price wars. The conventional Western answers lie in branding and innovation: teach the customer to value the brand; create products adding value for the customer wants faster than your competitors can. A less conventional answer pursued by Haier, China's largest appliance maker, lies in accounting and management systems diverting attention from growth targets by focusing relentlessly on profit. Rather than setting sales targets and rewarding sales in excess of targets, Haier's management and accounting systems allow individual employees to (a) discriminate profitable from unprofitable products, (b) discriminate profitable from unprofitable customers, (c) calculate individual and team contribution to bottom-line profitability, and (d) calculate individual and team earnings based on contribution to profitability. The management system is supported by some unusual managerial practices, including: (a) decentralization of target-setting to 2000 'self-owned' teams where setting and meeting aggressive yet realistic targets is rewarded (while setting low-ball targets or missing aggressive targets are punished), (b) inversion of the organizational hierarchy where teams have authority to choose members, compensate performance, and 'push back' on the organization for improved or redesigned products and services and responsibility to pay for improvements and new designs, as well as (c) a culture of 'end-to-end', reminding employees that the path to profitability lies in meeting end users' requirements.

There are many sources of resistance to accounting systems where profitability is calculated at every point in the organization, including support functions, and to managerial practices essentially gutting middle management and holding front-line employees responsible for bottom-line results. Little is known, for example, on whether market-like transactions are feasible in highly interdependent systems, and on whether focus on the bottom line will be conducive to teamwork or corrosive of it. It is possible, of course, that these issues will be rendered moot as China evolves towards more normal rates of economic growth and stronger corporate governance. But it is also possible that growth and fragmented markets have become deeply institutionalized and corporate governance not, in which case the burden will fall mainly on management to produce the sustainable profits firms require.

Summary

The most important point is that China is not a simple command economy where strategy, such as it is, emanates from the centre. Nor is China entirely capitalistic and profit-driven. Rather, China is a decentralized and regionalized economy where local units compete for growth, with unanticipated and, in some respects, untoward consequences for firms and markets. These consequences include economic fragmentation, hypercompetition and frequent price wars. Firms engage in diverse strategic responses to these conditions, including initiatives aimed at industry consolidation, extensive subcontracting, pursuit of favourable government policies, acquisition of foreign brands and so-called 'round-tripping', whereby domestic firms reorganize as foreign firms, in effect co-opting overseas institutions while operating in China. Some Chinese firms, additionally, are today looking inward to their accounting and management systems rather than outward to broad strategic initiatives in their quest for sustained profitability. These efforts are nascent and substantial departures from Western practice. Hence, it is possible that China will look to fundamental managerial innovation rather than to strategies drawn from Western experience to build firms of global scale and scope.

See Also

- ▶ Capitalism
- ▶ Decentralization
- ▶ Foreign Direct Investment (FDI) and Economic Development
- ▶ Strategic Planning

References

Amit, R., Ding, Y., Villalonga, B., and Zhang, H. 2010. The role of institutional development in the prevalence and value of family firms. Working paper, Harvard Business School. Available at http://www.hbs.edu/faculty/Publication%20Files/10-103.pdf. Accessed 29 Apr 2013.

Asia Times. 2007. China's entrepreneurs aspire to political office. Available at http://www.atimes.com/atimes/China_Business/IC01Cb01.html. Accessed 29 Apr 2013.

Bai, C.-E., Y. Du, Z. Tao, and S.Y. Tong. 2004. Local protectionism and regional specialization: Evidence from China's industries. *Journal of International Economics* 63: 397–417.

Battelle Institute. 2013. 2013 Global R&D funding forecast. Available at http://www.researchcaucus.org/schedule/2013/28Feb2013/2013%20Global%20RD%20Funding%20Forecast%20RD%20Caucus%20Presentation.pdf. Accessed 29 Apr 2013.

Hurun Report. 2013. The fifth year to release Hurun Wealth Report 2013. Available at http://www.hurun.net/usen/HRRL.aspx. Accessed 29 Apr 2013.

Li, H., and L. Zhou. 2005. Political turnover and economic performance: The incentive role of personnel control in china. *Journal of Public Economics* 89: 1743–1762.

Lyles, M., B. Flynn, and M. Frohlich. 2008. All supply chains don't flow through: Understanding supply chain issues in product recalls. *Management and Organization Review* 4: 167–182.

Meyer, M. 2011. Is it capitalism? *Management and Organization Review* 7: 5–18.

Meyer, M., and X. Lu. 2005. Managing indefinite boundaries: The strategy and structure of a Chinese business firm. *Management and Organization Review* 1: 57–86.

Poncet, S. 2003. Measuring Chinese domestic and international integration. *China Economic Review* 14: 1–21.

Poncet, S. 2005. A fragmented China: Measures and determinants of Chinese domestic market disintegration. *Review of International Economics* 13: 409–430.

Shih, V., C. Adolph, and M. Liu. 2012. Getting ahead in the communist party: Explaining the advancement of central committee members in China. *American Political Science Review* 106: 166–187.

The Telegraph. 2011. China's addiction to GDP highs could kill Five-Year plan. Available at http://www.telegraph.co.uk/finance/china-business/8363361/Chinas-addiction-to-GDP-highs-could-kill-Five-Year-plan.html. Accessed 29 Apr 2013.

Weitzman, M. 1980. The 'ratchet principle' and performance incentives. *Bell Journal of Economics* 11: 302–308.

Wu, J. 2005. *Understanding and interpreting Chinese economic reform*. Singapore: Thomson/South-Western.

Xu, C. 2011. The fundamental institutions of China's reforms and development. *Journal of Economic Literature* 49: 1076–1151.

Young, A. 2000. The razor's edge: Distortions and incremental reform in the People's Republic of China. *Quarterly Journal of Economics* 115: 1091–1135.

Zhang, X., J. Yang, and S. Wang. 2011. China has reached the Lewis turning point. *China Economic Review* 22: 542–554.

Choice Modelling

Timothy Devinney[1] and Nidthida Lin[2]
[1]University of Technology Sydney, Sydney, NSW, Australia
[2]The University of Newcastle, Newcastle, NSW, Australia

Abstract

Choice modelling represents the structured examination of individual decision-making among designed alternatives. Although most commonly used to examine individual choices – for example, among product alternatives where the features of the products vary – choice models can be utilized to examine any set of alternatives that can be decomposed into distinct parts, such as investments or market-entry alternatives. The most methodologically valid approach to choice modelling is discrete choice modelling, which has its basis in random utility theory (RUT) and relies on a number of simplifying assumptions to link its conceptual formulation to a specific empirical model.

Definition Choice modelling is a popular stated preference method used for understanding stated choice among discrete alternatives. A choice study uses experimental designs to create sets of alternatives that vary in their attributes and features and that statistically model the choices made and not made to yield measures of the relative importance of each attribute.

Choice modelling represents the empirical examination of systematically varied alternatives. The examination of what choices are made, controlling for alternatives that could have been chosen but were not, allows the investigator to empirically estimate a 'model of choice'. Such models can be estimated at the individual, group and organizational level, and can include choices made among currently available alternatives – such as products available in a grocery store environment – as well as alternatives that could potentially be made available, as in the design of new products. The two most common approaches to choice modelling are discrete choice modelling (DCM) and traditional conjoint analysis (CA).

Discrete Choice Modelling

DCM uses systematic factorial manipulation of independent variables – known as attributes or features – and their levels to elicit rank orders that capture the impact of the independent variables. The goal of discrete choice experiment is to characterize and predict individual choice among complex alternatives. Its most common application is to consumer choice (Carson et al. 1994), although it is increasingly finding its way into the managerial literature (Priem et al. 2011).

The conceptual heart of DCM is random utility theory (RUT), originally developed by Thurstone (1927), who was concerned with whether or not respondents could differentiate levels of psychological stimuli. Marschak (1960) extended Thurstone's work by showing that if one reinterpreted a stimulus as having 'utility', one could derive an RUT structure via maximization. McFadden (1973) generalized the use of DCM by extending the original model of paired comparisons (i.e., decisions between two choice alternatives) to multiple comparisons.

The Marschak approach to RUT follows from the fact that one can posit a latent construct called 'utility' that underlies individual decision alternatives. This latent utility can be characterized using by two generalizable components. The first is a systematic (observable) component, which consists of the utility values of the different choice alternatives and covariates explaining individual differences in choices. The second is a random (unobservable) component, which includes all unidentified factors influencing choices. Hence, the basic axiom of RUT can be written as:

$$U_{in} = V_{in} + \varepsilon_{in}$$

where U_{in} is the latent utility that individual n associates with alternative i, V_{in} is the

systematic, observable component of that utility and ε_{in} is the random component. In other words, ε_{ni} represents the difference between the true utility U_{in} and the representative utility that the researcher observes and captures in V_{in}. Hence, characteristics of ε_{in} depend critically on the specification of V_{in}.

It follows from basic economic theory that the individual chooses the choice alternative that provides the greatest utility. Hence, the behavioural model underlying RUT is that the individual will choose the alternative i if and only if $U_{in} : U_{jn} \forall j \neq i$. However, utilities are inherently stochastic. Researchers can predict the probability that individual n will choose alternative i, but not the exact alternative that individual n will choose. Thus, the probability that individual n chooses option i from a set of competing options – known as a choice set C_n – is:

$$P(i|C_n) = P\left[U_{in} > \text{Max}\left(U_{jn}\right)\right] = P\left[\left(V_{in} + \varepsilon_{in}\right) > \text{Max}\left(V_{in} + \varepsilon_{jn}\right)\right] \text{ for all } j \text{ options in choice set } C_n.$$

In DCM, each attribute will vary across two or more *levels*. A choice alternative described in terms of the attributes and levels in the study is called a *profile*, or choice configuration. In its experimental application, subjects are exposed to a series of *choice tasks*, which consist of two or more profiles, and asked to select the profile they consider to be the best or most preferred. In cases where forced choice is not realistic, researchers also include the option of choosing none of the offerings. The design of choice tasks is based on the experimental design literature (see Street and Burgess (2007) and Street et al. (2005)).

The effectiveness of DCM is in the link between its theoretical basis and its empirical estimation. By making different assumptions as to the probability distributions of error components, one can derive different probabilistic discrete choice models (e.g., multinomial logit, multinomial probit, nested logit, generalized extreme value models and mixed logit). The main probability axiom used to develop an operational model is known as the independence-from-irrelevant-alternatives (IIA) axiom, which states that *the ratio of the probabilities of choosing one alternative over another is unaffected by the presence or absence of any additional alternatives in the choice set*. IIA implies that the random components in the utility are independent across alternatives and are identically distributed; yielding a closed form of choice probabilities that can be characterized as a multinomial logit (MNL) model.

Discrete Choice Modelling Versus Traditional Conjoint Analysis

Within the management realm, it is not uncommon to find conjoint analysis being confused with DCM. Although both approaches aim to estimate the importance, or utility, of attributes, in traditional CA respondents evaluate the choice alternative independently, via an attractiveness rating scale. Rather than rating each alternative, respondents in DCM simultaneously consider two or more alternatives and are asked to select the most preferred or most important alternative, from which the researcher empirically derives the measure of attractiveness (i.e., the utility). DCM is an empirically more valid approach in that it removes artefacts that arise from the use of rating scales.

DCM and CA also differ in their theoretical foundation. While DCM is based on a long-standing, well-tested theory of choice behaviour, the axioms of CA have a very restrictive relationship to utility theory. Therefore, the emphasis in CA-related research has been predicting outcomes (i.e., how preferences are formed and what processes individuals use to form them), but CA methods have rarely been used to examine and model preference processes per se. In other words, one can view CA as a method of explaining the choice behaviour of numbers, while DCM, which can take interlinked behaviours into account, provides an explanation of the choice behaviour of humans (Louviere et al. 2011). Another difference lies in the RUT foundation of DCM and its direct link to different stages in the ▶ decision-making process. A DCM choice set always includes at least one feasible alternative (i.e., status quo, no-take or 'none' alternative), while CA

sometimes offers respondents an entire set of infeasible alternatives (Louviere et al. 2011).

Applications of Discrete Choice Modelling

DCM is used to examine choices made by both individuals and organizations (i.e., the decision unit assumed to be a person) in many different application areas, including problems that require decision makers to make trade-offs among attributes and alternatives. A non-exhaustive list of areas in which DCM has been used includes environmental science Adamowicz et al. 1998), health (Propper 1995), marketing (Kamakura and Srivastava 1984; Koelemeijer and Oppewal 1999; Moore et al. 1999; Auger et al. 2008), tourism (Haider and Ewing 1990; Crouch et al. 2009), transportation (Hensher 1989; Brandley and Gunn 1990), management (Brazell et al. 2005) and international business (Buckley et al. 2007).

Among the most common use of DCM is *attribute scaling*. Traditional surveys commonly ask respondents to evaluate statements, attributes or items on some type of scale (e.g., Likert) for which there are well-recognized methodological issues that can render the results questionable (Louviere et al. 2005; Drasgow et al. 2010; Spector and Brannick 2010). By decomposing an individual's response to one or more alternatives described by combinations of attribute levels that force respondents to make trade-offs, indirect methods such as DCM avoid many of the biases that arise from individual differences in scale use or interpretation.

Researchers have been using DCM for *competitive analysis*. Actions of competitors can be modelled and simulated by varying attribute levels of alternatives. Advance statistical methods for heterogeneity allow researchers to further understand how subgroups (e.g., segments) are likely to respond to actions taken by competitors (see Wedel and Kamakura (2000) and DeSarbo et al. (2006)). The classic application is the evaluation of pricing strategies (i.e., how choices respond to variations in price), in which incentive compatibility is particularly critical.

In many applications, the choice being modelled is not a single discrete choice, but more complex *multiple choices and allocation models*. For example, a firm may offer bundles of services, plus 'extras' that individuals have an option to select if desired. Individuals may decide to take not just one unit of offering, but, instead, multiple units spread across multiple providers. DCM can be designed to examine these complicated choice behaviours. Researchers also use DCM to *identify and optimize the offering or policy configurations* – searching for a profile (attribute-level combination) that maximizes the probability of choices. To identify an optimal profile, researchers need to take into account other competing alternatives (competitors).

See Also

- ▶ Decision-Making
- ▶ Expected Utility
- ▶ Rational Expectations

References

Adamowicz, W., P. Boxall, M. Williams, and J. Louviere. 1998. Stated preference approaches for measuring passive use values: Choice experiments and contingent. *American Journal of Agricultural Economics* 80: 64–75.

Auger, P., T.M. Devinney, J.J. Louviere, and P.F. Burke. 2008. Do social product features have value to consumers? *International Journal of Research in Marketing* 25: 183–191.

Brandley, M.A., and H. Gunn. 1990. Stated preference analysis of values of travel time in the Netherlands. *Transportation Research Record* 1285: 78–89.

Brazell, J., T. Devinney, and D. Midgley. 2005. Relating customer value to strategic competence: A discrete choice measurement approach. *Research in Competence-Based Management* 1: 15–45.

Buckley, P.J., T.M. Devinney, and J.J. Louviere. 2007. Do managers behave the way theory suggests? A choice-theoretic examination of foreign direct investment location decision-making. *Journal of International Business Studies* 38: 1069–1094.

Carson, R.T., J.J. Louviere, D.A. Anderson, P.B. Arabie, D.A. Hensher, R.M. Johnson, W.F. Kuhfeld, D. Steinberg, J. Swait, H. Timmermans, and J.B. Wiley. 1994. Experimental analysis of choice. *Marketing Letters* 5: 351–368.

Crouch, G.I., T.M. Devinney, J.J. Louviere, and T. Islam. 2009. Modelling consumer choice behaviour in space tourism. *Tourism Management* 30: 441–454.
DeSarbo, W.S., R. Grewal, and J. Wind. 2006. Who competes with whom? A demand-based perspective for identifying and representing asymmetric competition. *Strategic Management Journal* 27: 101–129.
Drasgow, F., O.S. Chernyshenko, and S. Stark. 2010. 75 years after Likert: Thurstone was right! *Industrial & Organizational Psychology* 3: 465–476.
Haider, W., and G.O. Ewing. 1990. A model of tourist choices of hypothetical Caribbean destination. *Leisure Sciences* 12: 33–47.
Hensher, D.A. 1989. Urban tollways and the valuation of travel time savings. *Economic Record* 66: 146–156.
Kamakura, W.A., and R.K. Srivastava. 1984. Predicting choice shares under conditions of brand interdependence. *Journal of Marketing Research* 21: 420–434.
Koelemeijer, K., and H. Oppewal. 1999. Assessing the effects of assortment and ambience: A choice experimental approach. *Journal of Retailing* 75: 319–345.
Louviere, J.J., T.C. Eagle, and S.H. Cohen. 2005. *Conjoint analysis: Methods, myths and much more*. CenSoC Working paper No. 05-001. Sydney.
Louviere, J.J., T.N. Flynn, and R.T. Carson. 2011. Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling* 3: 57–72.
Marschak, J. 1960. *Binary choice constraints on random utility indications*. Stanford: Stanford University Press.
McFadden, D. 1973. *Conditional logit analysis of qualitative choice behavior*. New York: Academic.
Moore, W.L., J.J. Louviere, and R. Verma. 1999. Using conjoint analysis to help design product platforms. *Journal of Product Innovation Management* 16: 27–39.
Priem, R.L., B.A. Walters, and S. Li. 2011. Decisions, decisions! How judgment policy studies can integrate macro and micro domains in management research. *Journal of Management* 37: 553–580.
Propper, C. 1995. The disutility of time spent on the United Kingdom's National Health Service waiting lists. *Journal of Human Resources* 30: 677–700.
Spector, P.E., and M.T. Brannick. 2010. If Thurstone was right, what happens when we factor analyze Likert scales? *Industrial & Organizational Psychology* 3: 502–503.
Street, D.J., and L. Burgess. 2007. *The construction of optimal stated choice experiments: Theory and methods*. Hoboken: Wiley-Interscience.
Street, D.J., L. Burgess, and J.J. Louviere. 2005. Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments. *International Journal of Research in Marketing* 22: 459–470.
Thurstone, L. 1927. A law of comparative judgement. *Psychological Review* 34: 273–286.
Wedel, M., and W.A. Kamakura. 2000. *Market segmentation: Conceptual and methodological foundations*. Boston: Kluwer Academic.

Clausewitz, Carl von (1780–1831)

Andreas Herberg-Rothe
Munster, Germany

Abstract

Clausewitz's concept of war at the beginning of chapter one of *On War* is based on a hierarchical understanding, whereas his concept of the trinity at the end of the same chapter is more like a floating balance between three different tendencies. The common understanding of Clausewitz's concept of strategy has nonetheless followed only the hierarchical structure, which is not in line with the concept of a floating balance in his trinity. This article examines the relation of purpose, aims and means (*Zweck*, *Ziel* and *Mittel*) in Clausewitz's theory and emphasizes that this relation is methodologically comparable to the floating balance of Clausewitz's trinity. Clausewitz's concept of strategy can be characterized as the attempt of combining a hierarchy with the floating balance of purpose, aims and means (ends, ways, means).

In summarizing Carl von Clausewitz's different concepts of strategy, one could say that strategy 'must therefore define an aim for the entire operational side of the war that will be in accordance with its purpose' (Clausewitz 1984: 177). War is thus an act of force to compel our enemy to do our will, according to Clausewitz's definition of war at the beginning of his famous first chapter of *On War* (Clausewitz 1984: 75). He continues: 'Force . . . is thus the means of war; to impose our will on the enemy is its purpose. To secure that purpose we must make the enemy defenceless and that in theory is the true aim of warfare. That aim takes the place of the purpose, discarding it as something not actually part of war' (p. 75). This seemingly simple last sentence reveals the core problem: What does it mean that the aim 'takes the place' (in German: *vertritt*) of the purpose? Are they identical? Are they different expressions of the same?

At the beginning of the second chapter of book one Clausewitz states that 'if for a start we inquire into the objective of any particular war, which must guide military action if the political purpose is to be properly served, we find that the object of any war can vary just as much as its political purpose and its actual circumstances' (Clausewitz 1984: 90). It is true, nevertheless, that not every aim and means serve the given purpose. And finally, the problem is that each tendency of the purpose-aims-means rationality has a rationality on its own, which Clausewitz emphasizes in his proposition that war has its own grammar, although not its own logic. Clausewitz comprehends war here in the context of a purpose-aims-means rationality. Because of his different war experiences (see Herberg-Rothe 2007), however, he deals with various differentiations within this concept. These are systematically justifiable if, within the purpose-aims-means rationality, different dimensions are distinguished.

The main purpose of this article is to show Clausewitz's (and sometimes beyond him) distinctions between the rationality of the whole process, the rationality of the separate aspects of purpose, aims and means in warfare, and finally their conflicting tendencies.

In the first part, we examine the importance that purpose-aims-means rationality has to Clausewitz's concept of war, and the conflicting rationality of these different aspects. The internal contradictions 'of modern warfare', as seen by Clausewitz, will be explained in the second part. In a summary we will attempt to combine the two aspects in order to accentuate the different inherent tendencies within warfare. If a single one of these rationales is portrayed as an absolute, it will result in exceptions, or a delimitation of war. Clausewitz's approach to a solution is the 'trinity', in which he defines war by different, even opposing, tendencies, each with its own rules. Nevertheless, since war is 'put together' in this concept of three tendencies, it is necessary to consider these tendencies themselves and their interactions and conflicts at the same time, rather than declaring one of the three as absolute. To summarize: if we go to war, there is a purpose: why we go to war. Different purposes are possible. Each of these possible purposes is connected with different possible aims, and each aim can be achieved by different means. Clausewitz's concept of strategy can be characterized as the attempt of combining a hierarchy with a floating balance of purpose, aims and means (ends, ways, means).

Clausewitz explains this dynamic relationship of purpose, aims and means in war in chapter two of book one. He writes, for example: 'We can now see that in war many roads lead to success, and that they do not all involve the opponent's outright defeat.' Clausewitz summarizes the wide range of routes (Clausewitz 1984: 94) by which one can reach the aim of war, and states that it would be an underrating to think of these differentiations as rare exceptions. After having explained some of these other strategies besides the destruction of the enemy's armed forces (for example, a number of other strategies are linked to the will of the opponent, not his military forces), which, for Clausewitz, remains 'the first-born son of war', he concludes that all we need to do for the moment is to admit the general possibility of their existence, the possibility of deviating from the basic concept of war under the pressure of particular circumstances (p. 99). But the main conclusion is, he argues, that in war many ways lead to success.

Purpose and Rationality in Clausewitz

If we emphasize the purpose within the total process (or compare different purposes for going to war), we are close to what Max Weber called value rationality of purposes. Although Weber sometimes seems to overemphasize the difference between the rationality of purposes and that of ways, his differentiation is useful in shedding light on Clausewitz's theory. Rationality of aims (ways, goals), in contrast, is a principle of action exclusively oriented to work towards a particular goal, presupposing the most effective means and rational consideration of possible consequences and side effects. Value rationale, as compared with the former, is legitimized only by conscious belief in an overarching purpose, and is relatively independent of the standards of success. Value rationality is primarily about the relationship of

different purposes with one another, and their classification into a hierarchy. The dependence of warfare to the shaping of international order, as Clausewitz puts it, is 'value-rational' as defined by Max Weber. If, in contrast, we place the aim within the war in the first place, then it results in process or action rationality. In the means rationality, the given means may be examined to see whether they correspond to a given purpose or not.

Clausewitz differentiates the purpose (ends), aims (goals, ways) and means rationality in a value hierarchy of the purposes, a process rationality of the goals and a means rationality. He makes this distinction at times only implicitly on the basis of different connotations of the concept of purpose. He uses the same concept of terms throughout, providing various contexts from which distinctions may be deduced. In his earlier research, however, Clausewitz differentiated the purpose-aims-means rationality explicitly in this sense: 'As a result each war is raised as an independent whole, whose entity lies in the last purpose whose diversity lies in the available means, and whose art therein exists, to connect both through a range of secondary and associated actions in the shortest way' (Clausewitz 1990: 64). The 'last purpose' of war is bound to the hierarchy of values, the connection of actions 'in the shortest way' can be understood as process rationality, and the 'diversity' of the means indicates the rationality, in which the means are examined to determine if they meet the given goals and purposes.

Additionally, one has to take into account the counter-actions of the opponent. Clausewitz emphasizes this difference in his chapter about the theory of war in book two: 'The essential difference is that war is not an exercise of the will directed at inanimate matter, as in the case with the mechanical arts, or at matter which is animate but yielding, as in the case with the human mind and emotions in the fine art. In war, the will is directed at an animate object that reacts' (Clausewitz 1984: 149). Hence, Clausewitz's final achievement is not a 'cookbook' strategy which could be applied to all kinds of war, but reflections on the art of warfare, the performance of warfare within a political purpose.

If, as it seems for Clausewitz, the purpose of war lies outside warfare, and war is determined only as means for this purpose, then a technical, instrumental understanding of the war is thereby intended. But this is not the whole of Clausewitz: he also examines praxis, performance and practical knowledge. If the purpose lies within warfare, this doesn't contain a full identity of the goal of martial action with its execution. In this case too, the purpose is not war for war's sake. Ernst Vollrath asserts, with regard to Aristotle, that the works of genuine practice are not achieved because of the goal, which is their execution. No one wages a war for war's sake, but whoever does this is thoroughly bloodthirsty (Vollrath 1984: 14). My thesis is that Clausewitz is trying to combine the Aristotelian difference between poiesis and praxis in his writings – an instrumental view of war for political purposes with the performance of the conduct of war, not just with the execution of the political will. If we would understand war just as the execution of the political will, this last approach would in fact cancel the dynamic relationship between the purpose and the conduct of war.

Further Dimensions of the Concept of Purpose in Clausewitz

In Clausewitz's distinction between an existential and instrumental view of war, Herfried Münkler emphasized that war is an instrument in both cases. The distinction does not refer thus to the instrumentality as such. In fact, certain actions are distinguished in contrast to others as suitable for the purpose and are instrumental, while others are not apparently comparable in their suitability for purpose and their instrumentality. Clausewitz summarized the difference between both forms of purpose as follows: 'Where there is a choice of purpose, one may consider and note the means, and where only one purpose may be, the available means are the right ones' (Clausewitz 1894: 76). For Clausewitz, the relationship of purpose, aims and means is bound to the prerequisite that one has a choice between different purposes, an approach, which we introduced as a value rationality of purpose, following Max Weber. If, on the other hand, there is only one purpose, the preservation of an individuality such as the political and physical existence of a society for example, the

purpose-aims-means relationship applies in quite another way. In such a case the crucial point is only the effectiveness of the process. A pure process rationality can lead to the fact that the means of warfare become the end in themselves. This understanding of the instrumentality of war therefore is one of the sharpest criticisms concerning (the early) Clausewitz. He adopted the Napoleonic model from Jena, trying to seize its successes systematically and, without considering the social background of France, to generalize it as an abstraction. In his criticism of Clausewitz, Keegan said that the military develops war cultures which correspond with their social environment. If, however, the war is seen as purely instrumental and the connection to this environment is cut, then the danger of blurring the military boundaries is potential endless violence. In this view the roots of Clausewitz's image of war refer back to the origins of the modern age: full possession of civil rights, the general right to vote and compulsory military service completed the portrait of the citizen soldier and the 'battle scenes' of the people's army.

The French model was in fact somewhat adapted for Prussian circumstances: a revolutionary people's army in the service of the national interest – but without 'republic' (meaning a civil and democratically constituted system of government). In this form Clausewitz's theory was proved, and began to be used later for multiple purposes, starting its triumphant advance through the general staff and war ministries. In Keegan's view the result of this process was the general armament of Europe during the nineteenth century and its excessive increase in the twentieth century. Keegan left unmentioned the fact that Clausewitz's theory of war had yet to be instrumentalized to fulfil this function, especially by the German general staff during the First World War. Nevertheless his criticism revealed a fundamental problem of modern war: the separation of potential options for warfare from socially meaningful purposes.

Historical Development of Clausewitz's Definition of Purpose

While the inconsistent definition of purpose made by the early Clausewitz was essentially based on a pure process rationality, merely action rationality within warfare, the definition of purpose of the later Clausewitz is based on the prerequisite political purpose outside warfare. There are still passages in the final version of *On War* in which Clausewitz does not differentiate clearly between purpose and aims. His uncertainty regarding a concrete definition of both was already expressed in the fact that between his last draft and the published text the first two chapters underwent significant revisions. In the last draft the first chapter was titled 'Purpose of war' and the second 'Means'; in the final text the second chapter was headed 'Purpose and means' while the first was headed with the question 'What is war?' This apparently small difference is symptomatic (Clausewitz 1984: 630 and 636).

Purpose, Rationality and an Instrumental View of War

To grasp the meaning of this development we need to have a look at Clausewitz's historical analyses. Since the French Revolution and since Napoleon, war has been fundamentally changed. Revolutionizing war cost millions of dead, and brought the whole planet in the era of atomic deterrence to the edge of self-destruction. While Clausewitz was sorrowful early on, but publicized what he saw as a necessary adjustment to this new form, to 'modern warfare', he expressed a different emphasis later. For late Clausewitz, the limitation of force as well as the orientation of warfare at the later formation of international peace were in the foreground.

For a whole range of critics 'modern warfare' was essentially caused by the French Revolution, a new form of war, which seemed to find in Clausewitz its most important theoretician. This point of view does not just ignore the difference between the concepts of adjustment and limitation in early and late Clausewitz, but also underestimates the narrowness of the influences of the French Revolution on modern warfare. The revolution was the cause of this development and partially determined some of its aspects – one in particular: the inclusion of ever larger sections of the population.

A further aspect is, however, the industrialization of warfare with ever more advanced technologies. The question to be asked, moreover, is to which of these aspects the historical development of disciplined soldiers in armies should be added: the idealization of the entire population as part of one nation or race, or the cultural process of the adjustment of soldiers to the newly developed weapons. The relationship of the historical level of industrialization of warfare and the human concept of soldiers can be measured by comparison. In about 1800, in the context of the ancient cult, finely formed marble for the representation of the hero body appeared, while in about 1900 the 'steel body from a casting' replaced it: '*Je heißer der Jugend/Stahlfeuerfluß/desto härter formt sich/der männliche Guß*' (The hotter the youth/river of molten steel/the harder the form of the manly casting).

Finally, 'modern warfare' did not only have a delimiting effect but also a limiting one. Modern civil law, conventions of war and the outlawing of certain weapon systems, The Hague Naval Warfare and Land Warfare Convention as well as the Law of Neutrality, are all part of the 'modern war', as is its undeniable delimitation.

In the view of Clausewitz's critics, his instrumental view of war is the theoretical expression of 'modern warfare'. In this view the war is an instrument, a means, with which military aims are to be achieved in favour of any definable superordinate purpose. For them, war for Clausewitz seems merged into a basic purpose-aims-means relationship, which is determined by purely rational considerations. Since the instrumental view of war is located in the centre of the writings of Clausewitz, it is located also in the centre of the criticism. Nevertheless, it draws attention to the fact that there are criticisms from completely opposite points of view. Thus the instrumentality of war is made responsible with Clausewitz, on the one hand, for the delimitation of war and force in 'modern' wars (by John Keegan among others). Clausewitz's political theory of the war appears here as a 'ready-made philosophy of military extremism', as Keegan is indicting Clausewitz. On the other hand, it is postulated that instrumental and purpose-rational warfare is limited by definition – limited by the relation of purpose, aims and means (among others by Martin van Creveld). Clausewitz's instrumental view of war appears in this criticism as strategy, with which the wars of the future could not have successful outcomes.

In my interpretation, Clausewitz's instrumental view of war could be understood as delimiting only if it were replaced from its implicit conditions: the acknowledgement of the opponent and of an international state system as well as the existence of war conventions and social conditions, all of which are opposed to an escalation to the extreme. These pre-requisites are mostly not explicitly indicated by Clausewitz, and may only be deduced from the historical background.

As has been shown, Clausewitz points out that there is not a complete delimitation in war, because the organization of the international peace after the war retroacts 'by the calculation' on warfare (Clausewitz 1984: 78–9). A pure, abstract calculus, which orients itself exclusively at the aims of the war, would, in contrast to this, have a fully delimiting effect. Such a reduction of warfare by mathematical calculation, however, would halve the theoretical approach of Clausewitz and would remove its pre-requisites. Clausewitz's theory is not by any means about an abstract mathematical calculation but a calculus, which after the weapons are silenced already includes the forming of international peace into the conduct of war.

'War is more than a Chameleon' begins Clausewitz in his 'Resultat für die Theorie' (Results for the Theory), the last paragraph of the first chapter, the conclusion of his whole work. This chapter is the only one which he could revise shortly before he returned to military service, during which he died. War is more than a chameleon because it is a trinity of different tendencies, says Clausewitz. These three tendencies are the primordial violence of war, the 'free soul activity' of the commander and his army as well as the subordinated nature of the war as a political tool, whereby it is transmuted into pure intellect. These three are consequently attributed to differing rationalities. As Clausewitz stated for the three tendencies, the following applies: a theory, in which one of these tendencies would be ignored or an arbitrary relationship would be created between them, thereby bringing the theory into conflict with

reality, so that one or more of these tendencies must therefore be considered 'as destroyed'. In contrast to such an approach, Clausewitz argues that a theory must maintain all the tendencies, in a floating balance between three points of attraction and repulsion (Clausewitz 1984: 89). In this 'trinity', Clausewitz consistently attempts to reconcile their various rationalities and to use their conflict as the basis of his political theory of war.

The absolutization of one of those rationalities has led to catastrophe in the history of warfare – be it the warfare of Napoleon, the mechanized war of the First World War, or the Blitzkrieg of the Second World War. We have barely escaped the conceivable planetary catastrophe of an absolutization of deterrence with weapons of mass destruction. In a bipolar world, nevertheless, it proved successful, and a worldwide 'superpowers' weapon monopoly. Whether this deterrent is rational in a multipolar world, and whether the increase in nuclear powers will likewise prevent war, is still moot.

The instrumentality of war in the later Clausewitz is a necessary prerequisite for the limitation of wars; however, it is not sufficient. After the reacknowledgement of the Prussian state, Clausewitz's instrumental view of war is linked with the acknowledgement of the opponent as equal as well as the international peace. Only through these additional assumptions is instrumental warfare able to have a limited effect, in Clausewitz's view.

It could be said that Clausewitz's purpose-aims means rationality opens a new window of opportunity by acknowledging that it involves different rationalities: initially, it incorporates a thorough reasoning of the purposes for which we go to war. Further, it does not imply the disconnection of the purpose and the aims and means in warfare, just the reverse. He emphasizes the difference between poiesis and praxis in warfare as different kinds of instrumentality. Purpose, aims and means are bound within a dynamic relationship, but German history especially has made evident the fact that the primacy of military aims and means over meaningful purposes results in deadly nightmares.

Clausewitz's statements about strategy are to a certain degree context-dependent on the war and the particular situation he is analysing, but this does not mean that there is no general treatment of strategy in his work. To make creative use of Clausewitz's thought, the reader should take the following approach: in those passages in which he is explicitly engaged with strategy, Clausewitz is treating *particular* strategies which could be applied depending on the given situation. His assessments of strategy in general are embodied in his reflections on warfare, and especially on defence and attack. It is essential to bear this methodological approach in mind, because otherwise the result is a futile argument about whether his propositions and recommendations are topical or not. He is not providing a particular doctrine which could be applied in any given situation, as many of his admirers as well as his critics have believed: 'Theory should be study, not doctrine' (Clausewitz 1984: 142). Instead, he is providing an universe of strategy by itself, or, to make a comparison, a toolbox of strategic thoughts. Sometimes it might be the right course of action to use a hammer, but sometimes a pair of tweezers is appropriate. His legacy is not the proposition that we should use a hammer in any given situation, as the early Clausewitz seems to argue with respect to Napoleon's kind of warfare; his legacy is a number of reflections on the problem of what it means to use a hammer or a pair of tweezers and, furthermore, which instrument should be applied in a specific context. In short, for Clausewitz there is no universal strategy which could be applied in all possible circumstances: the ends for which war is being waged determine the appropriate strategy.

Clausewitz's eminent contribution to strategic thought does not consist of a doctrine, a sample of meagre propositions to be applied in any given situation. Just the reverse: Clausewitz is rejecting such an approach and is instead reflecting on the policy and politics of the polities, the warring communities. Additionally, he emphasizes the great difference between strategies of self-preservation and gaining advantages, of any kind, which he introduces with the dialectics of attack and defence: for him, defence is the stronger form of war with a negative purpose, and attack the weaker form with a positive purpose. Clausewitz reveals his methodology in those parts

of *On War* in which he is treating the relation of defence and attack (Herberg-Rothe 2007) Although he favoured only one of these concepts during various stages of his life, his real legacy concerning strategy is the dialectics of attack (gaining) and defence (self-preservation), as well as the interdependency and the reciprocal interaction with the political dimension before, during and after a war.

See Also

- ▶ Military Strategy
- ▶ Strategic Objectives
- ▶ Sun Tzu

References

Herberg-Rothe, A. 2007. *Clausewitz's puzzle: The political theory of war*. Oxford: Oxford University Press.

Vollrath, E. 1984. Neue Wege der Klugheit. Zum methodischen Prinzip der Theorie des Handelns bei Clausewitz. *Zeitschrift für Politik* 1.

Selected Works

1922. Historisch-politische Aufzeichnungen von 1809 [Historical-political records of 1809]. In *Politische Schriften* [Political Literature], ed. H. Rothfels. Munich: Drei Masken.

1984. *On war*, trans. and ed. P. Paret and M. Howard. Princeton: Princeton University Press.

1990. *Schriften, Aufsätze, Studien, Briefe*, vol. 3, ed. W. Hahlweg. Göttingen: Vandenhoeck & Ruprecht.

1992. *Historical and political writings*, ed. P. Paret and D. Moran, 1992. Princeton: Princeton University Press.

Closed Vs Open Innovation

Andreas Panagopoulos
University of Crete, Rethymnon, Greece

Abstract

Whereas in closed ▶ innovation the firm creates, incubates and develops the needed technologies in lieu of external support and through in-house R&D, in ▶ open innovation the firm stops being the powerhouse of innovative activity, increasingly acting as the coordinator of R&D efforts carried out-of-house through start-ups and spin-offs.

Definition Closed innovation describes how firms self-develop their R&D competencies. By contrast, open innovation illustrates how firms utilize third-party R&D capabilities that extend beyond their core competencies and try to offshoot technologies that fall outside the domain of their business model.

The notion of ▶ open innovation was first advanced by Henry Chesbrough in 2003 in his book *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Open innovation acknowledges the new industry trends, where start-ups and spin-offs are increasingly credited with developing innovations. Accordingly, the firm stops being the powerhouse of innovative activity, acting instead as the coordinator of efforts carried out not in-house, as the traditional model of ▶ innovation (closed innovation) assumes, but out-of-house, through start-ups and spin-offs. Therefore, the firm is able to utilize an array of capabilities that frequently extend beyond its core competencies. Moreover, through the use of spin-off companies firms are able to market products that fall outside the domain of the firm's business model, which further enhances its capacity to profit from technologies that would have hitherto remained unexploited. Frequent examples of firms resting on open innovation include Intel, IBM, Proctor & Gamble and Sun Microsystems.

Specifically, in Chesbrough et al. (2008: 1), Chesbrough defines open innovation as:

> the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation, respectively. Open innovation is a paradigm that assumes that firms can and should use external ideas as well as internal ideas, and internal and external paths to market, as they look to advance their technology.

Contrasting with open innovation, closed innovation describes the paradigm behind innovative activity that predominated throughout most of the twentieth century. To offer an example, as Alfred

D. Chandler has chronicled (Chandler, 1990), American industrial firms (such as GM and DuPont) emerged in the first half of the twentieth century through an integrated value chain that linked R&D, manufacturing and distribution. In doing so they followed in the footsteps of the nineteenth-century German chemical industry, which fashioned science laboratories that worked alongside the firm in providing the needed technology. In this respect, firms work within their boundaries in order to have control over the ideas they employ in producing novel products and processes. Subsequently, in closed innovation, the aim of the firm is to self-develop the competencies needed in order to become a market leader.

Obviously, such a system allows the company to form an *impregnable castle* that can produce technologies independently of the scientific community. Thus, the firm does not need to wait for the needed technologies to become available, since it can instead develop them on its own. The obvious advantage of closed innovation is the speed through which a firm can incorporate its own technology into a final product, unhindered by exogenous factors such as ▶ licensing negotiations, multiple owners and users of a technology, and the legal disputes that can arise, sometimes unexpectedly, when the firm is not the singular owner and developer of its technology.

An apparent drawback of closed innovation is that it requires its R&D efforts to be focused on the specific technologies that the firm perceives to be needed. Consequently, it comes as no surprise that, due to insufficient resources and the inherent inability to manage pet projects for which the firm fails to find sufficient value added, closed innovation can produce a number of valuable yet unwanted by-products that do not fit inside the firm's business model. Consider, for example, the well-noted historic paradigm of Fairchild Semiconductor, which led to a series of spin-offs and start-ups that became famous along the way through products that developed into household names without Fairchild Semiconductor being able to profit from them; see Berlin (2001).

Fairchild Semiconductor was itself a spin-off from Shockley Semiconductor Laboratory, as, in 1957, a group of talented engineers left Shockley to form Fairchild Semiconductor. Even though the birth of Fairchild Semiconductor owed more to William Shockley's autocratic management style than to new ideas that could not be fully developed within Shockley Semiconductor, the same is not true for the firms that spawned out of Fairchild Semiconductor. A list of the semiconductor firms that were created by Fairchild's own staff is indicative of the engineering drainage that firms can suffer when they cannot generate a profit from the pet projects of their employees, acting in turn as nothing more than entrepreneurial "academies" for new start-ups and spin-off firms. The list includes (in alphabetic order) among others: Advanced Micro Systems (AMD), Cirus Logic, Intel, LSI and National Semiconductor, all key Silicon Valley players.

Xerox, one of the most innovative firms of the second part of the twentieth century, is perhaps the firm that outperformed all others in giving away future market-leading technologies such as the laser printer, windows, and the portable document format (PDF). For example, the inability of Xerox to pursue and further develop the PostScript technologies, proposed by John Warnock (co-founder of Adobe Systems, and an employee of Xerox's Palo Alto research centre until 1982), that eventually led to the now-familiar PDF, speaks volumes about the restrictions that firms face when a new technology (albeit outside the realm of the company's business model) lands in their lap. In this case, John Warnock, who was unable to convince the management of Xerox to commercialize the InterPress graphics language for controlling printing, had to leave Xerox in 1982 (together with Charles Geschke – also a co-founder of Adobe Systems) and start a new firm, Adobe. In turn, it was Adobe that developed the equivalent PostScript technology, bringing it to the market in 1984 as an application for Apple's LaserWriter.

According to Chesbrough, the following factors have contributed to the current erosion of the closed innovation paradigm: (a) the increase in the mobility of skilled workers, (b) the availability of venture capital, which can aid the funding and managing of start-ups and spin-offs, enhancing their ability to develop the competencies that larger firms need, and (c) the existence of an

array of options (e.g., the formation of spin-off firms) for the use of technologies that cannot be further captured by their current developer.

The message stemming from points (a) to (c) and the examples of Xerox and Fairchild Semiconductor are too important to pass by, as they indicate that knowledge is no longer proprietary to the company. From this it follows that knowledge dwells in actors who are either exogenous to the firm (such as suppliers, customers, competitors and universities) or, at best, are not under the full control of the firm (e.g., its employees).

Furthermore, points (a) to (c) inevitably point to the main culprit behind the prominence of the open innovation model, which is the ever-present availability of ▶ outsourcing firms that can provide the needed technology.

If the culprit is to be found in outsourcing firms, the following question is of merit: Why are such firms (mostly start-ups and spin-off firms) better at providing the needed technologies than the individual firm, who, after all, knows exactly what it's aiming for? Explanations based on: (i) incentives, (ii) inability to innovate and (iii) choosing risky projects have been proposed in addressing the above question. Let us briefly elaborate on these issues.

Even though large firms are well versed in providing appropriate incentives to their staff, the incentives on offer frequently fail to compare to the ones that start-ups provide. Consider, for example, the incentives that large firms offer to R&D personnel wishing to pursue unorthodox and unpredictable discoveries. As Josh Lerner (2009: 48) humorously notes "large firms are notorious for offering employees little more than a gold watch for major discoveries." The contrast with the stock option compensations that start-ups offer could not be starker.

Large firms equally get a bad rating in their ability to innovate beyond their single-minded world, which frequently focuses on existing and well-developed customer relations. Consequently, blind spots are the norm rather than the exception. Therefore, the ability of start-ups to identify and exploit new market opportunities comes as a surprise to no one.

Lastly, new firms are more likely to pursue high-risk strategies than established incumbents. The reason behind this divergence, as economic theory suggests, rests in the fact that new entrants have little to lose if they fail, compared with their much larger and established counterparts who stand to lose quite a lot from failure. Subsequently, while start-up firms are frequently "shooting stars" that only shine for a very limited time period and eventually fail and exit the market, when they succeed they can be more effective in creating new technologies.

It has to be said that the above differences in competencies between large and small firms would not come as a surprise to older generations. In fact, it has been argued that open innovation is far from being a new concept, and, as Mowery (2009) suggested, closed innovation might have been the exception in a history characterized mostly by open innovation practices; in fact, it can be said that in the past open innovation was the norm. An example frequently cited is Allen's (1983) discussion of the English nineteenth-century iron production industry. However, what made Chesbrough's early 2000 works so attractive for both scholars and practitioners is that Chesbrough assigned a single term to a collection of developments. As Huizingh (2011) notes, by giving it a label, open innovation acquired a face, and the stream of studies that followed Chesbrough's original contribution gave it a body too. In this respect, open innovation has become an umbrella that encompasses, connects and integrates a range of already existing activities. As von Hippel has argued, many firms have in the past successfully found ideas for commercially important innovations outside the firm. An example of such a related activity is user innovation, which focuses on buyer innovation, particularly at the individual level as in the open source software; see von Hippel (2005).

Examples

The example of Millennium Pharmaceuticals has been frequently employed in outlining the merits behind the use of open innovation. Millennium supplies information and analysis of biological compounds useful to large pharmaceutical firms in drug development. In the past, similar firms

acted as R&D contractors for larger pharmaceutical firms, with the pharmaceutical firms owning the resulting technology. This model had two obvious drawbacks as, (1) the potential value of the technology generated by the contractor is not fully uncovered because it is confined to the specific market of the pharmaceutical firm on whose behalf the R&D is carried out, and (2) such a pay-per-service restricts the growth potential of the contracting firm because of its inability to utilize economies of scale.

Faced with such prospects, Millennium Pharmaceuticals developed a business model that allowed it to retain the ownership of the technology that it developed, by licensing (instead of selling) it to larger pharmaceuticals. In this fashion, the full potential value of Millennium's technology could be captured as it was not structured and purposely developed for use in the core business of one firm. In fact, it would seem that a win-win situation resulted from this strategy, allowing both Millennium and the pharmaceuticals to benefit because Millennium retained ownership of its technology (allowing the firm to generate additional revenue by licensing it to other firms for use in other markets), and the large pharmaceuticals acquired the needed technology at a fraction of the cost they would have paid had they obtained complete ownership over it.

See Also

- ▶ Innovation
- ▶ Licensing
- ▶ Management of Technology
- ▶ Open Innovation
- ▶ Outsourcing
- ▶ Research and Development (R&D) Organization
- ▶ Technology Transfer

References

Allen, R. 1983. Collective invention. *Journal of Economic Behaviour & Organization* 4: 1–24.

Berlin, L. 2001. Robert Noyce and Fairchild semiconductor, 1957–1968. *Business History Review* 75: 63–101.

Chandler, A. 1990. *Scale and scope: The dynamics of industrial capitalism*. Cambridge, MA: Belknap.

Chesbrough, H. 2003. *Open innovation: The new imperative for creating and profiting from technology*. Boston: Harvard Business School Press.

Chesbrough, H., W. Vanhaverbeke, and J. West. 2008. *Open innovation: Researching a new paradigm*. Oxford: Oxford University Press.

Huizingh, E. 2011. Open innovation: State of the art and future perspectives. *Technovation* 31: 2–9.

Lerner, J. 2009. *Boulevard of broken dreams: Why public efforts to boost entrepreneurship have failed – And what to do about it*. Princeton: Princeton University Press.

Mowery, D. 2009. Plus ca change: Industrial R&D in the "third industrial revolution". *Industrial and Corporate Change* 18: 1–50.

Von Hippel, E. 2005. *Democratizing innovation*. Cambridge, MA: The MIT Press.

C

Clusters and Foreign Direct Investment (FDI)

Nigel Driffield[1], Yama Temouri[1] and Stefano Menghinello[2]

[1]Aston University, Business School, Birmingham, West Midlands, UK

[2]Istituto nazionale di statistica, Rome, Italy

Abstract

Multinational enterprises (MNEs) agglomerate in specific locations not only to reduce transactions costs and benefit from other foreign companies' experience, but also to exploit specialized local assets, such as a skilled labour market, specialized suppliers and access to new knowledge. The proximity to research facilities and/or spatial agglomeration of production plants can potentially generate spatially bounded knowledge spillovers and thus act as a strong catalyst to attract foreign direct investment (FDI). This is linked to the international business (IB) literature on the determinants of MNE location choice, and in particular to knowledge sourcing as a leading motivation for FDI targeted to well-known business clusters.

Definition Foreign direct investment (FDI) is an international capital flow designed to generate both ownership and strategic control of income-generating assets. A business cluster is a particular form of spatial agglomeration of activity, characterized by intense interaction and cooperation between co-located firms. Localized spillovers are the positive externalities in either productivity or output generated by interaction of these co-located firms.

The purpose of this contribution is to examine the relationship between the agglomeration of economic activity, in the form of business clusters, and ▶ foreign direct investment (FDI). We seek to make a key distinction that is often rather blurred in the literature, which is the relationship between the location of internationally mobile capital and pre-existing clusters. In doing this, we draw on earlier work, published as De Propris et al. (2005) and Menghinello et al. (2010). Subsequently, we briefly examine the importance of interactions between inward investors and local firms in the context of these agglomerations, and what the implications for strategic management are, building on some widely cited literature concerned with spillovers and FDI.

These issues are best examined through the lens of international business (IB) scholarship, concerning the importance of location for MNEs. Where to locate internationally mobile activities and how to benefit from FDI in specific areas are also crucial strategic management decisions. Equally, policymakers seek to influence these decisions, based on their local development agenda, and the perceived benefits of attracting inward investment to a given region (De Propris and Driffield 2006). This is particularly important in the context of capital being internationally mobile, but (less skilled) labour being relatively immobile, particularly in lagging regions. This issue is complicated by the fact that individual production processes often do not exist in isolation but as part of local, national and global supply chains, such that both location decisions, and the likely benefits of that investment to the local economy depend on the nature of these supply links, both internal and external (Driffield 2001; Driffield et al. 2004; Menghinello et al. 2010).

It is well understood that certain industries have a tendency to cluster in particular locations, gaining from/engendering agglomeration economies. These externalities are typically categorized as either technological or pecuniary, according to the mode by which the external benefits are appropriated. Pecuniary externalities related to the presence of a specialized labour market or connected to forward and backward linkages generated by the local market for intermediate goods, are considered to be more sensitive to geographical distance than technological externalities. In contrast to this view, established contributions to the economic literature also emphasize the key role of localized knowledge externalities as a primary source of firm's ▶ competitiveness (Maskell and Malmberg 1999). This is linked to an in-depth analysis of the local and global components in the process of knowledge creation (Antonelli 2000).

Recent work by Pitelis (2012) offers a governance-based perspective on clusters. Pitelis draws on several theories, such as transaction cost, resource-knowledge-capabilities and perspectives from industrial organization to propose a theoretical framework which highlights an entrepreneurial co-evolutionary theory of clusters. This framework compares clusters with alternative forms of economic organization and systems in order to explain the emergence, evolution and co-evolution of clusters. Rather than viewing clusters as having merely absolute advantages in co-location, he suggests that clusters may offer significant appropriation of co-created value which entrepreneurs support and help in creating, as long as value-appropriation in clusters is higher than in alternative economic organizations, such as markets, hierarchies and other inter-firm cooperation.

The literature on ▶ regional development and externalities comprises both theoretical and empirical contributions. Grossman and Helpman (1991) demonstrate that geographically bounded knowledge externalities can generate a self-reinforcing process that supports the creation of enduring dynamic comparative advantages,

regardless of a country's or region's relative factor endowments. This is consistent with Markusen and Venables (1999), and the voluminous applied literature that has followed it, focusing on the relationship between external effects and the process of knowledge creation, such as, for example, Haskel et al. (2007) and Smarzynska-Javorcik (2004).

The IB literature, pioneered by the works of Hymer (1976), Vernon (1966) and Dunning (1977), focuses on the determinants of MNE behaviour. This stream of literature links FDI motivations to MNE location choice. This has become a basis for the literature concerned with the analysis of spillovers, focusing on the interaction between foreign affiliates and domestic companies. In turn, the importance of interactions and linkages locally highlight the importance of clusters in this process.

FDI Location in a Business Cluster: Expected Benefits and Potential Threats

Local business clusters – that is, geographic concentrations of interconnected companies, specialized suppliers, service providers and associated institutions – have received increasing attention from academics and policymakers because, rather than wiping out the influence of space, firms in the globalized knowledge economy are relying more and more on their local environment for aspects of their competitiveness, while innovation and entrepreneurship activity is significantly concentrated across space (Potter and Miranda 2009). The analysis of the peculiar characteristics of a business cluster is intrinsically linked to the conceptual debate on the nature of business clusters. The current literature is dominated by a large number of highly heterogeneous definitions. Among the most commonly used definitions of business clusters, there are those developed by Porter (1990, 1998), Rosenfeld (1997), the OECD (1999) and Becattini (1990). In addition, contributions made by Markusen (1996) and Gordon and McCann (2000) are of particular interest, since these authors attempt to define an exhaustive taxonomy of different forms of spatial clustering.

From a strategic management perspective, a further development was made by Michael Porter (1990, 1998), who referred to factor conditions, demand conditions, related industries and inter-firm rivalry as the drivers of growth in clusters, which favour innovation, competitiveness, and productivity gains at the local level. Around this seminal work has developed an extensive literature, primarily based on case studies, which has discussed at length the internal dynamics and external relationships behind successful clusters around the world (Schmitz and Nadvi 1999; Giuliani et al. 2005). Much of the evidence is, however, of anecdotal nature, generally explaining success by some key factor, whether this is vertical or horizontal cooperation amongst firms, government support, industry–university relationships and so on. Quantitative evidence about the performance of business clusters is more limited in the literature (Temouri 2012). How do clusters compare in terms of employment and turnover? What clusters are on the rise and what others are on the decline? Does being part of a cluster help firms to keep a steadier performance?

The debate on the conceptualization of business cluster is not purely academic since it is well known that local benefits arising from different local contexts vary widely both in scope and magnitude. Indeed, all existing definitions mirror dissimilar conceptualizations of what a business cluster really represents. Three different overall approaches can be distinguished in this respect.

First, the so-called spatial approach has its roots in the German School on optimal location choice (von Thünen 1826; Weber 1962) and can be linked to more recent contributions such as the New Economic Geography pioneered by Krugman (1991). According to this approach, a business cluster has a purely spatial dimension and therefore can be applied to any forms of firm agglomeration while only pecuniary externalities, essentially based on the reduction of transaction costs due to firm proximity, are expected to arise from firms located in these areas.

Second, the geographical approach encompasses a wide range of different contributions in the

literature, such as the growth pole approach by Perroux (1950) and the innovative milieu (Aydalot 1986; Camagni 1991) as well as transaction costs and agglomeration approaches (Scott and Storper 1987). All these contributions, sometimes very different in their theoretical assumptions, recognize the unique nature of a business cluster as a peculiar place characterized by a dense interaction between a wide range of local actors (firms, institutions, universities) that build upon a stock of knowledge deeply influenced by local, specific competences.

Third, the so-called *industrial economics perspective* has its roots in the pioneering contribution made by Marshall (1890, 1919) and further extended by Becattini (1979, 1990) and Piore and Sabel (1984). Alfred Marshall identified a market for intermediate inputs, a skilled labour force and technology spillovers as the three key sources of externalities that benefit firms working close to each other in related industries. The concept was subsequently adapted to industrial districts, and used to analyse the significant growth of certain sectors of the Italian economy in the post-Second World War period. This attributed such growth to a model of production resting on 'flexible specialization', where each small firm would specialize in a specific input and cooperate with others in the same locality to deliver a final product of quality to international markets (Piore and Sabel 1984). This approach emphasizes the presence of production organization and local governance, as well as social and economic embeddedness in local firms' behaviour, as key factors that can discriminate peculiar forms of business clusters from other, simpler forms of firms' agglomeration. This approach also emphasizes the role of cooperation across local firms and the presence of local institutions and local values as factors that can discriminate these peculiar contexts from other forms of firm agglomeration. A wide range of potential benefits, including knowledge-related dynamic externalities, are associated to this peculiar form of firm spatial agglomeration.

From a strategic management perspective, the delination of the business cluster definition according to different approaches highlights the fact that very variable benefits can be derived from firms operating in dissimilar business clusters, whereas business clusters' characteristics play a crucial role in assessing the scope and magnitude of expected externality effects.

Another relevant issue for strategic management concerns the circumstances under which a foreign affiliate can benefit the most from the location in a business cluster and minimize, at the same time, sources of potential threats. The literature clearly shows that a certain degree of local embeddedness, usually achieved by balancing competition with cooperative behaviour in both the production and innovation-related activities, usually minimizes threats and substantially increases the chances for the MNE to gain the most from location advantages. Foreign affiliates embedded in business clusters are usually considered to be more embedded into the host country than footloose ones, since they are more deeply engaged in cooperative and long lasting relationships with indigenous firms. Crone and Roper (2001) show that the magnitude of spillovers depends on the strength of backward linkages with local suppliers. Gereffi (1989, 1994) shows that most FDI into developing countries presents a poor degree of local embeddedness, since it is exclusively devoted to exploit location advantages in terms of labour costs. As a result, footloose FDIs are very likely to realize 'enclave economies' with limited impact on the host economy and a high probability of leaving the country in the short–medium term. Bellandi (2001) represents one of the few theoretical contributions in the economic literature that explicitly focuses on the role of large firms, including MNEs, within business clusters. He stresses the relevance of 'embedded' behaviour as the key conduct for the MNE to access local tacit knowledge. In contrast, predatory behaviour is recognized as a threat by local firms, which may react by obstructing the share of local-based externalities.

Much of this literature is based on the seminal work of Krugman (1991) on trade and agglomeration, although this author does not explicitly refer

to business clusters. The extent to which an outside agent – that is, a foreign firm – can enter this rather organic system and facilitate mutual benefits is then a feature of the precise nature of the system. However, the basic assumption is that clusters are sustained by external economies of scale, and other forms of externality, that render the productivity of an individual firm as higher within the system than it would be outside it. As such, the attraction for MNEs with access to global supply chains is clear, as is their contribution through access to frontier technology. However, much of the subsequent work in economic geography, for example, by Dicken (1998), or more recently Boggs and Rantisi (2003) and Bellandi (2001), highlight the importance of place in terms of these interactions, while the IB literature has focused on issues such as the location of R&D (Cantwell and Iammarino 2000) and the importance of agglomeration more generally in explaining the location of internationally mobile investments, see Basile (2002, 2004), and Devereux and Griffith (1998). Nachum (2000), for example, offers a link between models based on economic geography and IB analysis, by suggesting that FDI can be a force for agglomeration, while this is tested more explicitly in Driffield and Munday (2000, 2001). Pantzalis (2001) demonstrates that the location of foreign subsidiaries can contribute significantly to the value of the parent company, while Zaheer and Manrakhan (2001) address this issue of agglomeration more explicitly. Expected benefits from the location of a foreign affiliate in a business cluster are associated with the presence of a skilled labour market, a pool of specialized suppliers and potential access to localized knowledge assets. Potential threats are related to the presence of a large pool of highly competitive firms specialized in similar market segments or niches.

Some scholars, such as Florida (1997) and Pearce (1999), provide some empirical evidence on the relationship between new trends in FDI direction and changing patterns in MNE behaviour. Florida (1997), by investigating the determinants of growing FDI into the USA, related to research and development activities, found evidence that the technology sourcing motivation is more important with respect to the traditional MNE behaviour to adapt the superior foreign technology to the host country. In particular, he shows that access to skilled and highly qualified human capital represents a dominant motivation for foreign entry. In a similar vein, Pearce (1999), by focusing on the activities of foreign-owned research laboratories in the UK, highlights that these laboratories show a much more intense involvement in original product development and are useful to better exploit location advantages.

Conclusion

In conclusion, the key issue in these related literatures concerns the importance of location or 'place' as an essential determinant of MNE performance. Strategic decisions on geographical location of affiliates abroad and position of the affiliates within the MNE supply chain cannot be divorced from the assessment of the nature of interactions between inward investors and local firms. The interaction between spatial location of affiliates, local or regional institutions and affiliate productivity contribute to parent firm performance. However, one would expect that this would not be independent of physical distance between the cluster and the foreign parent firm.

It should also be stressed that the scale and scope of interactions between foreign affiliates and local companies can also facilitate technology sourcing activity. Arita et al. (2002) show that MNEs have to undertake peculiar location policies, depending on the different typologies of geographical agglomeration, in order to make knowledge sourcing more effective. Andersen and Christensen (2005) more explicitly address the issue of knowledge sourcing from regional innovation systems. They point out that the external coordination mechanisms between foreign affiliates and 'local carriers of knowledge' play a central role in this respect. In particular, they emphasize the differences between MNEs and regional clusters in terms of both organizational settings and knowledge generation, representation and dissemination processes. The differences are

very likely to determine relevant problems when foreign affiliates attempt to assess external knowledge embedded in local contexts. Andersen and Christensen (2005) also highlight a list of conditions that make knowledge sourcing from local contexts easier. These include socialization and task participation, credible commitment, and the presence of complementary and overlapping competencies. Dupuy and Gilly (1999) propose a more integrated approach where complex business organizations, such as large enterprise groups, actively interact with the territories where they are located, shaping the geography of economic activity but also deeply influencing their dynamics. Finally, Rugman and Verbeke (2003) try to encompass the MNE within the 'diamond-model' elaborated by Porter (1990, 1998) for geographical clusters. Within this extended framework, the MNE is simultaneously a knowledge generator and a knowledge seeker.

See Also

- ▶ Competitiveness
- ▶ Cooperation and Competition
- ▶ Competitive Strategy
- ▶ Foreign Direct Investment (FDI)
- ▶ Foreign Direct Investment (FDI) and Economic Development
- ▶ Regional Development

References

Andersen, P.H., and P.R. Christensen. 2005. *From localized to corporate excellence: How do MNCs extract, combine and disseminate sticky knowledge from regional innovation systems?* DRUID Working paper No. 16. Aalborg: Danish Research Unit of Industrial Dynamics.

Antonelli, C. 2000. Collective knowledge communication and innovation: The evidence of technological districts. *Regional Studies* 34: 535–547.

Arita, T., I.R. Gordon, and P. McCann. 2002. Industrial clusters, transactions costs and the institutional determinants of MNE location behaviour. *International Business Review* 11: 647–663.

Aydalot, P. (ed.). 1986. *Milieux innovateurs en Europe*. Paris: GREMI.

Basile, R. 2002. The locational determinants of foreign-owned manufacturing plants in Italy: Preliminary results. *Rivista di Politica Economica* 93: 89–134.

Basile, R. 2004. Acquisition versus Greenfield investment: The location of foreign manufacturers in Italy. *Regional Science and Urban Economics* 34: 3–25.

Becattini, G. 1979. Dal settore industriale al distretto industriale: Alcune considerazioni sull'unità di indagine dell'economia industriale. *Rivista di Economia e Politica Industriale* 1: 56–78.

Becattini, G. 1990. The Marshallian industrial district as a socio-economic notion. In *Industrial districts and inter-firm co-operation in Italy*, ed. G. Becattini, F. Pyke, and W. Sengenberger. Geneva: International Institute for Labour Studies.

Bellandi, M. 2001. Local development and embedded large firms. *Entrepreneurship & Regional Development* 13: 189–210.

Boggs, J.S., and N.M. Rantisi. 2003. The relational turn in economic geography. *Journal of Economic Geography* 3: 109–116.

Camagni, R. 1991. *Innovation networks: Spatial perspectives*. London: Belhaven Press.

Cantwell, J., and S. Iammarino. 2000. Multinational corporations and the location of technological innovation in the UK regions. *Regional Studies* 34: 317–332.

Crone, M., and S. Roper. 2001. Local learning from multinational plants: Knowledge transfers in the supply chain. *Regional Studies* 35: 535–548.

De Propris, N., and N. Driffield. 2006. The importance of clusters for spillovers from foreign direct investment and technology sourcing. *Cambridge Journal of Economics* 30: 277–229.

De Propris, L., N. Driffield, and S. Menghinello. 2005. Local industrial systems and the location of FDI in Italy. *International Journal of the Economics of Business* 12: 105–121.

Devereux, M., and R. Griffith. 1998. Taxes and the location of production: Evidence from a panel of US multinationals. *Journal of Public Economics* 68: 335–367.

Dicken, P. 1998. *Global shifts: Transforming the world economy*. London: Paul Chapman Publishing.

Driffield, N. 2001. The impact on domestic productivity of inward investment in the UK. *Manchester School* 69: 103–119.

Driffield, N., and M. Munday. 2000. Industrial performance, agglomeration, and foreign manufacturing investment in the UK. *Journal of International Business Studies* 31: 21–37.

Driffield, N., and M. Munday. 2001. Foreign manufacturing, regional agglomeration and technical efficiency in UK industries: A stochastic production frontier approach. *Regional Studies* 35: 391–399.

Driffield, N., M. Munday, and A. Roberts. 2004. Inward investment, transaction linkages and productivity spillovers. *Papers in Regional Science* 83: 699–722.

Dunning, J. 1977. Trade, location of economic activity and the MNE: A search for an eclectic approach. In

The international allocation of economic activity: Proceedings of a Nobel symposium held at Stockholm, ed. P. Hesselborn, B. Ohlin, and P. Wijkman. London: Macmillan.

Dupuy, C., and J.-P. Gilly. 1999. Industrial groups and territories: The case of Matra-Marconi-space in Toulouse. *Cambridge Journal of Economics* 23: 207–223.

Florida, R. 1997. The globalization of R&D: Results of a survey of foreign-affiliated laboratories in the USA. *Research Policy* 26: 85–103.

Gereffi, G. 1989. Rethinking development theory: Insights from east Asia and Latin America. *Sociological Forum* 4: 505–533.

Gereffi, G. 1994. Global commodity chains. In *Commodity chains and global capitalism*, ed. G. Gereffi and M. Korzeniewicz. Westport: Greenwood Press.

Giuliani, E., C. Pietrobelli, and R. Rabellotti. 2005. Upgrading in global value chains: Lessons from Latin American clusters. *World Development* 33: 549–573.

Gordon, I.R., and P. McCann. 2000. Industrial clusters: Complexes, agglomeration and/or social networks? *Urban Studies* 37: 513–532.

Grossman, G.M., and E. Helpman. 1991. *Innovation and growth in the global economy*. Cambridge, MA: The MIT Press.

Haskel, J.E., S.C. Pereira, and M.J. Slaughter. 2007. Does inward foreign direct investment boost the productivity of domestic firms? *Review of Economics and Statistics* 89: 482–496.

Hymer, S. 1976. *The international operations of national firms: A study of direct foreign investment*. Cambridge, MA: The MIT Press.

Krugman, P. 1991. *Geography and trade*. Cambridge, MA: The MIT Press.

Markusen, A. 1996. Sticky places in slippery space: A typology of industrial districts. *Economic Geography* 72: 293–313.

Markusen, J.R., and A.J. Venables. 1999. Foreign direct investment as a catalyst for industrial development. *European Economic Review* 43: 335–356.

Marshall, A. 1890. *The principles of economics*. London: Macmillan.

Marshall, A. 1919. *Industry and trade*. London: Macmillan.

Maskell, P., and A. Malmberg. 1999. Localised learning and industrial competitiveness. *Cambridge Journal of Economics* 23: 167–185.

Menghinello, S., L. DePropris, and N. Driffield. 2010. Localised competitive advantages: Local industrial systems in Italy as engines of international trade and economic development. *Journal of Economic Geography* 10: 539–558.

Nachum, L. 2000. Economic geography and the location of TNCs: Financial and professional service FDI into the USA. *Journal of International Business Studies* 32: 367–385.

OECD (Organisation for Economic Co-operation and Development). 1999. *Boosting innovation: The cluster approach*. Paris: OECD.

Pantzalis, C. 2001. Does location matter? An empirical analysis of the geographic scope and MNC market valuation. *Journal of International Business Studies* 32: 133–155.

Pearce, R.D. 1999. Decentralised R&D and strategic competitiveness: Globalised approaches to generation and use of technology in multinational enterprises (MNEs). *Research Policy* 28: 157–178.

Perroux, F. 1950. Economic space: Theory and applications. *Quarterly Journal of Economics* 64: 89–104.

Piore, M.J., and C.F. Sabel. 1984. *The second industrial divide: Possibilities for prosperity*. New York: Basic Books.

Pitelis, C. 2012. Clusters, entrepreneurial ecosystem co-creation, and appropriability: A conceptual framework. *Industrial and Corporate Change* 21: 1359–1388.

Porter, M.E. 1990. *The competitive advantage of nations*. New York: Free Press.

Porter, M.E. 1998. Clusters and the new economics of competition. *Harvard Business Review* 76: 77–90.

Potter, J., and G. Miranda. 2009. *Clusters, innovation and entrepreneurship*. Paris: OECD.

Rosenfeld, S.A. 1997. Bringing business clusters into the mainstream of economic development. *European Planning Studies* 5: 3–23.

Rugman, A.M., and A. Verbeke. 2003. Multinational enterprises and clusters. *Special Issue, Management International Review* 43: 151–169.

Schmitz, H., and K. Nadvi. 1999. Clustering and industrialization. *World Development* 27: 1503–1514.

Scott, A.J., and M. Storper. 1987. High technology industry and regional development: A theoretical critique and reconstruction. *International Social Science Journal* 112: 215–232.

Smarzynska-Javorcik, B. 2004. Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *American Economic Review* 94: 605–627.

Temouri, Y. 2012. *The cluster scoreboard: Measuring the performance of local business clusters in the knowledge economy*. OECD Local Economic and Employment Development (LEED) Working paper No. 2012/13. Paris: OECD. Available at: https://doi.org/10.1787/5k94ghq8p5kden. Accessed 1 Aug 2012.

Vernon, R. 1966. International investment and international trade in the product cycle. *Quarterly Journal of Economics* 80: 190–207.

Von Thünen, J. H. 1826. *Von Thünen's isolated state*. Trans. C. M. Wartenburg [Ger. orig. *Der Isolierte Staaat in Beziehung auf Landtschaft und Nationalokonomie*]. Oxford: Pergamon Press.

Weber, A. [1909] 1962. *Theory of the location of industries* (trans: Friedrich, C.) [Ger. orig. *Über den Standort der Industrien*]. Chicago: University of Chicago Press.

Zaheer, S., and S. Manrakhan. 2001. Concentration and dispersion in global industries: Remote electronic access and the location of economic activities. *Journal of International Business Studies* 32: 667–686.

Coalitions

Ilja van Beest[1] and J. Keith Murnigham[2]
[1]Tilburg University, Tilburg, The Netherlands
[2]Northwestern University, Kellogg School of Management, Evanston, IL, USA

Abstract

Groups often need to reach collective decisions when the group's members do not all agree. In these situations, the subset of a group that agrees on the group's final decision can be considered a winning coalition. Game theory, social psychology and political science have all addressed the dynamics of coalition formation; organizational behaviour has also recognized the importance of coalition formation processes, although it has not utilized the other literatures' advances in theory and research findings to increase the understanding of important strategic processes.

Definition Coalitions include two or more parties who act jointly to affect their own and others' outcomes, often because the parties are in conflict over the allocation of scarce resources but need each other to achieve their goals.

Coalitions include two or more parties who act jointly to affect their own and others' outcomes, often because the parties are in conflict over the allocation of scarce resources but need each other to achieve their goals. Coalitions form throughout and between organizations and societies: teams form, firms merge, joint ventures develop and governments ally. As Thompson (1967: 126) noted, 'coalition behavior is undoubtedly of major importance to our understanding of complex organizations'. Even though coalitions and coalition formation are part of the conceptual foundations of classic theories of organization (e.g., Cyert and March 1963; Pfeffer and Salancik 1978), the field has paid scant attention to the extensive literature on coalitions in social psychology, game theory and political science (Murnighan 1978a, 1994).

Social psychologists have adopted a game-theoretic approach to studying coalitions (Kahan and Rapoport 1984; Komorita 1984). Their models, which assume that parties are all self-interested, have generated a wide array of theoretical predictions; their empirical research has tested these predictions by investigating the interactions of individual players who control different amounts of resources in coalition games in which they seek to increase their own outcomes. The two main questions in this line of research are 'Which players will be included in the coalition?' and 'How will they distribute the outcomes?' In addition to tangible outcomes, being included as a member of a successful coalition can bring significant personal benefits, and being excluded can lead to significant costs. In addition, the tangible benefits for the members of winning coalitions may vary as a function of which of the parties are included in the winning coalition.

As an example, social-psychological research might investigate a five-party interaction in which players A, B, C, D and E hold 10, 7, 5, 3 and 2 votes, respectively, and know that they must control 14 votes to obtain a fixed, total payoff of (a usually fictitious) $100. (Alternatively, the payoff might vary as a function of the coalition's membership.) This structure allows player A to form a two-party coalition with B or C and a three-party coalition with D and E; it also allows B and C to form a three-party coalition with either D or E. Social-psychological and game-theoretic models all acknowledge that the underlying structure of this interaction, as determined by the distribution of the parties' resources, gives the most power to player A and the least to players D and E who, like B and C, are interchangeable as coalition members even though they control different amounts of resources. Theoretical models differ, however, in attributing power to each of the players. Strategically, almost all of the models suggest that coalitions with fewer members tend to be preferred to coalitions with more members, as smaller coalitions give the included players larger portions of the fixed payoff. Thus, the selection of coalition partners and the distribution of the coalition outcomes over its members tend to be strongly related, with the preferred distribution

of coalition outcomes determining who the parties prefer as coalition partners.

Examples of the Theories' Predictions

Minimum resource theory (Caplow 1956; Gamson 1961) pays strong attention to the actual value of the players' resources: it assumes that coalition members will distribute a payoff in proportion to those resources. Thus, in the 10-7-5-3-2 example, it predicts that a 7-5-2 coalition will form and distribute the payoffs 50%–36%–14%. In contrast, *bargaining theory* (Komorita and Chertkoff 1973) predicts that a 10-5 coalition will form because it reduces the size of the coalition and helps to maximize the payoffs of the included players (in comparison to the outcomes they could obtain in other coalitions): it predicts that the outcomes for the ten and five players will range from 58–42% to 65–35%. A third model, *minimum power theory* (Gamson 1964), assumes that the parties will negotiate an agreement on the basis of whether particular parties are truly needed to form a winning coalition: it predicts that a 10-3-2 coalition will form and distribute the payoffs 74%–13%–13%.

A Critique

These theories assume that the parties all agree on the basis for distributing the outcome and that they will smoothly move to form the predicted coalitions. In practice, however, dynamic negotiation processes involve jockeying for position and intense negotiations over inclusion and payoff distributions. Self-interested parties, for instance, tend to argue that the most appropriate factors for determining coalition outcomes and inclusion are those that, not by chance, favour them (Van Beest et al. 2004). Thus, parties with many resources typically argue that resources are all that matter and that outcomes should be distributed in proportion to the resources each member possesses. Parties with fewer resources, in contrast, typically argue that their membership is all that matters and that resources should be distributed equally among all coalition members, regardless of the number of resources that they contribute. (An analogue for these arguments resulted in the formation of the two houses of Congress in the United States, one which gives each state two votes and the other which gives states varying numbers of votes depending on their population.) Thus, the different strategic positions that parties take during coalition bargaining are often predictable.

The dynamics of self-interest and a strong desire to avoid being excluded can lead to a wide variety of outcomes in coalition negotiations (Murnighan 1978b; Van Beest and Van Dijk 2007). Valuing more than one's own outcomes (i.e., social utilities) can also lead to the formation of coalitions that are larger than necessary, as can a desire for solidarity or the fear of attack from those who are ultimately excluded (Van Beest et al. 2003).

The dynamics of self-interest that are an inherent assumption in coalition theories can lead, however, to an interesting phenomenon that is known as the *strength-is-weakness effect*. Indeed, many coalition studies have shown that appearing weak increases a party's attractiveness to other potential partners and, as a result, leads to a higher likelihood of being included in the winning coalition; it can also lead to higher-than-expected payoffs. These results are usually observed in settings in which an apparent advantage or strength is not associated with real power (Kelley and Arrowood 1960; Murnighan 1978b). That is, when power bases vary, strength-is-weakness only results when the parties that hold differential resources are structurally interchangeable. Then, partners who have just enough resources to do the job appear optimal because fewer resources imply smaller outcome demands and greater attractiveness. When parties are not interchangeable, however, true structural strength is extremely valuable and no longer the harbinger of exclusion (i.e., weakness).

Suggestions for Organizations

Organizational founders have the challenge of building an effective, functioning coalition

(Polzer et al. 1998). As a result, effective founders tend to possess a diverse network of weak ties rather than strong links to only a few others, as the coalitional strength of a new organization may rest on infrequent, non-repetitive interactions with many others rather than on frequent, well-established interactions with a few close contacts. Political models (e.g., De Swaan 1973) also suggest that coalitions form incrementally, as interconnected sets of interacting dyads. Put simply, coalitions form one party at a time. Also, once a coalition is successful at achieving a critical mass, continued growth becomes considerably easier.

Being surreptitious during the coalition-formation process can also be critical in the success of an organizational coalition: keeping quiet helps blunt the formation of an organized opposition, that is, counter-coalitions, which von Neumann and Morgenstern's (1947) classic, original model assumed would be the parties' natural reactions to the formation of a coalition. Successful coalitions tend to be fluid: forming quickly, expanding, and then often bursting at the time of decision and quickly disappearing (Murnighan and Brass 1991).

Political models suggest that founders add similar members to ensure their own centrality in the final coalition (De Swaan 1973). This can reduce a coalition's flexibility when it comes to performance, suggesting that new coalition partners might be better selected when they can balance each other's ideologies on either side of a founder's ideological position. This can allow a coalition to grow until it is just large enough to win, while keeping the range of its ideologies at a minimum and increasing the likelihood that its final policy positions will most closely resemble the founder's. This kind of political strategy, which may be well understood by astute organizational tacticians, has not found its way into the organizational literature.

Within organizations, executives who are involved in many productive projects (i.e., organizational coalitions) are viewed as politically powerful. A few organization members may be in several dominant coalitions: they represent Thompson's (1967) concept of *the inner circle*, a select few whose interconnectedness provides them with considerable influence. These are the parties who typically wield considerable coalitional and political influences in organizations.

See Also

- ▶ Alliances
- ▶ Bargaining Models
- ▶ Inter-firm Cooperation
- ▶ Markets as Networks

References

Caplow, T. 1956. A theory of coalitions in the triad. *American Sociological Review* 21: 489–493.

Cyert, R., and J.G. March. 1963. *A behavioral theory of the firm*. Englewood Cliffs: Prentice Hall.

De Swaan, A. 1973. *Coalition theories and cabinet formation*. San Francisco: Jossey-Bass.

Gamson, W.A. 1961. An experimental test of a theory of coalition formation. *American Sociological Review* 26: 565–573.

Gamson, W.A. 1964. Experimental studies of coalition formation. In *Advances in experimental social psychology*, vol. 1, ed. L. Berkowitz. New York: Academic.

Kahan, J.P., and A. Rapoport. 1984. *Theories of coalition formation*. Hillsdale: Lawrence Erlbaum.

Kelley, H.H., and A.J. Arrowood. 1960. Coalitions in the triad: Critique and experiment. *Sociometry* 23: 231–244.

Komorita, S.S. 1984. Coalition bargaining. In *Advances in experimental social psychology*, vol. 18, ed. L. Berkowitz. New York: Academic.

Komorita, S.S., and J.M. Chertkoff. 1973. A bargaining theory of coalition formation. *Psychological Review* 80: 149–162.

Murnighan, J.K. 1978a. Models of coalition behavior: Game theoretic, social psychological, and political perspectives. *Psychological Bulletin* 85: 1130–1153.

Murnighan, J.K. 1978b. Strength and weakness in four coalition situations. *Behavioral Science* 23: 195–208.

Murnighan, J.K. 1994. Game theory and organizational behavior. In *Research in organizational behavior*, ed. B.M. Staw and L.L. Cummings. Greenwich: JAI Press.

Murnighan, J.K., and D.J. Brass. 1991. Intraorganizational coalitions. In *Research on negotiation in organizations*, ed. M. Bazerman, R. Lewicki, and B. Sheppard. Greenwich: JAI Press.

Pfeffer, J., and G. Salancik. 1978. *The external control of organizations*. New York: Harper & Row.

Polzer, J.T., E.A. Mannix, and M.A. Neale. 1998. Interest alignment and coalitions in multiparty negotiation. *Academy of Management Journal* 1: 42–54.
Thompson, J.D. 1967. *Organizations in action*. New York: McGraw-Hill.
Van Beest, I., and E. Van Dijk. 2007. Self-interest and fairness in coalition formation: A social utility approach to understanding partner selection and payoff allocations in groups. *European Review of Social Psychology* 18: 132–174.
Van Beest, I., H. Wilke, and E. Van Dijk. 2003. The excluded player in coalition formation. *Personality and Social Psychology Bulletin* 29: 237–247.
Van Beest, I., E. Van Dijk, and H. Wilke. 2004. Resources and alternatives in coalition formation: The effects on payoff, self-serving behavior, and bargaining length. *European Journal of Social Psychology* 34: 713–728.
Von Neumann, J., and O. Morgenstern. 1947. *The theory of games and economic behavior*. Princeton: Princeton University Press.

Coase, Ronald (Born 1910)

Anoop Madhok
York University, Schulich School of Business, Toronto, ON, Canada

Ronald Coase's work is centred around economic organization and, more specifically, the comparative costs of organizing in firms and markets. In contrast to the prevailing notion of the firm as a production function, he importantly conceptualized firms and markets as ▶ governance structures characterized by authority- and price-based coordination respectively. Yet there is another important though under-appreciated element in his work which focuses on the question of the distribution of economic activity among firms, asking why an activity is organized within one firm and not another. Coase's explanation emphasizes the importance of the interrelatedness of a firm's set of activities. Both these issues are integral to strategic management. It has been almost 75 years since Coase's classic 'The nature of the firm' (NF) was first published in 1937. Coase also published another classic, 'The problem of social cost', in 1960, which was developed around a set of principles akin to those in the 1937 article. However, the former in particular has had a deep influence on academic research in strategy, because it was concerned with economic organization and the boundaries of the firm – a central concern in corporate strategy research – whereas the latter was more centrally concerned with the law. I therefore focus primarily on the contributions of NF.

It is only in the past 25 years or so that the enormous import of NF has been more fully realized in strategic management, and even then arguably insufficiently so. In the article, Coase famously asked the deceptively simple question 'Why are there firms?' This was a marked departure from the then economic orthodoxy and prevailing focus on applied price theory, which was interested in the efficient functioning of markets and which approached the firm as a production function. Coase questioned this assumption and proposed an alternative conceptualization of the firm as a governance structure. He argued that, whereas price movements direct and coordinate production outside the firm, coordination within the firm was accomplished not through prices but through conscious direction by the 'entrepreneur-coordinator'. Within the firm, the employment contract, by which employees agree to follow the directions of the entrepreneur within agreed limits, substitutes for the multiple contracts that characterize market relations and price-based coordination. The firm and the market are therefore alternative methods of coordinating production, 'the distinctive mark of the firm being the supersession of the price mechanism' (Coase 1937: 389).

Coase was perhaps ahead of his time with this radical proposal, since it took a long time for his ideas to take hold and be appreciated. In fact, in his approach to economics and economic organization he was much closer to what is today called strategic management than to the field of economics itself. Perhaps his unconventional background and training (going by the standards of economics at the time) enabled him to approach the phenomenon in the manner that he did. As he himself mentions, in his early years he was more interested in history and science, but, unable to pursue the former and discovering a distaste for mathematics on commencing the latter, he ended up studying commerce (business administration

in today's parlance) by default at the London School of Economics, where he took a wide variety of courses but little economics per se. Towards the end of his studies, simultaneously unconvinced and stimulated by the lack of theoretical explanation on the topic, he became interested in the different ways in which industries were organized, and spent much of his final year visiting businesses and industrial plants in the US, where he was on scholarship. Thus, his ideas were shaped first through broad academic study and second, but equally importantly, through a deep interest in issues relevant to business and extensive interactions with businessmen. This is what makes his work so significant and relevant to business and strategic management scholars.

In NF, by asking why there are firms, he specifically linked organization with cost and put forward the argument that 'the most obvious cost of "organizing" production through the price mechanism is that of discovering what the relevant prices are' (Coase 1937: 390). Although his primary focus was on why firms exist, for theoretical completeness he also asked the corollary: 'Why is all production not carried out by one big firm?' This he attributed to costs associated with 'diminishing returns to management' of a firm. Therefore, to be more precise, his key focus was not on just cost per se but, more specifically, on the comparative costs of organizing across different governance structures. In a series of reflective essays celebrating the 50th anniversary of NF, he reiterated his view that the gains in internal organization come from a reduction in transaction costs (Coase 1988b: 32), where 'whether a transaction would be organized within the firm (integration) or carried out on the market by independent contractors depended on a comparison of the costs' (Coase 1988a: 17). A few decades after the original article, Williamson (1975, 1985) fleshed out Coase's argument more fully through a detailed comparative assessment of alternative governance structures – specifically markets and hierarchies – now known as ▶ transaction cost economics.

However, Coase's contribution to strategic management goes much further than this. What he actually postulated is that 'at the margin the costs of organizing within the firm will be equal *either* to the costs of organizing in another firm *or* to the costs involved in leaving the transaction to be organized by the price mechanism' (Coase 1937: 30; emphasis added). The transaction cost approach, as popularized by Williamson and his adherents, focuses only on the second part of the comparison – vis-à-vis the price mechanism of the market – ignoring the costs of organizing within another firm.

Quite clearly, this other and more neglected comparison is deserving of further attention. The comparison effectively amounts to why an activity is organized within one firm and not another, which appears to be the very basis of the ▶ resource based view of the firm, arguably the theory with the most impact today in the strategic management discipline. In this sense, the field of strategic management owes a greater intellectual debt than has been fully recognized to Coase's original paper and, as I point out below, his subsequent reflections.

Moreover, in additionally raising and addressing the question of why all economic activity is not organized within one large firm, Coase stated that, at a certain point:

> [T]he entrepreneur fails to place the factors of production in the uses where their value is greatest, that is, fails to make the best use of the factors of production. Again, a point must be reached where the loss through the waste of resources is equal to the marketing costs of the exchange transaction in the open market *or* to the loss if the transaction was organized by another entrepreneur … It would appear that the costs of organizing and the losses through mistakes will increase with an increase in … the dissimilarity of the transactions. (Coase 1937: 394, 397; emphasis added)

Once again, the emphasis is on the three-way comparison with the market as well as another firm. Additionally, the argument anticipates the resource-based argument of a firm not straying away from its basic area of competence since it would then be rendered uncompetitive. Coase's argument above has direct implications for diversification, acquisition and other such issues related to firm boundaries that are central to the strategic management discipline.

In reflecting on NF 50 years after it was published, Coase acknowledged in hindsight that the way he presented his original argument had

weaknesses, as a result of which 'a good deal of what is said on the subject seems to be wide off the mark' (Coase 1988c: 36). He notably mentions that what he meant by entrepreneur is 'the hierarchy in a business which directs resources' (Coase 1988b: 31) but the way he elaborated it led to undue attention instead on the employer–employee relationship. He adds:

> For the purpose in Nature of the firm, i.e. why there are firms, my exposition was adequate. But if one is concerned with the further development of the analysis of the firm's activities, the way in which I presented my ideas has, I believe, led to or encouraged an undue emphasis on the role of the firm as a purchaser of the services of factors of production and on the choice of the contractual arrangements which it makes with them. As a consequence of this concentration on the firm as a purchaser of the inputs it uses, economists have tended to neglect the *main* activity of a firm, *running a business*. And this has tended to submerge what is to me the key idea in 'The Nature of the Firm': the comparison of the costs of coordinating the activities of factors of production within the firm with the costs of *bringing about the same result* by market transactions *or* by means of operations undertaken within some other firm. (Coase 1988c: 38, emphasis added)

Two issues are worth noting. First, if the pivotal issue is not just cost but instead the comparative costs *of bringing about the same result*, then Coase's emphasis on cost is 'logically equivalent to resource-based theorists' arguments on competitive advantage in that such advantage is the logical outcome of a superior cost position (of bringing about the same result)' (Madhok 2002: 539). Second, as he himself mentioned in NF, given that the majority of economic activity is carried out by firms, the market for the most part is an abstract representation of other firms. Simon (1991) concurs and terms this as the organizational economy. To underscore this point, Coase elaborated that what he meant in his original paper was that 'what emerges from this interfirm competition . . . is a situation in which, apart from the purchase of the services of factors of production and retail trade, most market transactions will be interfirm exchanges' (Coase 1988c: 40).

If market exchange mostly amounts to exchange between firms, then the division of labour between the firm and the market actually reflects the division of labour between firms. Putting it differently, it has to do with the distribution of economic activity between firms. Importantly, the resource-based view of the firm is fundamentally concerned with the very same issue, even if framed differently in terms of performance differences across firms (Madhok 2002). Thus, although both the research traditions on comparative governance (i.e., transaction cost economics) and competitive advantage (i.e., the resource-based view and related perspectives) have largely tended to go their separate ways, one can see that they are interrelated at an underlying level and, though not explicitly recognized, both share an intellectual heritage originating in Coase's work.

In a general sense, Coase was more interested in understanding better the institutional structure of production as a whole. In later seeking to counter what he considered a overly narrow focus on transaction costs, he asserted that 'the dominant factor determining the institutional structure of production will in general no longer be transaction costs but the relative costs of different firms in organizing particular activities . . . for which it is necessary to uncover the reasons why the cost of organizing particular activities differs among firms' (Coase 1990: 11). However, the kernel of the idea underlying interfirm comparison was incipient in NF, on which he expanded subsequently, stating that:

> [T]he costs of organizing an activity within any given firm depends on what other activities it is engaged in. A given set of activities will facilitate the carrying out of some activities but hinder the performance of others. It is these relationships which undermine the actual organization of production. (Coase 1972: 64)

Once again, two issues are worth noting. First and foremost, the argument shifts the lens on to firms themselves rather than the comparative assessment of firms and markets. In this respect, if the main activity of firms is running a business (rather than deliberating upon why they exist), then this becomes very relevant to scholars and students of strategic management, since running a business successfully is precisely what they are concerned about.

Second, capability considerations are shown to significantly influence boundary decisions. In NF,

by stating that an entrepreneur (firm) can fail to place factors of production to best use, resulting in possible losses through both higher organization costs as well as mistakes, Coase initially introduced – though only limitedly so – the idea of limits to the scope of firms arising from limits to their capabilities relative to other entrepreneurs. His subsequent elaboration underscores that the difference between firms which can undertake a particular activity at the lowest cost relative to others is largely dependent on other activities undertaken by them. This anticipates the argument of path-dependence that underpins much of the capabilities perspective in strategic management, which views the firm as a bundle of resources and capabilities where individual skills, organization and technology are inextricably woven together over time through dynamic and interactive firm-specific processes (Nelson and Winter 1982).

Since Coase's primary focus in NF was on cost, in particular comparative costs of alternate governance structures, he considered it sufficient to merely introduce the argument of limits to firms' scope but did not develop it further. In actuality, as we can see from the above, the argument implicitly shifts the focus away from just cost to firms' skills, capabilities and knowledge. In a somewhat neglected paper, Richardson (1972: 888) emphasized this very issue and claimed that 'we cannot construct an adequate theory of industrial organization and in particular to answer our question about the division of labor between firm and market unless the elements of organization, knowledge, experience, and skill are brought back to the foreground of our vision'.

Besides cost, this also emphasizes productivity benefits tied to skills and knowledge. Similarly, Chandler, a noted historian of industry, argues that 'an understanding of the changing boundaries of the firm required an awareness of the specific capabilities of the firm' (Chandler 1992: 89).

Judging by Coase's reflections, as a result of its very focus, strategic management research is on a promising path that addresses important issues necessary for more completely understanding the institutional structure of production. Besides being a governance structure, each firm is, as Demsetz put it:

> a bundle of commitments to technology, personnel and methods, all contained and constrained by an insulating layer of information that is specific to the firm, and this bundle cannot be altered or imitated easily or quickly. The components of this bundle that are emphasized by TC are important, but not exclusively so. (Demsetz 1988: 148)

It is due to this bundle that one firm is able to deliver a certain outcome more cost-effectively than another.

In a sense, strategic management scholars have been approaching the question of the organization of economic activity from the opposite angle to transaction cost economics scholars. Whereas the latter begin from the default of the market and argue that transactions are organized within firms when markets fail, strategic management scholars begin from firm success wherein a transaction is not organized by a given firm when the firm 'fails', in the sense of securing competitive advantage. These two, respectively, relate to the questions of: (a) why an activity is organized within firms and not purchased through the market, and (b) why particular economic activities are organized within different firms. Yet both these sets of arguments are based on Coase's monumental work(s).

Coase's Dream

In fact, as work in the area progresses, the very notion of a firm is evolving. A firm is many (and all) things – a legal entity, an administrative entity, a bundle of resources (and capabilities and knowledge) – all of which, generally speaking, enable it to transform inputs into outputs in the form of goods and services. However, different conceptualizations of the firm enable scholars to understand it through different lenses. Scholars such as Ghoshal and Moran (1996) even argue that firms are distinct institutions governed by distinct principles and mechanisms that have nothing to do with the market-hierarchy continuum, which they criticize as restricting our understanding of firms. The essential point is that, through collective efforts, scholars may be able to attain a more complete theoretical understanding of what a firm is and what it does (and how).

Coase recently celebrated his 100th birthday. In his reflections a half-century after the original article, he mentioned:

> My dream is to construct a theory which will enable us to analyze the determinants of the institutional structure of production. In NF, the job was only half done – it explained why there were firms but not how the functions which are performed by firms are divided up among them. My dream is to help complete what I started some fifty-five years ago and to take part in the development of such a comprehensive theory. (Coase, 1988c: 47)

He dreamt this almost a quarter of a century ago. With ongoing interest in Coase's work, and with recent developments in strategic management scholarship, I anticipate that strategy and economics scholars together will be well positioned to fulfil Coase's dream within the next quarter-century.

See Also

- ▶ Governance
- ▶ Managerial Resources and Capabilities
- ▶ Resource-Based View
- ▶ Scope of the Enterprise
- ▶ Theory of the Firm
- ▶ Transaction Cost Economics

References

Chandler, A.D. 1992. Organizational capabilities and the economic history of the multinational enterprise. *Journal of Economic Perspectives* 6: 79–100.

Demsetz, H. 1988. The theory of the firm revisited. *Journal of Law, Economics & Organization* 4: 141–161.

Ghoshal, S., and P. Moran. 1996. Bad for practice: A critique of transaction cost theory. *Academy of Management Review* 21: 13–47.

Madhok, A. 2002. Reassessing the fundamentals and beyond. *Strategic Management Journal* 23: 535–550.

Nelson, R., and S. Winter. 1982. *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.

Richardson, J.B. 1972. The organization of industry. *Economic Journal* 82: 883–896.

Simon, H.A. 1991. Organizations and markets. *Journal of Economic Perspectives* 5: 25–44.

Williamson, O.E. 1975. *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.

Williamson, O.E. 1985. *The economic institutions of capitalism*. New York: Free Press.

Selected Works

1937. The nature of the firm. *Economica* 4, 386–405.

1960. The problem of social cost. *Journal of Law and Economics* 3, 1–44.

1972. Industrial organization: A proposal for research. In *Policy Issues and Research Opportunities in Industrial Organization*, ed. V. R. Fuchs. New York: National Bureau of Economic Research.

1988a. The nature of the firm: Origin. *Journal of Law, Economics, & Organization* 4, 3–17.

1988b. The nature of the firm: Meaning. *Journal of Law, Economics, & Organization* 4, 19–32.

1988c. The nature of the firm: Influence. *Journal of Law, Economics, & Organization* 4, 33–48.

1990. Accounting and the theory of the firm. *Journal of Accounting and Economics* 12, 3–13.

C

Cognition and Strategy

Sarah Kaplan
University of Toronto, Rotman School, Toronto, OH, USA

Abstract

A cognitive perspective on strategy suggests that managers' interpretations of the environment shape strategic choices and, therefore, organizational action. This perspective has developed over the past 25 years through various phases, first establishing cognition as a legitimate factor in strategic management (alongside the traditional explanations of capabilities and incentives), second elucidating the causal relationship between cognition and strategic outcomes, and, recently, developing more complex models of the dynamic relationships between cognition, capabilities and incentives. The latest work expands from the firm level of analysis down to the psychological foundations of strategy and up to the impact of interpretive processes at the field level.

Definition A cognitive perspective on strategy suggests that managers' interpretations of the environment shape strategic choices and, therefore, organizational action.

A cognitive perspective on strategy suggests that managers' interpretations of the environment shape strategic choices and, therefore, organizational action. The premise is that interpretive processes at individual, group, organizational and field levels influence strategic outcomes. The interpretive process involves noticing or attending to signals from the environment, interpreting or making sense of that information, and then making choices and taking action based on those interpretations. This short entry, providing a summary of the field and its development, is drawn in part from a longer treatment of cognition and strategy in Kaplan (2011). A more detailed appraisal and a long list of relevant citations to work in the area are provided there.

Origins and 'proof of concept'

Research on cognition and strategy emerged in the 1980s. Scholars have developed this line of thinking for two reasons. First, in attempting to understand the match between the organization and its environment, the cognitive perspective suggests that the environment is not purely exogenous and therefore organizations' responses to the environment are mediated by the interpretations of that environment made by managers. While the underlying attribution for the necessity of interpretation differs – some (behaviouralists) attributing it to the cognitive limits of managers who therefore have to use simplifying heuristics, and others suggesting that it is the inherent (Knightian) uncertainty of the environment that requires managers to make interpretations of the unknown – the effect is to endogenize the environment. Cognitive frames are the means by which managers make sense of the environment, and such ▶ sensemaking shapes strategic choice and action. Second, the cognitive perspective brings attention to the actions of managers in the field of strategic management that has tended to privilege firm-level position and situation. The cognitive approach suggests that structural features do not determine outcomes but rather that organizations only 'act' through the choices and actions of managers within them. These two moves can be seen as a reaction to the rational actor model of utility maximization in explaining strategic choice.

'Cognition' should be seen as an umbrella term underneath which lies many different concepts developed over time. Walsh (1995) and Kaplan (2011) catalogue nearly 100 terms, including attention, frames, mental models, ▶ dominant logic, world view, cognitive maps, information processing and so on. Their use depends on what aspect of the interpretive process is the focus of study and the theoretical tradition on which the authors draw.

The first phase of the development of the field was aimed at creating legitimacy for the line of inquiry and establishing the existence of shared cognitive frames ('proof of concept'). To do so, scholars developed useful measures of cognitive constructs and demonstrated their presence in multiple empirical settings (these approaches were well represented in Huff's 1990 edited volume). One important focus was on identifying cognitive strategic groups in an industry and developing a socially constructed view of rivalry among firms. This approach reversed the causality proposed by economists in suggesting that inter-firm monitoring creates the structure of competition rather than results from competition. Other studies examined different features of the environment from perceptions of industry changes to evaluations of external instability and munificence. Some scholars were concerned with measuring the degree of accuracy or the direction of bias of the frames, while others were focused on understanding the sources of interpretations and how they shape organizational outcomes. Methodologically, scholars mainly used interviews and surveys to develop the cognitive maps of managers and evaluated these data in cross-sectional analyses.

More Recent Developments and Future Research

Once scholars established the proof of concept, interest grew in connecting cognitive frames to strategic outcomes. Doing so implied longitudinal analyses using consistent measures of cognitive constructs over time and across firms. This requirement tested the limits of interviews and surveys as

sources of data, and thus scholars turned to texts such as the letters to shareholders in company annual reports that eliminate the retrospective reconstruction that can accompany surveys and interviews. Such studies have demonstrated that different interpretations of the environment are associated with different subsequent strategic choices. Later studies ruled out reverse causality (in which prior events might only have been noticed later by senior managers and reported retrospectively in the annual reports) and controlled for the important alternative explanations of organizational capabilities and incentives proposed in the strategic management literature.

The next phase brought both case study and large-sample analyses exploring the relationship of cognition, capabilities and incentives. These studies have shown, for example, that organizational response to changes in the environment may be impeded by inertial cognitive frames even if organizational capabilities and incentives are aligned to take advantage of the opportunity. Conversely, shifts in managerial attention can overcome gaps in capabilities and incentives, or facilitate the later development of organizational factors required for action. In-depth studies inside organizations began to explore the processes underlying these effects. These studies moved away from static representations of cognition in which actors are depicted as constrained by rigid mental models and, instead, showed strategy-making as a dynamic, purposive and politically charged process of meaning construction. These analyses demonstrated that interpretation is both an individual and a social process, and that individuals need social interaction in order to understand and interpret their environments, while at the same time using these social interactions to shape the interpretations of others. From this perspective, making strategy can be comprehended as a product of contests over which a cognitive frame should guide the understanding of an ambiguous environment.

Future research in cognition and strategy could focus profitably on three emerging research areas. The first is to understand how collective frames and categories emerge, both inside organizations and at the field level (where they are negotiated between producers, users and institutional actors). The second is to examine the reciprocal dynamics between cognition and materiality, that is, how new products or technologies might be seeds for new categories or how actions to create collective frames might enable the creation of new markets and systems of exchange. The third is to make links to the burgeoning research in behaviouralist and psychological foundations of strategy, which has, to date, developed somewhat independently from the cognitive tradition.

See Also

- ▶ Cognition and Technical Change
- ▶ Dominant Logic
- ▶ Sensemaking
- ▶ Social Cognition

References

Huff, A.S. 1990. *Mapping strategic thought*. Chichester: Wiley.

Kaplan, S. 2011. Research in cognition and strategy: Reflections on two decades of progress and a look to the future. *Journal of Management Studies* 48: 665–695.

Walsh, J.P. 1995. Managerial and organizational cognition: Notes from a trip down memory lane. *Organization Science* 6: 280–321.

Cognition and Technical Change

Sarah Kaplan
University of Toronto, Rotman School, Toronto, OH, USA

Abstract

A cognitive perspective on technical change suggests that actors' interpretations of a technology shape their choices and actions about it and, as a result, influence the technology's trajectory. Applying a cognitive perspective to the technology cycles suggests that diverse technological frames are a source of variation

in the era of ferment, that framing activities help drive the achievement of a dominant design when one emerges, and that the intertwining of the technological frames with the industry and organizational architectures in the mature phase can explain why transitions to new technologies are difficult.

Definition A cognitive perspective on technical change suggests that actors' interpretations of a technology shape their choices and actions about it and, as a result, influence the technology's trajectory.

Evolutionary models of technical change invoke a lifecycle metaphor to describe technical change. The basic model posits that an era of ferment characterized by high levels of technical variation follows a technological discontinuity. Next, selection among competing technologies leads to the retention of a ▶ dominant design – a set of technologies and associated problem-solving heuristics embodied in a particular product design. Convergence on a dominant design is followed by a mature phase of incremental progress that is ultimately disrupted by another technological discontinuity as the cycle repeats itself. Research over many decades has addressed critical research questions at each phase: Where does technical variation originate? Does a dominant design emerge, and if so, when, and which design becomes dominant? What drives the rate and direction of technical change during the mature phase? When does a technological discontinuity occur? For the most part, the answers have come from economic or organizational perspectives, and cognitive explanations have only recently begun to take hold.

A cognitive perspective on technical change suggests that actors' interpretations of a technology shape their choices and actions about adoption, investment and development, and, as a result, influence the technology's trajectory. Scholars in this stream of research argue that cognition is essential to understanding technical change because technologies are inherently unpredictable and equivocal. As a result, the interpretations made by various actors – be they producers, users or industry institutions – mediate their choices and actions, and these choices (such as whether to invest in or adopt one variant or another) shape the nature of the technology and its trajectory. This short article is drawn in part from a longer treatment of cognition and technical change in Kaplan and Tripsas (2008) that provides a more detailed appraisal and a longer list of relevant citations to work in the area.

Technological Frames and Technical Change

In research on managerial cognition, a plethora of terms have been used to characterize the interpretations actors make: Walsh (1995) and Kaplan (2011) catalogue nearly 100 different terms used in the field. In the context of understanding technical change, one term that has come to the fore is 'technological frame', which Orlikowski and Gash (1994) say shapes how actors categorize a technology, identify specific applications and anticipate the consequences of use.

'Technological frame' is a multi-level construct that operates at the individual and collective levels. Because various collective actors such as firms, institutions and user groups have an impact on technical change, it is sometimes useful to treat these actors as 'cognizers' who 'think' and 'act' on their own. However, a cognitive perspective on technical change assumes that the frames of each organization are the product of interactions of individuals and groups within the organization, each of whom have their own frames (Porac et al. 2001). Collective technological frames may also exist at the industry or field level of analysis in what are sometimes called 'industry recipes' (Spender 1989) or 'field frames' (Lounsbury et al. 2003).

Actors' technological frames are the encoding of the prior histories of the individuals and the organizations within which they operate. Individuals are trained in different fields (e.g., engineering, or chemistry or business) and work in different functional areas (e.g., R&D, marketing, finance). In addition, the common experiences of members create a shared understanding of technology that will be unique to the organization given its distinctive history. Even start-up firms

with no prior organizational history have founders whose unique backgrounds become imprinted in the firm and inform its technological frame. Further, an organizational actor's external affiliations (with industry associations, customers, competitive groups, user groups etc.) provide a flow of certain kinds of information and ideas. Thus, actors' multiple histories lead to knowledge accumulations that are the source of their technological frames.

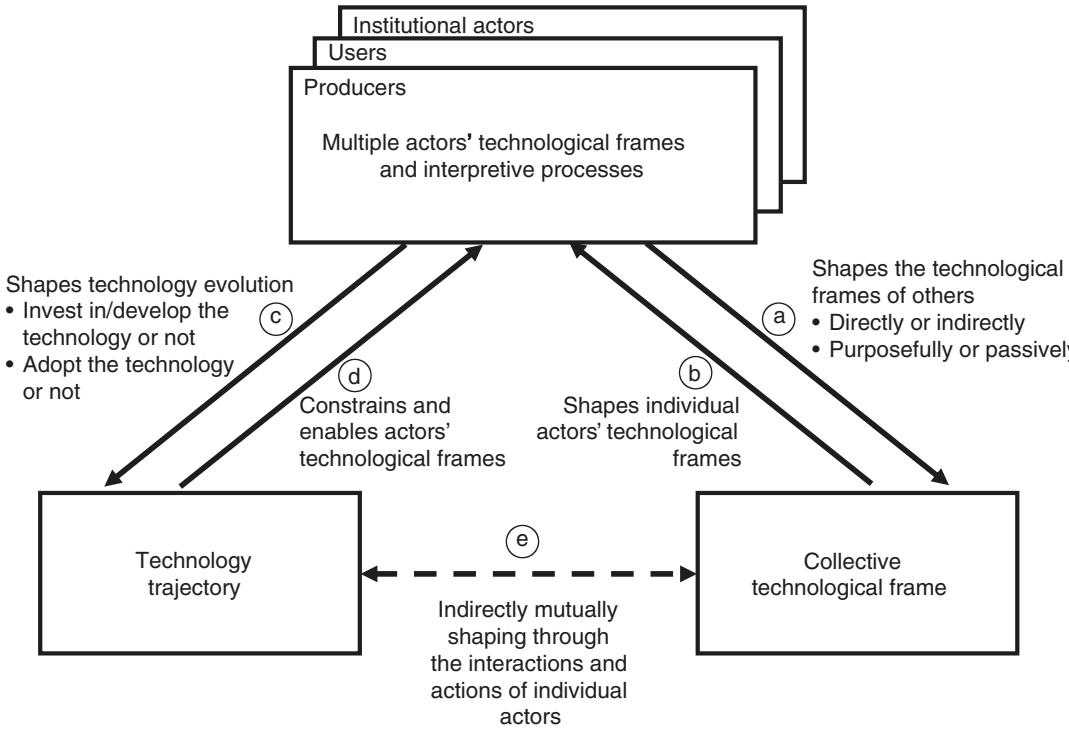
To understand the evolution of technology, one must pay attention to the technological frames of multiple sets of actors. Scholars have highlighted the independent roles of producers, users and institutional actors, though only rarely taking into account the ways that their interpretations shape their actions. While this field of study is still emerging, research has shown that, for producers, only if managers notice a new technology (Kaplan 2008) or perceive it as a threat (Gilbert 2006) will their organizations respond. Further, the beliefs that scientists and managers hold about what will be technically feasible (Garud and Rappa 1994), what business models are required for new technologies (Tripsas and Gavetti 2000) and what solutions will create value (Kaplan and Murray 2010) influence which technologies get developed. Research on users has shown that they impose their assumptions about familiar technologies on new products and use them accordingly (Orlikowski and Gash 1994), and that discontinuities in users' preferences can trigger ▶ technological change (Tripsas 2008). Some studies have shown how producers, users and institutional actors interact in shaping new technological frames, in particular how new categories are negotiated through the media (Rosa and Porac 2002; Kennedy 2008) and how markets are constructed by generating a collective frame about what a new business would be (Weber et al. 2008).

Thus, the underpinnings of a cognitive model of technical change are the technological frames of multiple sets of actors and the interpretive processes that connect these frames to action. Yet a cognitive perspective on technical change is not wholly socially constructionist in the sense that it does not make the claim that interpretations construct the technology. Rather, the view is that the interpretations of actors shape and are shaped by technologies indirectly through a recursive process associated with the choices and actions of these actors in the field. Kaplan and Tripsas (2008) have proposed a model for these dynamics (see Fig. 1).

This model focuses on the reciprocal and mutually constituting dynamics among three components: actors' technological frames and interpretive processes, field-level collective technological frames and the technology itself. The technology trajectory is only indirectly shaped by the collective technological frame (arrow e). The underlying mechanisms are associated with the ways that the technological frames of the multiple actors in the field are developed and how they choose to act in relation to the technology. First, the various actors in the field are likely to have different frames, and through their interactions they produce a collective frame (a). At the same time, existing collective frames in the field will be inputs to individual actors' frames (b). These actors will then make choices about whether to invest in or develop a technology (if they are producers), whether to support or regulate certain technologies (if they are institutional actors), or whether to adopt or adapt a technology (if they are users) (c). These choices are shaped by actors' own technological frames but are not wholly unconstrained. The technology as it exists at the moment, with its particular affordances, constrains and enables the frames that actors can develop and the actions they can take (d). This is a dynamic process. If a technology does not ultimately meet the implied performance criteria in a technological frame, then that incongruity will lead actors to new understandings of what the technology is and could do. Similarly, technology evolution might enable users to discover new applications for the technology, resulting in new interpretations by other actors.

Applying a Cognitive Model to the Technology Lifecycle

The *era of ferment* is empirically characterized by a great deal of variation, but neither the



Cognition and Technical Change, Fig. 1 Role of cognition in technical change (Source: Kaplan and Tripsas 2008)

organizations nor economics literatures has traditionally had much to say about the source of these variations other than that they are stochastic technological breakthroughs, exogenous to the system. A cognitive perspective can contribute to a more robust theory of the sources and types of variation that are produced by showing how the diverse technological frames of various actors produce these outcomes. In an era of ferment, no collective or pre-established frames exist to help make sense of the technology. In their absence, actors draw on their prior frames or search for new ones based on related industries or technologies, categorizing new technologies based on their perceived similarities with existing technologies and applying performance criteria borrowed from other frames (Greve and Taylor 2000). These frames define the solution spaces that producers use to focus their technology development efforts, the kinds of uses that users identify for the new technologies, and the standards established or market definitions created by institutional actors such as regulatory bodies or trade associations. Because producers focus their efforts based on resource allocation processes and because users adopt or innovate based on their needs and preferences, their technological frames shape which technologies are introduced, developed and adopted in the era of ferment. Producers with similar technical capabilities but different technological frames are likely to develop different technologies. Users with different technological frames are likely to implement the same technology in different ways.

The questions for the next phase of the lifecycle are whether a *dominant design* is achieved and, if so, which technical variant becomes dominant. Economists attribute the crystallization of a dominant design to economies of scale or increasing returns to scale resulting from network externalities. Organizational scholars, being informed by a stream of research on the social construction of technology (SCOT), argue that selection among technologies is adjudicated

through competing social groups and coalitions. While not explicitly excluding a cognitive explanation, neither do organizational theorists typically call out the interpretive processes that might be associated with the selection of a dominant design.

The cognitive perspective on technical change suggests that interactions among various actors with different technological frames will determine whether a dominant design emerges and which one wins. These interactions may be passive, taking place in the form of market exchanges: producers offer some technologies and not others, and users adopt them or not, each gaining new information in the exchange and updating how they think about the technology. These interactions may also be purposeful and even political. Producers and users may interact in more formal arrangements such as through joint experimentation and rapid prototyping (von Hippel 1986; Garud and Karnoe 2003). Producers may attempt to influence regulators and users through the media and advertising (e.g., Lampel 2001; Pollock and Rindova 2003). The press itself often seeks a role in defining technology categories (Lounsbury and Rao 2004; Kaplan and Radin 2011). Institutions such as standards setting bodies and technical committees can be arenas in which various actors come together to influence each other's interpretations and shape the collective frame that emerges (Rosenkopf et al. 2001; Rosa and Porac 2002). These various forms of interaction facilitate the emergence of a collective technological frame that, if it is shared widely enough, can enable the materialization of a dominant design. If a collective technological frame does not arise, then it is unlikely that a dominant design will be possible.

In the *mature phase*, technical change becomes incremental as the dominant design is solidified. Economists argue that the inertia in this phase is driven by network externalities such that movement to potentially more socially optimal solutions is thwarted. Organizational scholars attribute inertia to routines. Actors engage in local search, constrained by organizational routines and problem-solving heuristics associated with the dominant design. A cognitive perspective does not contradict these economic and organizational views. Instead, it suggests that a dominant design embodies a collective frame that is reproduced in day-to-day organizational routines, usage patterns and preferences. The system stabilizes as the capabilities, routines, incentives and technological frames align with each other in organizational and industry architectures (Henderson and Clark 1990). Thus, in this phase, technological frames matter, not distinctively, but as an essential part of this self-reinforcing system.

Discontinuities break the inertial system by introducing new technologies. Economists and organizational scholars have mainly argued that this occurs when an old technology reaches its natural limits, but others have shown that a technology's performance limits are as cognitive as they are technical (Henderson 1995). The cognitive perspective suggests that the sources of variation previously described in the era of ferment are the forces that cause old technological frames and technologies to lose their salience. Startup firms are often said to be the source of new technologies, and this is attributed to their flexibility as small entities. From the cognitive viewpoint, entrants – both *de novo* and *de alio* – come up with variations because they see the world through different frames. These new technologies are often ignored by the incumbent players in an industry, not only because of their dependence on existing power and resource structures, but also because the existing technological frame leads actors to perceive the new variant as inferior because it does not match existing performance criteria. Thus, technological discontinuities are more likely to be introduced by industry outsiders or peripheral actors who possess not only different capabilities and incentives but also different technological frames from those of existing industry participants.

Thus, applying a cognitive perspective to the ▶ technology cycles suggests that diverse technological frames are a source of variation in the era of ferment, that framing activities help drive the achievement of a dominant design when one emerges, and that the intertwining of the technological frames with the industry and organizational architectures in the mature phase can explain why transitions to new technologies are

difficult. The latest trends in research in this field have been to delve further into the era of ferment, attempting to understand the emergence of new technological frames (sometimes called schemas or categories) and the sources of discontinuities.

See Also

- Dominant Design
- Innovation
- Management of Technology
- Sensemaking
- Social Cognition
- Technological Change
- Technological Inertia
- Technological Paradigms and Technological trajectories
- Technology Cycles

References

Garud, R., and P. Karnoe. 2003. Bricolage versus breakthrough: Distributed and embedded agency in technology entrepreneurship. *Research Policy* 32: 277–300.

Garud, R., and M.A. Rappa. 1994. A socio-cognitive model of technology evolution: The case of cochlear implants. *Organization Science* 5: 344–362.

Gilbert, C.G. 2006. Change in the presence of residual fit: Can competing frames coexist? *Organization Science* 17: 150–167.

Greve, H.R., and A. Taylor. 2000. Innovations as catalysts for organizational change: Shifts in organizational cognition and search. *Administrative Science Quarterly* 45: 54–80.

Henderson, R. 1995. Of life cycles real and imaginary: The unexpectedly long old age of optical lithography. *Research Policy* 24: 631–643.

Henderson, R.M., and K.B. Clark. 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly* 35: 9–30.

Kaplan, S. 2008. Cognition, capabilities, and incentives: Assessing firm response to the fiber-optic revolution. *Academy of Management Journal* 51: 672–695.

Kaplan, S. 2011. Research in cognition and strategy: Reflections on two decades of progress and a look to the future. *Journal of Management Studies* 48: 665–695.

Kaplan, S., and F. Murray. 2010. Entrepreneurship and the construction of value in biotechnology. *Research on the Sociology of Organizations* 29: 109–149.

Kaplan, S., and J. Radin. 2011. Bounding an emerging technology: Para-scientific media and the Drexler–Smalley debate about nanotechnology. *Social Studies of Science* 41: 457–486.

Kaplan, S., and M. Tripsas. 2008. Thinking about technology: Applying a cognitive lens to technical change. *Research Policy* 37: 790–805.

Kennedy, M.T. 2008. Getting counted: Markets, media, and reality. *American Sociological Review* 73: 270–295.

Lampel, J. 2001. The core competencies of effective project execution: The challenge of diversity. *International Journal of Project Management* 19: 471–483.

Lounsbury, M., and H. Rao. 2004. Sources of durability and change in market classifications: A study of the reconstitution of product categories in the American mutual fund industry, 1944–1985. *Social Forces* 82: 969–999.

Lounsbury, M., M.J. Ventresca, and P.M. Hirsch. 2003. Social movements, field frames and industry emergence: A cultural–political perspective on US recycling. *Socio-Economic Review* 1: 71–104.

Orlikowski, W.J., and D.C. Gash. 1994. Technological frames: Making sense of information technology in organizations. *ACM Transactions on Information Systems* 2: 174–207.

Pollock, T.G., and V.P. Rindova. 2003. Media legitimation effects in the market for initial public offerings. *Academy of Management Journal* 46: 631–642.

Porac, J.F., M.J. Ventresca, and Y. Mishina. 2001. Interorganizational cognition and interpretation. In *Blackwell companion to organizations*, ed. J. A.C. Baum. Malden: Blackwell.

Rosa, J.A., and J.F. Porac. 2002. Categorization bases and their influence on product category knowledge structures. *Psychology & Marketing* 19: 503–531.

Rosenkopf, L., A. Metiu, and V.P. George. 2001. From the bottom up? Technical committee activity and alliance formation. *Administrative Science Quarterly* 46: 748–772.

Spender, J.C. 1989. *Industry recipes: An enquiry into the nature and sources of managerial judgement*. Oxford/New York: Blackwell.

Tripsas, M. 2008. Customer preference discontinuities: A trigger for radical technological change. *Managerial and Decision Economics* 29: 79–97.

Tripsas, M., and G. Gavetti. 2000. Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal* 21: 1147–1161.

Von Hippel, E. 1986. Lead users: A source of novel product concepts. *Management Science* 32: 791–805.

Walsh, J.P. 1995. Managerial and organizational cognition: Notes from a trip down memory lane. *Organization Science* 6: 280–321.

Weber, K., K.L. Heinze, and M. DeSoucey. 2008. Forage for thought: Mobilizing codes in the movement for grass-fed meat and dairy products. *Administrative Science Quarterly* 53: 529–567.

Cognitive Dissonance

Wilfred Amaldoss[1] and Huihui Wang[2]
[1]Duke University, The Fuqua School of Business, Durham, NC, USA
[2]Duke University, Durham, NC, USA

Abstract

Cognitive dissonance theory has a long history in psychology, but has been sparsely examined in the economics literature. It can help answer several questions of practical significance, such as: what causes the formation of social underclass (Oxoby R, Econ J 114:727–749, 2004)? When does preaching morality increase immoral behaviour (Rabin M, J Econ Behav Organ 23:177–194, 1994)? Should the government regulate hazardous industries to ensure worker safety (Akerlof G, Dickens W, Am Econ Rev 72:307–319, 1982)? In this entry, we provide an overview of Festinger's original formulation (1957), outline the subsequent revisions to cognitive dissonance theory and review the related economics literature.

Definition Cognitive dissonance is the unpleasant feeling that arises from holding incompatible cognitions in one's mind. The dissonance can be reduced by modifying the beliefs to be compatible with the individual's actions.

Festinger's Formulation of Cognitive Dissonance

According to Festinger (1957), individuals experience cognitive dissonance if they hold two or more elements of knowledge that are inconsistent with one another. The perceived incompatibility between the cognitions causes a state of psychological unpleasantness and motivates the individual to reduce the resulting discomfort.

To understand Festinger's original work, note that in his experiment participants were first asked to perform tasks that were counter to their personal preference. After participants performed the counter-attitudinal task, they were requested to lie about the desirability of the task to a fellow participant in exchange for either a small financial reward or a high financial reward. A common finding in these experiments is that participants in the high-reward condition attribute the reason for lying to the financial reward. By contrast, participants in the low-reward condition experience cognitive dissonance as the low financial reward is not sufficient to justify their lying; they reduce the dissonance by modifying their preference for the task to be consistent with their action.

Advances in Cognitive Dissonance Theory in the Last 40 Years

In the last 40 years the theory has been revised in the light of the evidence gathered from the hundreds of additional experiments on the phenomenon. We next review the major refinements to the original theory (see also Greenwald and Ronis 1978; Fischer et al. 2008).

Self-consistency. In Festinger's formulation, the mere inconsistency between two cognitions is enough to motivate an individual to reduce dissonance. Aronson (1969), however, shows that dissonance arises only when there is inconsistency between an individual's behaviour and her self-concept. A review of the empirical evidence on the significance of self-esteem for inducing cognitive dissonance presents a mixed picture: while high esteem is conducive to creating dissonance, low self-esteem might not dampen the motivation to reduce dissonance (see Stone 2003, for a review).

Self-affirmation. Steele (1988) suggests that individuals have an overall self-image and are motivated to affirm the integrity of this self-image. Consequently, when two self-relevant cognitions are inconsistent an individual is motivated to affirm the integrity of the self by maintaining a perception of moral and adaptive adequacy (see Sherman and Cohen 2006, for a review).

Self-standards. Stone and Cooper (2003) advance the idea that self is a multi-dimensional construct, and, as such, inconsistency in self-relevant cognitions will induce dissonance,

depending on the accessibility of the different aspects of self.

Aversive consequences. Cooper and Fazio (1984) posit that inconsistency between cognitions will lead to dissonance when an individual feels responsible for producing a negative consequence. While having an aversive consequence can help induce cognitive dissonance, subsequent research shows that aversive consequence is not necessary to evoke dissonance (Harmon-Jones 2000).

Action-based model. Cognitive inconsistency evokes an aversive psychological state that could interfere with effective and unconflicted action (Beckman and Irle 1985). By reducing dissonance an individual can engage in unconflicted action. At times, however, dissonance reduction could lead to prolonged commitment to a harmful course of action (see Harmon-Jones et al. 2010, for a review).

In sum, cognitive inconsistency motivates an individual to reduce the resulting dissonance. Although self-consistency, self-affirmation, self-standard and aversive consequences accentuate cognitive dissonance, an individual can experience dissonance even in circumstances where such factors are weak. Furthermore, although the cognitive dissonance following a decision distorts an individual's beliefs, it helps the individual to adapt to the decision and engage in unconflicted action.

Economic Analysis of Cognitive Dissonance

To facilitate the economic analysis of cognitive dissonance, it is essential to define the premises of the theory and translate them into a tractable formulation. Here we discuss three different formulations of cognitive dissonance.

Intra-person formulation of cognitive dissonance. Akerlof and Dickens' (1982) formulation rests on three premises: (1) individuals have preferences over states of the world and also over their beliefs about the states of the world; (2) individuals can manipulate their own beliefs by selecting sources of information likely to confirm their desired belief; and (3) beliefs, once chosen, persist over time. Based on these premises, they propose a two-period model to explain how cognitive dissonance affects the behaviour of workers in a hazardous industry. In the first period, workers make a choice between pursuing a safe job and working in a hazardous job. Those individuals who choose to work in the hazardous industry face the prospect of an accident as no safety equipment is available. Given their choice in the first period, workers in the hazardous industry reduce their cognitive dissonance by believing that their job is safe, provided the cost of making a wrong decision is not too high. In the second period, even when cost-effective safety equipment is available, workers in the hazardous industry do not buy it as they have come to believe that their job is safe. In this context, legislation can force workers to purchase the safety equipment and thereby attain a Pareto-superior outcome.

To appreciate the modelling of cognitive dissonance, let the probability of accident in the hazardous industry be q, with the cost of an accident being c_a. Although safety equipment is not available in the first period, a worker in the hazardous industry can purchase the equipment in the second period at a cost c_s and thereby eliminate any chance of an accident. The cost c_s is such that $qc_a > c_s$. Each worker in the hazardous industry is afraid of the likelihood of accident and this fear constitutes the non-monetary cost of working in the hazardous industry. Let this non-monetary cost be given by $c_f f$, where c_f is the unit cost of fear and f is the level of fear. The level of fear, in turn, is related to both the objective and the perceived probability of accident in the hazardous industry. Specifically, $f = \frac{q*}{q}$ where $0 \leq q^* \leq q$ is the perceived probability of accident and is malleable to subjective distortion. Considering both the monetary and the non-monetary costs, a worker in the hazardous industry will buy the safety equipment in the second period only if $q * c_a + \frac{q*}{q} c_f > c_s$. This implies that a worker in the hazardous industry should buy the safety equipment in the second period if $q* > \frac{qc_s}{qc_a + c_f}$. Recall that the perceived probability of an accident is malleable for distortion, and thus q^* is

a decision variable. If so, what should be the optimal level of q^* for workers buying the safety equipment and for those not purchasing the safety equipment? A worker not purchasing the equipment will choose $q^* = 0$ to make her belief about the possibility of an accident consistent with her action and thereby reduce cognitive dissonance. A worker who buys the safety equipment in the second period could also end up distorting her belief but to a much lesser extent. In an attempt to maximize the monetary and non-monetary benefits of buying the safety equipment in the second period, this person will reduce the perceived probability of an accident to $q* = \frac{qc_s}{qc_a + c_f}$ so that it is justifiable to purchase the equipment in the second period and yet keep the fear of accident in the first period low.

Akerlof and Dickens' formulation is consistent with the notion of self-affirmation, in that a worker is changing her belief to reaffirm that she had made a smart choice by opting to work in the hazardous industry. Clearly, working in the hazardous industry has aversive consequences and the worker is responsible for the choice. One can also argue that by reducing the perceived probability of accident to zero an individual could focus on her work without any psychological conflict. Next we discuss two formulations that allow cognitive dissonance, which is essentially an intra-person phenomenon, to be tempered by inter-person and inter-group factors.

Inter-person model of cognitive dissonance. Rabin (1994) advances a formulation that is better equipped to study social phenomena. Specifically, dissonance is induced by the gap between one's action and one's perception about the socially acceptable norm, which is endogenous to the model. An individual can distort her perception of the objective social norm to reduce dissonance, but doing so comes at a cost. The cost of holding perceptions divergent from the objective social norm increases as the gap increases. Also, the higher the society's average acceptable norm, the lower both the absolute and marginal cost of convincing oneself that an activity is moral. In equilibrium, the perceived social norm is equal to the society's average belief. Using this simple formulation, Rabin establishes a counter-intuitive result: increasing people's distaste for immoral activities will actually increase the level of immoral activities. The intuition for this result is that when the distaste for immoral activity increases, an individual could either reduce the level of immoral activity or, alternatively, convince herself that the activity is less immoral. Now if everybody believes that an activity is morally more defensible, the socially acceptable level of the immoral activity increases, meaning that it becomes more permissible for each individual to actually engage in the immoral activity. This indirect effect means that we could actually see a higher level of immoral activity. This perverse effect could not occur with an isolated individual, but only when members of a society learn about and care about each other's beliefs.

Inter-group model of cognitive dissonance. Building on Rabin's work, Oxoby (2004) proposes a formulation that views cognitive dissonance as an intergroup phenomenon. Consider a setting where individuals care about their relative position, but can choose the dimension on which their position is measured. In this setting, Oxoby shows how cognitive dissonance could lead to the formation of an underclass in a society. In principle, as status is assigned to individuals based on their ability to consume at an above-average rate, one would expect all individuals to exert greater effort. Yet we observe an underclass whose members do not care about the consumption-based social rank. Why is this so? Oxoby explains that when individuals fail in their efforts to attain high social status, they experience dissonance between their desire and the reality. Such individuals change their beliefs about what constitutes status. For example, instead of engaging in above-average consumption, they might choose to enjoy above-average leisure and view leisure consumption as the measure of true status. This facilitates the formation of a social underclass whose beliefs and habits are different from the mainstream. Note that Oxoby's formulation is consistent with the earlier cognitive dissonance literature based on self-consistency, self-affirmation and self-standard.

In sum, the psychology literature informs us about the different factors that moderate the magnitude of cognitive dissonance. Building on the extant economic models of cognitive dissonance, future work can explore how these psychological factors shape the functioning of competitive markets.

See Also

- ▶ Behavioural Strategy
- ▶ Cognition and Strategy
- ▶ Nash Equilibrium

References

Akerlof, G., and W. Dickens. 1982. The economic consequence of cognitive dissonance. *American Economic Review* 72: 307–319.

Aronson, E. 1969. The theory of cognitive dissonance: A current perspective. In *Advances in experimental social psychology*, ed. L. Berkowitz. New York: Academic.

Beckman, J., and M. Irle. 1985. Dissonance and action control. In *Action control: From cognition to behavior*, ed. J. Kuhl and J. Beckman. Berlin: Springer.

Cooper, J., and R. Fazio. 1984. A new look at dissonance theory. In *Advances in experimental social psychology*, ed. L. Berkowitz. Orlando: Academic.

Festinger, L. 1957. *A theory of cognitive dissonance*. Stanford: Stanford University Press.

Fischer, P., D. Frey, C. Peus, and A. Karenmueller. 2008. The theory of cognitive dissonance: State of the science and directions for future research. In *Clashes of knowledge: Orthodoxies and heterodoxies in science and religion*, ed. P. Meusburger, M. Welker, and E. Wunder. New York: Springer Science and Business Media.

Greenwald, A., and D. Ronis. 1978. Twenty years of cognitive dissonance: Case study of the evolution of a theory. *Psychological Review* 85: 53–57.

Harmon-Jones, E. 2000. Cognitive dissonance and experienced negative affect: Evidence that dissonance increases experienced negative affect even in the absence of aversive consequences. *Personality and Social Psychology Bulletin* 26: 1490–1501.

Harmon-Jones, E., D. Amodio, and C. Harmon-Jones. 2010. Action-based model of dissonance: On cognitive conflict and attitude change. In *Psychology of attitudes and attitude change*, ed. J. Forgas, J. Cooper, and W. Crano. New York: Psychology Press.

Oxoby, R. 2004. Status, cognitive dissonance, and the growth of the underclass. *The Economic Journal* 114: 727–749.

Rabin, M. 1994. Cognitive dissonance and social change. *Journal of Economic Behavior and Organization* 23: 177–194.

Sherman, D., and G. Cohen. 2006. The psychology of self defense: Self-affirmation theory. In *Advances in experimental social psychology*, ed. M. Zanna. San Diego: Academic.

Steele, C. 1988. The psychology of self-affirmation. In *Advances in experimental social psychology*, ed. L. Berkowitz. San Diego: Academic.

Stone, J. 2003. Self-consistency for low self-esteem in dissonance processes: The role of self-standards. *Personality and Social Psychology Bulletin* 29: 846–858.

Stone, J., and J. Cooper. 2003. The effect of self-attribute relevance on how self-esteem moderates attitude change in dissonance processes. *Journal of Experimental Social Psychology* 39: 508–515.

Collaborative Innovation

Haluk Demirkan[1], James C. Spohrer[2] and Nicholas Donofrio[3]

[1]University of Tacoma, Tacoma, WA, USA

[2]IBM University Programs World-Wide, Almaden Research Center, San Jose, CA, USA

[3]IBM Corporation, USA

Definition Collaborative innovation is a dynamic capability of entities for sensing, seizing and managing new levels of socioeconomic transformation, based on opportunity sharing within and between multidisciplinary, multi-sector, multicultural researchers, practitioners and policymakers, in order to realize open innovations more effectively and efficiently.

Exploiting routine collaborations based on division of labour or comparative advantage is fundamental to socio-economic evolution. In contrast, collaborative innovation is a dynamic capability (Teece et al. 1997) of entities exploring necessary modes of socio-economic transformation and evolution. As a result of collaborative innovation, people work together more synergistically, collaborate outside their 'silos', become invigorated and gain a deeper, more common, understanding of shared ▶ value. As a result, new but unpredictable ideas, solutions and opportunities are generated,

speed is improved significantly and desired outcomes are reached more frequently.

Collaborative innovation is about how people with collective vision interact with each other, how they offer ideas and intuitive thinking and how they share knowledge and experience for a common goal. It is a value co-creation process that not only results in a desired outcome for all collaborators but simultaneously co-elevates the capabilities of entities. In its simplest form, this can be about recursive interaction of co-creativity, information and knowledge sharing, and mutual learning between two or more people working together towards a common goal of generating new sources of growth or wealth in an organization (Donofrio et al. 2010). On a larger scale, this can be about many people and entities interacting (e.g., professionals, citizens, businesses, educational institutions, nations), forming global networks, defining open standards to reduce transaction costs, and continuously benefiting from finding new business models to collaborate and improve each other's capabilities.

In the 1990s, several researchers emphasized the collaborative aspect of innovation processes, which goes beyond the organizational boundary of the firm and involves the formation of alliances, cooperation and collective arrangements between several actors (Akrich et al. 1988; Freeman and Soete 1997; Freeman 1991). Most formally, Chesbrough (2003) introduced an 'open innovation' approach as a way of breaking out of 'silos' and breaking through boundaries to achieve collaborative innovation. The open innovation approach has been adopted by information technology companies as a strategy to combine internal and external sources of ideas as well as internal and external channels to market in order to advance their technological platforms.

Historically, technical and organizational barriers frequently made higher levels of collaboration difficult. Today, new management approaches and emerging information communication technologies (ICTs) overcome some of these barriers.

To foster innovation, organizations need to develop semi-structured procedures and places where they can come together, work creatively, search data and share information (Nambisan 2009). For example, in manufacturing, computer-aided-design (CAD) and computer-aided engineering (CAE) tools, and project scheduling programmes support structured collaboration (Swink 2006). The reason we called these 'semi-structured' procedures is because unstructured collaboration promotes creativity, while structured collaboration promotes efficiency.

Even some ICT solutions have started to overcome barriers associated with distance (most participants can be geographically dispersed), access to information and resources, lack of institutional memory and awareness. There are still barriers with individual and organizational behaviours (i.e., sense of direction, goal sharing and alignment, commitment, opportunism, trust, competence, capability building, division of power, culture and communication).

The best companies recognize that collaborative innovation is one of the most powerful means of creating new ideas that impact on revenues as well as expenses. Successful collaborative innovation requires a non-adversarial mind-set, a multi-level and multifunctional organizational approach, the ability to learn to speak 'another language', new metrics and the willingness to share intellectual property. First, organizations need to clearly formulate their internal collaborative innovation strategy before starting to collaborate with external parties.

CEOs, government officials, academics and community leaders around the world are all counting on 'innovation' to be the fundamental driver of economic opportunity, job creation, business competitiveness and advances in education, health care and a vast range of other disciplines. Investing in innovation, they say, is the surest way to survive and thrive in today's complex, connected world. For collaborative innovation to become part of our collective DNA, we must accept the notion that the surest way to make progress and solve problems is to tap into the collective knowledge of the team. Networked

enterprises are the future. No individual enterprise, however large and talented, can afford to go it alone in today's highly competitive, globally integrated marketplace.

For entities (e.g., companies, governments and educational institutions), the choice is either innovation or commoditization. Most important, collaborative innovation today should be about new growth opportunities in both economic and societal activity. Seizing the opportunities demands unique foresight and capability. As collaborative innovation takes hold, the availability of professional talent, infrastructure and investment are increasing everywhere, making the world more tightly integrated and dependent on collaborative innovation to survive and thrive.

See Also

- ▶ Autonomous Innovation
- ▶ Innovation
- ▶ Product Innovation
- ▶ Systemic Innovation
- ▶ Value

References

Akrich, M., M. Callon, and B. Latour. 1988. A quoi tient le succès des innovations? 1: L'art de l'intéressement, gérer et comprendre. *Annales de Mines* 11: 4–17.

Chesbrough, H.W. 2003. *Open innovation: The new imperative for creating and profiting from technology.* Boston: Harvard Business School Press.

Donofrio, N., C. Sanchez, and J. Spohrer. 2010. Collaborative innovation and service systems. In *Holistic engineering education*, ed. D. Grasso. New York: Springer.

Freeman, C. 1991. Network of innovators: A synthesis of research issues. *Research Policy* 20: 499–514.

Freeman, C., and L. Soete (eds.). 1997. *The economics of industrial innovation*. New York: Psychology Press.

Nambisan, S. 2009. Platforms for collaboration. *Stanford Social Innovation Review* 7: 44–49.

Swink, M. 2006. Building collaborative innovation capability. *Research Technology Management* 49: 37–47.

Teece, D.J., G. Pisano, and A. Shuen. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* 18: 509–533.

Collusion and Cartels

Matthew Olczak[1] and Luke Garrod[2]
[1]Aston University, Birmingham, UK
[2]Loughborough University, Loughborough, UK

Abstract

Collusion refers to conduct where firms cooperate over time to raise prices above competitive levels. Preventing collusion is one of the main aims of ▶ competition policy, and there is a distinction between explicit and tacit collusion. Explicit collusion refers to a cartel that colludes by directly communicating with each other. Tacit collusion is where firms collude without such explicit communication. Usually, only explicit collusion is considered illegal. Competition authorities attempt to deter cartels through sanctions on the firms and the individuals involved, and leniency programmes are an important method in which cartels are detected. Firms are encouraged to establish compliance programmes to avoid breaches of the law by their employees.

Definition Collusion is conduct in which rival firms cooperate with each other over time to raise prices above competitive levels through coordinated action. A cartel is a group of firms that conspire to reach an agreement over such conduct by explicitly communicating with each other.

Collusion

▶ Competition ensures that products are sold at low prices, so firms commonly have a collective incentive to raise profits by dampening competition. Collusion refers to conduct where firms act on such incentives by cooperating over time to raise prices above competitive levels. For example, such cooperation may involve price fixing, sales quotas, bid rigging, and/or mutual forbearance. Collusive conduct comes at the expense of any downstream firms and final consumers, and it

causes the economic inefficiencies associated with ▶ market power, including a dead-weight loss to society (see Motta 2004: 41–44).

Sustaining collusion is not easy. Colluding firms must agree upon acceptable conduct, and each firm must resist a short-term unilateral incentive to increase their market share by deviating from that conduct. Such temptations can be eliminated by the threat of engaging in a sufficiently fierce price war whenever a rival is suspected of deviating. Consequently, firms have to be able to monitor their rivals' actions to some extent to ensure that any such deviations from the collusive agreement can be detected and punished accordingly.

Using the theory of repeated games, which models strategic interactions with the above ingredients, economists have discovered a number of factors that facilitate collusion (see Ivaldi et al. 2003). For example, a now well established result is that collusion is most likely to occur in markets with a small number of similar firms.

Cartels

Preventing collusion is one of the main aims of competition policy, and there is an important distinction between explicit and tacit collusion. Explicit collusion is where a group of firms, commonly referred to as a cartel, reach an agreement over their conduct by directly communicating with each other. Such communication may take place in face-to-face meetings, over the telephone, via email or by any other means. Tacit collusion is where firms reach a collusive understanding without such explicit communication. This can arise naturally through repeated interaction in a market, and reaching an agreement is likely to involve some trial and error (Harrington and Zhao 2012).

Anticompetitive agreements, such as collusion, are commonly prohibited by competition laws and hard evidence of a cartel (e.g., recorded conversations, minutes of meeting, or emails) is usually required to prove guilt of collusion. Consequently, tacit collusion is not usually considered illegal, despite causing similar effects as explicit collusion. This approach to the law is arguably desirable because it ensures legal certainty over illegal conduct, and it prevents competitive behaviour from being punished erroneously, which could undermine the market mechanism across the economy (Motta 2004: 185–190).

C

Sanctions

Competition authorities impose sanctions on detected cartels in an attempt to deter such conduct. The main sanctions include (International Competition Network 2008):

- Pecuniary fines on cartel members. The level often depends upon the value of sales and the duration the cartel was active, with adjustments due to aggravating and mitigating factors. They can then be further adjusted to avoid bankrupting firms.
- Incarceration of culpable individuals. This aims to deter employees involved in the day-to-day running of the firm from participating in cartel activity, as they may be less concerned about company fines than the firm's shareholders.
- Payment of damages to victims. This can involve litigation in which the victims must prove the harm they suffered as a result of the cartel. In the US, victims can receive compensation of as much as three times the harm. The harm is commonly calculated by determining how much buyers are overcharged due to the existence of the cartel. This calculation is non-trivial and often controversial because it involves estimating some hypothetical price that buyers would have been charged had the cartel not existed (Connor 2014).

Since the mid-1990s, cartel sanctions have increased substantially. For example, 2014 was a record year for fines with $5.3 billion imposed globally (Financial Times 2014), and jail terms of over 2 years in the US are increasingly common (The Economist 2014). Despite this, there are still doubts over the extent to which cartels will be deterred by these sanctions (Connor and Landes 2012).

Detection

A number of techniques can be used by competition authorities to screen for cartels, but there remains scepticism of the value of such techniques (The Economist 2014). As a result, most cartels are detected through an involved party admitting to the illegal activity (Stephan and Nikpay 2015). Such 'whistleblowing' is encouraged through leniency programmes that grant immunity to the first firm to come forward, thereby creating a race to the authorities to avoid sanctions. Hard evidence may be provided by the whistleblowers as part of the conditions for leniency, and competition authorities have powers to undertake dawn raids of suspected cartel members in the hope of seizing incriminating documents and messages. Recent evidence suggests that the introduction of leniency programmes has significantly increased detection of cartels, though it is estimated that less than a fifth of cartels are detected by competition authorities (Miller 2009; Ormosi 2014).

Compliance Programmes

Competition authorities place the responsibility of compliance with the law on the board of directors and senior management by encouraging firms to establish compliance programmes (Baer 2014). Such programmes aim to prevent breaches of the law and ensure early detection, by providing training about and incentivising appropriate conduct, facilitating anonymous reporting systems, auditing employee conduct, and disciplining breaches (Murphy and Kolansky 2012). While some jurisdictions treat the presence of a compliance programme as a mitigating factor in terms of fine calculations, it is often argued that the tough sanctions and the scope for leniency should provide sufficient incentives for effective compliance programmes to be put in place (for example, see Almunia 2011).

See Also

- ▶ Competition
- ▶ Market Power
- ▶ Perfect Competition
- ▶ Prisoner's Dilemma
- ▶ Rivalry and Collusion

References

Almunia, J. 2011. Cartels: The priority in competition enforcement. Speech by Vice President of the European Commission responsible for Competition Policy, 15th International Conference on Competition, Berlin, 14 April. Available at http://europa.eu/rapid/press-release_SPEECH-11-268_en.htm?locale=en. Accessed 22 June 2015.

Baer, B. 2014. Prosecuting antitrust crimes. Remarks as prepared for the Georgetown University Law Center Global Antitrust Enforcement Symposium, Washington, DC, 10 September. Available at http://www.justice.gov/atr/public/speeches/308499.pdf. Accessed 22 June 2015.

Connor, J.M. 2014. Price-fixing overcharges: Revised 3rd Edition, 24 February. Available at SSRN: http://ssrn.com/abstract=2400780. Accessed 22 June 2015.

Connor, J.M., and R.H. Landes. 2012. Cartels as a rational business strategy: Crime pays. *Cardozo Law Review* 34: 427–490.

Financial Times. 2014. Global fines for price-fixing hit $5.3bn record high, 6 January. Available at http://www.ft.com/cms/s/0/83c27142-95a8-11e4-b3a6-00144feabdc0.html#axzz3PYoUFpG9. Accessed 22 June 2015.

Harrington, J., and W. Zhao. 2012. Signaling and tacit collusion in an infinitely repeated Prisoners' Dilemma. *Mathematical Social Sciences* 64: 277–289.

International Competition Network. 2008. Setting of fines for cartels in ICN jurisdictions. Report to the 7th ICN Annual Conference, Kyoto, April. Available at http://www.internationalcompetitionnetwork.org/uploads/library/doc351.pdf. Accessed 22 June 2015.

Ivaldi, M., B. Jullien, P. Rey, P. Seabright, and J. Tirole. 2003. The economics of tacit collusion. Final Report for DG Competition, European Commission.

Miller, N. 2009. Strategic leniency and cartel enforcement. *American Economic Review* 99: 750–768.

Motta, M. 2004. *Competition policy: Theory and practice*. Cambridge: Cambridge University Press.

Murphy, J., and W. Kolansky. 2012. The role of anti-cartel compliance programs in preventing cartel behavior. *Antitrust* 26: 61–64.

Ormosi, P. 2014. A tip of the iceberg? The probability of catching cartels. *Journal of Applied Econometrics* 29: 549–566.

Stephan, A., and A. Nikpay. 2015. Leniency decision-making from a corporate perspective: Complex realities. In *Anti-cartel enforcement in a contemporary age: The leniency religion*, ed. C. Beaton-Wells and C. Tran. Oxford: Hart Publishing.

The Economist. 2014. Just one more fix, 29 March. Available at http://www.economist.com/news/business/21599799-trustbusters-have-got-better-detecting-cartels-and-bolder-punishing-them-incentives. Accessed 22 June 2015.

Commoditization

Glenn Hoetker
Arizona State University, W. P. Carey School of Business, Tempe, AZ, USA

Definition Commoditization refers to the transformation of a differentiated product into an undifferentiated product, and is typically accompanied by an increase in competition based on price rather than features.

Differentiation occurs when a product offers something unique that is valuable to buyers beyond simply offering a low price (Porter 1985) and allows firms to charge a price premium for its product. Commoditization refers to the process through which previously differentiated products become undifferentiated, removing the ability of the firm to charge a price premium and changing the terms of competition to price, rather the features. According to the industry life-cycle model of Utterback and Abernathy (1975), differentiation is most prominent early in an industry's lifecycle, as producers add new product features to address uncertain consumer demand. Over time, producers and users gain experience, reducing market uncertainty. At the same time, technological advances allow firms to develop product innovations aimed at specific user needs, increasing product differentiation. However, once most obvious improvements have been incorporated by the majority of producers, the demands of mass production lead to standardization, and innovation shifts primarily to create improvement in process. Given common features across producers, the terms of competition shift to price, margins shrink and industries often become oligopolistic (Cusumano and Rosenbloom 1987; Christensen 1997). The pursuit of economies of scale and optimized production processes serves to further reduce the incentives for product variation. The shift from product to process innovation is associated with the emergence of a ▶ dominant design (Abernathy and Clark 1985; Tushman and Anderson 1986), a widely accepted standard that crowds other designs out of the market. In economic terms, commoditization marks the shift away from monopolistic competition towards perfect competition.

In many cases, commodification is neither permanent, nor inevitable. At the industry level, technological discontinuities may lead to a new era of ferment (Anderson and Tushman 1990), leading again to uncertainty, product variation and a renewed opportunity to differentiate. Even if a product itself can no longer be differentiated, firms can sometimes resist commodification by differentiating associated features such as warranty, packaging, distribution, after-sales service or, in some cases, brand and trademark (Chamberlin 1962; Porter 1985).

See Also

- ▶ Dominant Design
- ▶ Innovation Strategies

References

Abernathy, W.J., and K.B. Clark. 1985. Innovation: Mapping the winds of creative destruction. *Research Policy* 14: 3–22.

Anderson, P., and M.L. Tushman. 1990. Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative Science Quarterly* 35: 604–633.

Chamberlin, E. 1962. *The theory of monopolistic competition: A re-orientation of the theory of value*. Cambridge, MA: Harvard University Press.

Christensen, C.M. 1997. *The innovator's dilemma: When new technologies cause great firms to fail*. Boston: Harvard Business School Press.

Cusumano, M.A., and R.S. Rosenbloom. 1987. Technological pioneering and competitive advantage: The birth of the VCR industry. *California Management Review* 1: 51–76.

Porter, M.E. 1985. *Competitive advantage: Creating and sustaining competitive performance*. New York: Free Press.

Tushman, M.L., and P. Anderson. 1986. Technological discontinuities and organizational environments. *Administrative Science Quarterly* 31: 439–465.

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

Comparative Advantage

Deepak Somaya
University of Illinois, College of Business, Champaign, IL, USA

Definition Comparative advantage is a theory of trade, traditionally attributed to David Ricardo, which argues that trade between two countries can be advantageous even when one country has an absolute advantage (higher efficiency) in producing all goods and services.

The theory of comparative advantage has its roots in the literature on international trade, and is classically attributed to David Ricardo (1821). Its central argument is that a country may gain from trade with another even if it is more efficient than the latter in the production of all goods. To use Ricardo's now famous example, England and Portugal would find it advantageous to trade Portuguese wine for English cloth even if Portugal was more efficient in the production of both goods, so long as it is *relatively* more efficient in the production of wine over cloth. Say, in line with Ricardo, that Portugal produced a quantity of wine with the labour of 80 person-years (versus 120 in England) and cloth with 90 person-years (versus 100), then Portugal would have a comparative advantage in wine production (and England in cloth) despite being more efficient in producing both goods. Ricardo explained that the rationale for trade arises 'because it would be advantageous to [Portugal] rather to employ her [resources] in the production of wine, for which she would obtain more cloth from England [than the cloth Portugal could produce by employing those same resources]' (Ricardo 1821). England would likewise benefit from freeing up her own resources to produce cloth, rather than wine.

Ricardo's logic was formalized in the twentieth century by Swedish economists Eli Heckscher and Bertil Ohlin, and was further extended by several scholars. The Heckscher–Ohlin model links comparative advantage to country factor endowments, typically capital and labour, and shows that free trade automatically results in specialization of production. The relatively capital-rich (labour-rich) country produces and exports more of the capital-intensive (labour-intensive) good, and the terms of trade are such that total surplus increases in both countries.

The implications of comparative advantage for the sourcing and location choices of multinational enterprises are foundational to the field of ▶ international business (see e.g., Caves 1996), but the theory is very rarely invoked in contemporary international business research (for instance, only three articles in the *Journal of International Business Studies* between 2003 and 2012 even mention the term comparative advantage). Plausibly, international business research draws on more immediate manifestations of comparative advantage, such as differences in factor prices or resource availability, and therefore abstracts away from the theoretical drivers of these differences (which are of relatively greater interest to international *economics* scholars).

Comparative advantage has also had an impact on strategic management research; specifically to explain specialization and market exchange among firms, which can be analogized to international trade. This research builds on the ▶ resource-based view (RBV), and a broader set of arguments that link firm boundaries to capability differences between firms. In contrast to governance-based explanations, RBV scholars posit that (and find empirical support for) a firm's advantages in unique and non-tradable resources may drive its choice of internalized and outsourced activities (e.g., Argyres 1996; Madhok 1996; Leiblein and Miller 2003; Hoetker 2005; Mayer and Solomon 2006).

Some scholars have suggested that firms may choose to engage in outsourcing even without such absolute differences in resource advantage (Jacobides 2005; Jacobides and Hitt 2005; Jacobides and Winter 2005). A key distinction drawn with resource based explanations is a 'simpler' focus on variation in raw productive abilities, without recourse to resources as a source of such variation. While this stream of work has helped advance comparative advantage as a potential explanation for firm boundaries, a

systematic examination of the theory's limits when it is extended from countries to firms awaits further research.

See Also

- ▶ Firm Size and Boundaries, Strategy
- ▶ International Business
- ▶ Outsourcing
- ▶ Resource-Based Theories
- ▶ Resource-Based View
- ▶ Trade Creation/Diversion
- ▶ Vertical Integration

References

Argyres, N. 1996. Evidence on the role of firm capabilities in vertical integration decisions. *Strategic Management Journal* 17: 129–150.

Caves, R.E. 1996. *Multinational enterprises and economic analysis*, 2nd ed. Cambridge: Cambridge University Press.

Hoetker, G. 2005. How much you know versus how well I know you: Selecting a supplier for a technically innovative component. *Strategic Management Journal* 26: 75–96.

Jacobides, M.G. 2005. Industry change through vertical disintegration: How and why markets emerged in mortgage banking. *Academy of Management Journal* 48: 465–498.

Jacobides, M.G., and L.M. Hitt. 2005. Losing sight of the forest for the trees? Productive capabilities and gains from trade as drivers of vertical scope. *Strategic Management Journal* 26: 1209–1227.

Jacobides, M.G., and S.G. Winter. 2005. The co-evolution of capabilities and transaction costs: Explaining the institutional structure of production. *Strategic Management Journal* 26: 395–413.

Leiblein, M.J., and D.J. Miller. 2003. An empirical examination of transaction- and firm-level influences on the vertical boundaries of the firm. *Strategic Management Journal* 24: 839–859.

Madhok, A. 1996. The organization of economic activity: Transaction costs, firm capabilities, and the nature of governance. *Organization Science* 7: 577–590.

Mayer, K.J., and R. Salomon. 2006. Capabilities, contractual hazard and governance: Integrating resource-based and transaction cost perspectives. *Academy of Management Journal* 49: 942–959.

Ricardo, D. [1817] 1821. *On the principles of political economy and taxation,* 3rd edn. London: John Murray. Available at http://www.econlib.org/library/Ricardo/ricP.html. Accessed 20 November 2011.

Competency Trap

Gautam Ahuja[1] and Vivek Tandon[2]
[1]University of Michigan, Stephen M Ross School of Business, Ann Arbor, MI, USA
[2]National University of Singapore, Singapore, Singapore

C

Abstract

It is frequently observed that many firms, especially those that are successful in the current environment, fail to change quickly enough when the environment changes significantly. One potent explanation for these observations is a phenomenon of learning called a 'competency trap'. In this article we briefly describe what competency traps are, the factors that lead to them, the consequences for the firm, including reduced adaptability and missed opportunities, and, finally, the various means through which a firm can avoid falling into the competency traps.

Definition Competency trap is a pathology of learning wherein an actor persists with current practices and does not learn alternatives that are superior in the long term because previous experience makes continued use of current practices more attractive than adopting new ones that yield smaller returns in the short term.

Definition and Relevance

It has been frequently observed by scholars that many firms that are quite competent in their current activities falter and fail to change quickly enough when the environment around them changes significantly. Furthermore, this propensity to fail in changed circumstances seems correlated with their success in the earlier environment (Barnett and Pontikes 2008). One potent explanation for these observations is a phenomenon of learning called a 'competency trap' (Levitt and March 1988). An entity falls into a competency trap when it persists with its current practices or procedures, even in the presence of superior

alternatives, because its prior experience with the current practices makes their continued use more attractive than adopting new ones which yield smaller returns in the short term. A competency trap can also be thought of in theoretical terms as an organization being stuck with a locally optimum configuration of decisions where moving to a better optimum entails adjusting the configuration in a way that could eventually yield better results but reduces the entity's performance in the short term (March 1991). Here we briefly list the antecedences and consequences of competency traps and the means to counter them.

Entities other than firms can also fall into competency traps. Individuals, through a process of experiential learning, can develop expertise with certain technologies or practices and then refrain from gaining experience in different but potentially more efficient technologies because continuing to utilize current skills can be more beneficial in the short term. Similar effects can be observed for other entities such as governments and even societies and other economic systems (consider, for instance, the persistence of the QWERTY keyboard even in the presence of superior ones). Given the differences in goals and the resources at the disposal of these entities, the factors that cause them to fall into competency traps, the consequences of falling into these traps and the means that can be employed to come out of the trap are likely to be considerably different across these entities. Owing to considerations of length, here we limit our scope to for-profit firms.

Factors Leading to Competency Traps

Learning Myopias

One of the primary reasons why firms are trapped into persisting with their current set of competencies is the process through which they learn and adapt. Most learning is local. Firms learn by making and evaluating changes close to their current set of activities (Nelson and Winter 1982). Because predicting the consequences of radical changes in complex conditions is very difficult, firms prefer to make only incremental changes that do not destabilize the organization. This propensity to experiment with only proximate changes tends to increase the facility with current practices and routines and prevents the firm from gaining radically different competencies. This propensity for proximate changes is also reflected in firms' tendency to ignore the consequences that are distant in time (Levinthal and March 1993). Short-term survival and the unpredictability of the long term leads firms to sacrifice long-term competitiveness for short-term gains (Levinthal and March 1993). Consequently, firms are less inclined to shift resources from learning and improving their current competencies in the interest of building new competencies that can yield superior returns in the future.

Organizational Factors

A number of organizational factors are also instrumental in trapping the firm into persisting with its current set of competencies and routines instead of developing and moving towards new ones. A firm's various competencies are based on various bundles of routines developed within the firm. Indeed, a firm can be thought to consist of these routine-bundles linked together in a complex interdependent system (Nelson and Winter 1982; Winter 2003). Avoiding the competency trap and building new competencies involves developing new routines, disrupting some of the existing routines and incorporating the new routine into the existing system. This disrupts the existing system of routines and, in an interdependent system, can cause unpredictable consequences far from the actual locus of change. This increases the cost of dramatically changing the status quo; disruptions from small incremental improvements can be contained and controlled. Interdependent routines therefore usually favour the deepening of existing competencies over developing new ones. The structure of the firm — the way different units are grouped together and communicate with each other and how the authority and decision-making is distributed among the units — also restricts the pattern of information flow and shapes the pattern of attention within the firm (Ocasio 1998). This restriction of attention and information thereby constrains the firm from recognizing the importance of new sets of competencies, especially those that require a new configuration of activities

and a new arrangement of the components of the system (Henderson and Clark 1990). In this way the current structure of the firm, by hiding the potential of new competencies, can keep the firm trapped in its current set of competencies. Another organizational factor preventing the firm from developing new sets of competencies is politics: certain influential organizational members may have vested interests in continuing with the old competencies and prevent the firm from adopting new competencies (Taylor 2010).

Environmental Factors

Firms also fall into competency traps in response to the demands placed on them by the environment. The expectations of many external constituents: customers, capital providers and institutions, as well as competitive pressures, focuses the firm on the short term and on fulfilling the current demands reliably, a focus that supports maintaining and improving the current competencies rather than taking the risk of building new and different capabilities. Christensen and Bower (1996) demonstrate the impact of the needs of a firm's major customers on its inability to allocate resources towards developing new markets through new technologies. Ahuja and Lampert (2001) identify three sets of traps that emerge from such 'rational' constraints on organizations – maturity traps, propinquity traps and familiarity traps. Similarly, scholars have shown that pressures from financial markets to produce financial results in the short term prevent firms from investing in R&D and thus building future technological capabilities (Bushee 1998). Other scholars (Barnett and Hansen 1996) have argued that competitive pressures force a firm to continually fine-tune their current set of competencies and make them efficient vis-à-vis the current environment, in effect ignore investing in developing a new set of competencies that can make them more flexible.

Consequences of Competency Traps

Adaptability to Changes

The most obvious and perhaps the most pernicious consequence of falling into a competency trap is that it makes a firm vulnerable to changes in the environment and thus can prove lethal in the long term. A competency trap slows a firm's response to adverse changes in the environment not only because superiority in the current practices increases the costs of moving away from them but also because increased competence in current practices delays the adverse impacts of external changes. Initially the superiority in current competencies may still continue to provide satisfactory returns until the adverse changes gather potency over time and then it may be too late for the firm to change; the incumbents' fallibility to disruptive technologies (Christensen and Bower 1996) is one example of this malignancy.

Missing Opportunities

Another related but subtly different consequence of competency traps is that they cause firms to miss out on many opportunities. Continuing with current competencies and unwillingness to invest in developing new ones can lead the firms to ignore many technologies that may exist within the firm itself, technologies that might later turn out to yield far greater returns than the original competencies (for instance, given the computer technologies developed in Xerox labs, Xerox could have dominated the computer revolution but its inability to invest in new markets caused it to lose out). Although missing out opportunities to enter new markets and fully utilize its latent resources may not necessarily kill a firm (unless the missed opportunities can evolve and later compete with a firm's current area of business), it does lead the firm to perform below its potential and be more vulnerable to exogenous shocks.

Means to Avoid or Escape Competency Traps

Structural Changes

As discussed earlier, a firm's structure considerably influences the patterns of communication and attention within a firm, thereby leading it to incrementally improve its current set of routines and systematically ignore new competencies. One

way to break out of this competency trap is to change the structure within the firm (Siggelkow and Levinthal 2003), forcing the organization to pay attention to different configurations of resources and develop new competencies. Another structural mechanism is to create separate units with different structures (such as skunk works) with the explicit mandate to develop different competencies (Fosfuri and Rønde 2009).

Breadth of Search

Competency traps arise from the localness of search for improvements. Firms can escape and avoid the traps by engaging in distant searches for solutions, such as by scanning technologies and competencies in different domains and in scientific communities such as universities (Ahuja and Katila 2004). Another organizational means of forcing the organizational members to look beyond their current competencies is raising the aspirations of the organization so that incremental proximate solutions are not enough to meet the aspirations (Levinthal and March 1993). Another means is to allow autonomous search processes at lower and middle managerial levels (Burgelman 1983). This allows the firm to explore different possibilities and discover many different competencies that can be potentially invested in. In effect, such a strategy allows 'multiple flowers to bloom' and enables the firm to identify new competencies that can be developed. Similarly, creating a corporate venture capital fund which invests in multiple different strategically relevant areas can also enable the firm to discover the new competencies that it needs to acquire (Dushnitsky and Lenox 2005).

Hiring and Collaborating with Partners

Firms can escape the myopia of search and break out of their competency traps by hiring new talent from the external labour market. The 'imported' talent brings with it new knowledge and skills which can be used to move the firm to different competence-building trajectories. This strategy, however, is not always successful and many factors, such as the presence of persistent trajectories in the focal firm (Song et al. 2003) and the distribution of power among the focal firm's existing employees (Tzabbar 2009; Taylor 2010), influence the effectiveness of this strategy. Similarly, collaborating with other firms in the industry also exposes the firm to different competencies and enables it to avoid the dangers of falling into a competency trap.

See Also

- ▶ Capability Development
- ▶ Dynamic Capabilities
- ▶ Dynamics of Resource Erosion, the
- ▶ Knowledge-Based Strategy
- ▶ Strategic Organization Design

References

Ahuja, G., and R. Katila. 2004. Where do resources come from? The role of idiosyncratic situations. *Strategic Management Journal* 25: 887–907.

Ahuja, G., and C.M. Lampert. 2001. Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal* 22: 521–543.

Barnett, W.P., and M.T. Hansen. 1996. The Red Queen in organizational evolution. *Strategic Management Journal* 17: 139–157.

Barnett, W.P., and E.G. Pontikes. 2008. The Red Queen, success bias, and organizational inertia. *Management Science* 54: 1237–1251.

Burgelman, R.A. 1983. A process model of internal corporate venturing in the diversified major firm. *Administrative Science Quarterly* 28: 223–244.

Bushee, B.J. 1998. The influence of institutional investors on myopic R&D investment behavior. *Accounting Review* 73: 305–333.

Christensen, C.M., and J.L. Bower. 1996. Customer power, strategic investment, and the failure of leading firms. *Strategic Management Journal* 17: 197–218.

Dushnitsky, G., and M.J. Lenox. 2005. When do incumbents learn from entrepreneurial ventures? Corporate venture capital and investing firm innovation rates. *Research Policy* 34: 615–639.

Fosfuri, A., and T. Rønde. 2009. Leveraging resistance to change and the skunk works model of innovation. *Journal of Economic Behavior & Organization* 72: 274–289.

Henderson, R.M., and K.B. Clark. 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly* 35: 9–30.

Levinthal, D.A., and J.G. March. 1993. The myopia of learning. *Strategic Management Journal* 14: 95–112.
Levitt, B., and J.G. March. 1988. Organizational learning. *Annual Review of Sociology* 14: 319–340.
March, J.G. 1991. Exploration and exploitation in organizational learning. *Organization Science* 2: 71–87.
Nelson, R.R., and S.G. Winter. 1982. *An evolutionary theory of economic change*. Cambridge, MA: Belknap Press of Harvard University Press.
Ocasio, W. 1998. Towards an attention-based view of the firm. *Strategic Management Journal* 18: 187–206.
Siggelkow, N., and D.A. Levinthal. 2003. Temporarily divide to conquer: Centralized, decentralized, and reintegrated organizational approaches to exploration and adaptation. *Organization Science* 14: 650–669.
Song, J., P. Almeida, and G. Wu. 2003. Learning-by-hiring: When is mobility more likely to facilitate interfirm knowledge transfer? *Management Science* 49: 351–365.
Taylor, A. 2010. The next generation: Technology adoption and integration through internal competition in new product development. *Organization Science* 21: 23–41.
Tzabbar, D. 2009. When does scientist recruitment affect technological repositioning? *Academy of Management Journal* 52: 873–896.
Winter, S.G. 2003. Understanding dynamic capabilities. *Strategic Management Journal* 24: 991–995.

Competition

Anne Marie Knott
Washington University in St Louis, Olin Business School, St Louis, MO, USA

Abstract

There are two forms of competition relevant to strategists. Static competition, which receives the bulk of attention, utilizes traditional ▶ industrial organization (IO) economics and game theory to link industry structure, firm behaviour and profits. While conditions exhibiting positive profits are a win for strategists (Porter's five forces), they are a loss for society (inefficient resource allocation) and are relatively rare (<4% of firms). The more prevalent form (>52% of firms) is dynamic or ▶ innovation competition (Porter's diamond). This dynamic competition occurs in structures which otherwise appear perfectly competitive. Moreover, it provides a win-win for firms (higher market value) as well as society (economic growth).

Definition Competition involves the structures and processes by which firms' behaviours determine market prices and the allocation of resources.

There are two forms of competition relevant to strategists. *Static competition* characterizes settings in which firms earn sustainable profits largely by virtue of their initial position. In contrast, *dynamic competition* characterizes market structures in which firms constantly fight merely to sustain profits. This entry discusses each of these separately, then draws inferences from the comparison.

Static Competition

A discussion of competition for a strategy audience is necessarily different from that for an economics audience. The latter is concerned with the link between structural conditions and price-cost margins for purposes of achieving allocative efficiency. If firms are experiencing textbook perfect competition, CEOs will need to price at marginal cost; under monopoly they price such that marginal revenue equals marginal cost; and under oligopolies their price is defined by Nash equilibria that vary with timing, commitment devices and demand heterogeneity. Strategic managers, in contrast, are interested in raising price-cost margins above competitive levels.

Conveniently, ▶ michael porter translated ▶ industrial organization economics into a set of heuristics to evaluate industry attractiveness. Using his five forces framework (1982), strategists can gauge the likelihood of achieving sustainable profits in a given industry. If *buyers*, *suppliers* and *rivals* are numerous and homogeneous, if *substitutes* are strong (high cross-price elasticity) and *potential entrants* unconstrained (low entry barriers), then industry behaviour will be perfectly competitive. (Note these Porter conditions match the standard definition of perfect competition in microeconomics texts: no buyer or seller can affect

price.) At the other extreme, if *buyers* and *suppliers* are again numerous and homogeneous; if *substitutes* are weak (low cross-price elasticity) and *potential entrants* constrained (via high entry barriers such as no alternative sources for a key input), then a single producer faces the entire demand curve and enjoys monopoly profits.

Both these extremes are relatively rare, so Porter derives his power from gauging where in the spectrum between perfect competition and monopoly a given industry lies. The sense conveyed in texts is that CEOs should characterize each of the five forces then construct some form of weighted average to assess where in the continuum their prospective industry lies. In fact, the appropriate use of the framework is to identify the factors, then construct a game to derive equilibrium structure and behaviour. This is quite difficult beyond the duopoly case.

What we know from the duopoly case, however, is:

1. Industries with undifferentiated products, where capacity can be expanded quickly and inexpensively (Bertrand competition) are likely to produce price competition and zero profits.
2. Having capacity constraints (Cournot competition) supports prices above marginal cost.
3. Product differentiation, either artificial (Bertrand differentiation) or through heterogeneous buyer tastes (Hotelling differentiation) is a mechanism for suppressing price competition.
4. Moving first in settings with either capacity constraints (Stackelberg competition), switching costs or product differentiation (Hotelling differentiation) can produce permanent share differences to first movers and sustainable profits for both the first mover and the follower.
5. In some cases the monopoly choices of price, configuration and output are also the best preemptive choices. In those cases there is no cost to behaving 'strategically'.

The empirics tend to support these expectations. For example, Bresnahan and Reiss (1990) show that isolated markets of homogeneous products are perfectly competitive with merely two firms (matching expectations from Bertrand competition). Borenstein et al. (1997) find that New Jersey electricity markets exhibit Cournot equilibria with prices substantially higher than perfectly competitive ones. Caruana and Einav (2008) find that production targets of the big three car manufacturers conform to expectations under 'Stackelberg warfare': each attempts to commit to high production levels early to obtain a Stackelberg leadership position, but as they approach the production horizon they decrease plans as a best response to opponents' high production targets.

The implicit prescription Porter draws from his five forces framework is that firms should choose concentrated industries to insulate themselves from competitive forces. Once they have chosen such an industry, they can rest on their laurels. Indeed, this prescription has merit. In a study of all publicly traded US firms, Hou and Robinson (2006) showed that firms in the top quintile of industry concentration had triple the accounting profits of those in the lowest quintile.

There are three concerns with this framework, however. First, as Lippman and Rumelt (1982) show, concentration is also an outcome of heterogeneous and causally ambiguous cost functions even when firms are price-takers. If that is true, then attempting to enter the industry de novo will most likely end in failure. Thus the only way to enjoy incumbent profits is through acquisition (which, of course, is priced at the net present value of those profits: Rumelt and Wensley 1981). Second, these settings are rare: firms in these industries comprise only 3.9% of firms in the US economy. Third, the US landscape is littered with firms that rested on their laurels (big steel, big auto and legacy airlines to name a few). Firms insulated from competition fall victim to firms who are not. Besides, the prescription is not interesting because there is no role for CEOs or their strategies once they have identified and entered the industry.

Dynamic Competition

Accordingly, the more important form of competition for strategists is dynamic competition.

Interestingly, Porter has a framework and prescriptions here as well (Porter 1990), but these are diametrically opposed to those in his earlier work (Porter 1980). Moreover, whereas his 1980 work is derived from industrial organization (IO) economics (yet turning it on its head), the newer work is entirely inductive. It is based on an in-depth case study of 500 firms across ten of the top trading nations, to understand why firms based in a particular nation come to dominate international markets in particular industries, such as robotics in Japan, pharmaceuticals in the US, cars in Germany.

Porter synthesized his observations from the case studies in the 'Four diamond determinants' of national advantage: (1) factor conditions – but not the natural resources of trade theory, rather advanced factors such as communications infrastructures; (2) demand conditions – diverse, sophisticated and powerful buyers with stringent needs force firms to perceive and satisfy new opportunity; (3) related and supplier industries that determine the pool of technology upon which firms in the focal industry can draw for ▶ innovation; and (4) firm strategy, structure and rivalry – large number of sophisticated rivals forces firms to create new advantages as fast as rivals replicate old ones. Note these observations are very similar to those in the work of Saxenian (1996), comparing the successful cluster of electronics firms in Silicon Valley to its less successful counterpart along Route 128 outside Boston. So geography is not strictly national.

While Porter's framework was derived inductively, his conclusions are anticipated by prior economic theory (outside IO) as well as management theory. The earliest discussion of dynamic competition is perhaps that by Schumpeter (1994), who described a process of creative destruction where firms compete not at the margin, but by displacing existing products and methods of production with newer, more effective ones. More recent theory in evolutionary economics, such as Nelson and Winter (1982), characterizes the dynamic competition that drives emerging industries towards concentration. Here firms are endowed with a cost function, but in each operating period they choose output as well as whether to innovate (with stochastic outcome). If they do not innovate they imitate last period's best practice. Under these conditions early entrants achieve durable scale and cost advantages that ultimately make later entry unattractive. The industry consolidates in a concentrated structure wherein innovation ceases. Note this equilibrium is similar to Lippman and Rumelt's equilibrium of persistent profits among price-taking firms (no market power). What differs from Lippman and Rumelt's work is the innovation dynamics, but these cease once the industry reaches steady-state.

In contrast, what Porter and Saxenian seem to characterize is a setting in which rivalry and innovation continue in steady-state. The theories coming closest to matching this form of competition are endogenous growth within macroeconomics and ▶ dynamic capabilities within the management literature.

Endogenous growth theory comprises stochastic models that cast innovation by profit-seeking firms as engines of growth. These models characterize knowledge as an intermediate good produced by profit-maximizing firms through imitation and invention. A major distinction between these models and IO models of innovation are the assumptions that: (1) firms differ in their level of knowledge, which determines the amount of imitation, and (2) firms are not forward-looking – they innovate even if there is a high likelihood their R&D investments will be expropriated. Thus, in these models, firms invest in R&D principally to regain eroded advantage rather than to pursue the new frontier (Aghion et al. 2001). Empirical tests of these models (Knott and Posen 2009) indicate that the intensity of this 'erosion-innovation cycle' is accelerated by capital intensity, market size, market growth, the number of rivals and the ease of expropriating spillovers.

While the endogeneous growth literature deals with 'representative firms', the management literature exploits firm differences and examines how these are maintained over time under intense rivalry. One piece that bridges economics and strategy is Adner and Levinthal (2001). These authors model innovation alternately by a monopolist and duopolists under demand heterogeneity.

They find monopolists cease innovating once the reservation utility of the marginal consumer has been satisfied. In contrast, under duopoly performance and price both improve in steady-state. Maintaining competitive position in their model is relatively straightforward – there is only one dimension of performance firms seek to improve. The real world seems to have more than one dimension. Accordingly, the dynamic capabilities literature (Teece et al. 1997) examines the complex process of maintaining competitive advantage under intense multidimensional rivalry. The theory makes it clear that substantial effort is required merely to maintain a competitive position. There is no resting on laurels as there is in static competition.

The implicit prescription for dynamic competition differs markedly from static competition. Rather than insulate themselves from competitive forces, firms wishing to preserve profits in the long run should actively seek vigorous competition. Given the divergence, it is worth asking which prescription has greater merit.

Interestingly, both strategies have merit. They differ, however, in the metric of merit. As mentioned previously, firms in insulated industries have higher accounting profits. What distinguishes the firms in competitive industries is that they have higher market returns. Firms in the least concentrated industries have average monthly returns 0.26 percentage points higher than those in most concentrated industries. Thus accounting profits come at the expense of market returns (and vice versa).

In sum, there are two forms of competition relevant to strategists. The one receiving the bulk of attention is *static* competition. The prescription in these settings is for firms to insulate themselves from competitive forces. While this prescription is a *win* for strategists, it is a *loss* for economists (inefficient resource allocation). Moreover, these conditions are relatively rare (comprising less than 4% of firms).

The more important form of competition is *dynamic* competition or innovation competition. The prescription in these settings is to seek out tough rivals and demanding customers and actively compete through innovation. There are three nice features of this alternative form of competition. First, the associated industries comprise the bulk (>52%) of firms. Second, their structures match those of perfect competition (thus achieving allocative efficiency). Third, they offer a *win* for firms (higher market value) as well as *win* for economists (economic growth).

See Also

- ▶ Dynamic Capabilities
- ▶ Industrial Organization
- ▶ Innovation
- ▶ Porter, Michael E. (Born 1947)
- ▶ Technology Strategy

References

Adner, R., and D. Levinthal. 2001. Demand heterogeneity and technology evolution: Implications for product and process innovation. *Management Science* 47: 611–628.

Aghion, P., C. Harris, P. Howitt, and J. Vickers. 2001. Competition, imitation and growth with step-by-step innovation. *Review of Economic Studies* 68: 467–492.

Borenstein, S., J. Bushnell, and C. Knittell. 1997. A Cournot-Nash equilibrium analysis of the New Jersey electricity market. New Jersey Board of Public Utilities. Review of Public Utilities Restructuring Petition.

Bresnahan, T., and P. Reiss. 1990. Entry in monopoly markets. *Review of Economic Studies* 57: 531–553.

Caruana, G., and L. Einav. 2008. Production targets. *RAND Journal of Economics* 39: 990–1017.

Hou, K., and D. Robinson. 2006. Industry concentration and average stock returns. *Journal of Finance* 61: 1927–1956.

Knott, A.M., and H. Posen. 2009. Firm R&D behavior and evolving technology in established industries. *Organization Science* 20: 352–367.

Lippman, S., and R. Rumelt. 1982. Uncertain imitability: An analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics* 13: 418–438.

Nelson, R., and S. Winter. 1982. *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.

Porter, M. 1980. *Competitive strategy*. New York: Free Press.

Porter, M. 1990. *Competitive advantage of nations*. New York: Free Press.

Rumelt, R., and R. Wensley. 1981. In search of the market share effect. Paper presented at Proceedings of the 41st Meeting of the Academy of Management.
Saxenian, A. 1996. *Regional advantage: Culture and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.
Schumpeter, J. 1994. *Capitalism, socialism and democracy*. London: Routledge.
Teece, D.J., G. Pisano, and A. Shuen. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* 18: 509–533.

Competition Policy

Myong-Hun Chang
Cleveland State University, Cleveland, OH, USA

Abstract

This article discusses competition policy in the United States and the European Union. A brief overview of the historical and institutional background is provided, followed by a review of the scholarly research on competition policy that has accumulated in the field of industrial organization. The emphasis is placed on the role that economic and strategic analyses have played in the development of competition policy in its modern form.

Definition Competition policy, also known as antitrust policy in the United States, is a body of legislated law designed to promote and maintain competition in markets.

► Competition policy, also known as antitrust policy in the United States, is a body of legislated law designed to promote and maintain competition in markets. While competition policies have historically been limited to controlling the behaviour of firms in markets within a country's own border, the growth of multinational corporations and international trade has necessitated taking a global perspective on their implementation. This article offers a brief description of competition policies in two major economies, the United States and the European Union (EU), followed by an overview of the scholarly research on such policies.

Theoretical Basis

The free market ideal is perfect competition, which satisfies four conditions: (1) there exists a large number of small buyers and sellers; (2) the product is homogeneous; (3) all buyers and sellers have full information about the available prices and the nature of the product; (4) there is freedom of entry and exit for the producers. Under these conditions, a market transaction takes place at a quantity where the price of a unit of the good is equal to its marginal cost of production, and available resources are allocated to the uses that generate the highest value for society; hence allocative efficiency is achieved.

Violation of one or more of these conditions could lead to misallocation of resources and, thus, a loss of potential value for society (often called 'deadweight loss'). Competition policy is motivated by the recognition of this potential market failure and the resulting deadweight loss. The failure that has received most attention is an 'insufficient' number of sellers. The extreme version is the case of a monopoly. A monopolist, by restricting the quantity below the competitive level, causes the price of the good to rise above its marginal cost of production and an inefficient allocation of resources.

Most markets lie between perfect competition and monopoly. In general, firms in a market with fewer firms tend to find it easier to collude, where explicit or implicit coordination of firms' production decisions leads to high prices for consumers and inefficient allocation of resources. Competition policy aims to reduce or eliminate such inefficiency by altering the structural features of the market or restricting the anticompetitive conduct of the firms.

Historical and Institutional Overview

In the US, the prosecution of antitrust law violations is carried out under the auspices of three major antitrust statutes: the Sherman Act of 1890, the Clayton Act of 1914 and the Federal Trade Commission (FTC) Act of 1914. The introduction of the Sherman Act was facilitated by the

growing populist sentiment against big business, which can be traced back to the severe recessions during the 1870s and the 1880s that resulted in fewer surviving firms with greater ▶ market power.

Section 1 of the Sherman Act prohibits firms from conspiring to engage in practices against public interest, while Section 2 addresses the problem of market monopolization. Because the Sherman Act emphasizes punishment rather than prevention, the Clayton Act and the FTC Act were introduced in 1914 to check monopoly in its incipiency. Sections 2 and 3 of the Clayton Act address price discrimination and tying practices (including bundling, exclusive dealing and other vertical restraints), while Section 7 focuses on restricting merger activity. These practices are deemed illegal only if they *substantially* lessen competition or create a monopoly. The FTC Act established the Federal Trade Commission in order to evaluate and enforce antitrust policies before going to court.

Antitrust law in the European Union (EU) addresses the issues of cartels and monopoly in Articles 101–109 of the Treaty on the Functioning of the European Union (TFEU). The constitutional basis of the European Union consists of two treaties: the Treaty on European Union (TEU; 'Maastricht Treaty') and the Treaty establishing the European Community (TEC; 'Treaty of Rome'). The Treaty of Lisbon (signed, 2007; in force, 2009) amended these two treaties, in the course of which TEC was renamed as the Treaty on the Functioning of the European Union (TFEU). As such, articles 101–109 of TFEU correspond to articles 81–89 of TEC with all the relevant modifications. Article 101 of TFEU prohibits collusion and anticompetitive practices and is comparable to Section 1 of the Sherman Act. Article 102 of TFEU is the counterpart to Section 2 of the Sherman Act. Article 102 of TFEU is also used in testing merger cases by asking whether the combined firms will create or reinforce a dominant firm in a given market. The control of merger activities that significantly impede competition in a market within a member state is addressed in Articles 2 and 3 of the European Commission's (EC's) Merger Regulation.

In the US, antitrust laws are enforced by the Antitrust Division within the Department of Justice (DOJ) and the Federal Trade Commission (FTC). The equivalent position in the EU is the Directorate General for competition (DG Comp) of EC. The federal enforcement in the US is supplemented by state-level antitrust statutes as well as private antitrust law suits. Similarly, the competition laws of the EU's member states supplement the articles in TFEU. However, private enforcement actions are rare in the EU.

Enforcement of Competition Law

Three broad types of anticompetitive conduct are subject to antitrust enforcement: (1) price-fixing, bidrigging or cartel formation; (2) an attempt to gain a monopoly for the purpose of exercising market power; (3) mergers that create excessive market power.

In cases of cartels and monopolization, the DOJ, the FTC or private parties can bring lawsuits. While the Sherman Act allows the DOJ to pursue criminal charges, the DOJ more often brings civil suits that seek injunctions. Private treble damage lawsuits can also be brought. The prevention of potentially anticompetitive mergers is enforced through lawsuits brought by the DOJ or the FTC; the objective of such a suit is an injunction to prevent the merger from proceeding. The FTC does not have the authority to pursue criminal charges.

The standard of proof can differ among the types of anticompetitive conduct. For mergers and monopolization the *rule-of-reason* standard is typically used. Successful prosecution requires extensive evidence based on a social cost–benefit analysis. Under this standard, there is a high resource cost imposed on both the plaintiff and the defendant due to the uncertainty as to what activities are violations. The other standard applied by the court is the per se standard which judges certain activities illegal, regardless of economic effect. One only needs to show that the act has been committed. An example is cartel behaviour

involving price-fixing, bid-rigging, market division or restriction of output.

The enforcement mechanism can be either public or private. Public enforcement entails fines, imprisonment or structural change, as in the cases of Standard Oil (1911) and AT&T (1982). In the US, private enforcement entails treble damage, which is three times the damage measured as the excess payments made by customers over what the prices would have been in the absence of the conspiracy. In the EU, the recovered amount tends to approximate the actual loss incurred, and this limits the incentive for private litigation. In addition, the burden of proof often rests with the plaintiff, thus significantly raising the cost of discovery.

Economic Analysis of Competition Policy

Modern research has made a substantial contribution to the development of competition policy in three areas: understanding the behaviour of cartels; identifying the market conditions under which predatory or exclusionary practices can be a rational strategy; and providing an analytic framework for evaluating mergers. This section summarizes the basic issues in these categories.

Cartels

Explicit cartels are illegal per se in many countries. As such, most existing cartels are tacit: the agreement must be self-enforcing in that each firm must have an economic incentive to abide by it. According to the economic theory of tacit collusion, this requires that, for each firm, the present value of discounted profits from adhering to the cartel agreement must exceed that from deviating (and, hence, destroying the cartel arrangement):

$$\frac{1}{1-\delta}\pi^C \geq \pi^D + \frac{\delta}{1-\delta}\pi^N$$

where π^C is the per-period profit to a firm under the cartel agreement, π^D is the one-time profit to optimally deviating from the agreement, π^N is the per-period profit earned by a firm when the industry reverts to oligopolistic competition (following dissolution of the cartel) and δ is the discount factor.

Note that the various profit levels are functions of the structural features of the market such as the number of firms, demand conditions and technological conditions. The past theoretical literature has focused on establishing the relationships between the various structural parameters and the degree of cartel stability (Levenstein and Suslow 2006). This line of research informs the antitrust authorities which types of industries are more or less conducive to the formation and maintenance of cartels.

A more recent line of research asks what observable patterns in firm behaviour may signal current or past cartel activity, hence turning the focus on *detecting* existing cartels. Based on detailed case studies, Harrington (2006) proposes a variety of collusive markers which can be used as a signal of collusive behaviour. A related issue, arising from the 'tacit' nature of cartels, is that we only get to observe cartels that are discovered; we do not observe cartels that are formed and stay undetected. Given that the detection activities affect the incentives of firms to form cartels, how do we evaluate the effectiveness of the policy if we only observe those cartels that are discovered? Harrington and Chang (2009) address this issue by modelling a population of heterogeneous industries in which cartels can be created and dissolved on the basis of stochastic market conditions as well as the detection activities of the antitrust authority. The enforcement is explicitly modelled and influences the stability of cartels. The time-series behaviour of cartels in this model provides markers of discovered cartels that can be used to *infer* the impact of competition policy on the population of all cartels. Chang and Harrington (2010) use this approach to evaluate the effectiveness of the Corporate Leniency Program.

Prior to the early 1990s, most cartels involved firms from a single country. The recent globalization of industries and the increasing number of

multinational corporations, however, have resulted in many cartels involving participants from multiple nations. Connor (2008) reports that the number of foreign defendants in US criminal cartel cases rose from less than 1% before 1995 to 40–70% after 1995. The magnitude of punishment has also increased dramatically during this period, reflecting a major shift in antitrust enforcement. The largest price-fixing fine prior to 1994 was $2 million (imposed on purely domestic cartels). Between 1994 and 1999, the record fine kept increasing each year, ultimately reaching $500 million in 1999 (imposed on Hoffmann-La Roche as part of their global vitamins cartel).

One cause of the increased likelihood of detection and prosecution of cartels is the Corporate Leniency Program, where the cartel participant that cooperates with the enforcement agency receives amnesty. The US DOJ instituted the programme in 1978 (and revised it in 1993 and 1994). The success of this programme has led to the adoption of similar programmes in the EC in 1996, soon followed by others, including Canada, UK, Japan, South Africa and Brazil. Connor (2008) reports that at least 300 international cartels have been discovered by authorities since 1990 and almost half of them since 2000. While the positive impact this programme has had on the rate of discovery is clear, its impact on the rate of *cartel formation* is not and remains a subject of ongoing research (Chang and Harrington 2010).

Monopolization

Anticompetitive conduct includes an array of strategies to drive out rivals or to deter possible entrants. To the extent that these actions are successful, they have efficiency implications and are subject to anti-trust enforcement. Two types of anticompetitive practices in this category have been identified: (1) predatory pricing and (2) exclusionary practices.

Traditionally, predatory pricing involves a predator firm increasing its output with the intention of driving down the market price to impose losses on its rivals. A sustained period of losses eventually forces the rivals to exit the market. With the newly acquired monopoly position the predator firm reduces its output and raises the price to the monopoly level. As long as the present value of the extra profits enjoyed by the monopolist exceeds that of the extra losses incurred during the period of predatory pricing, it is rational for the firm to engage in the predatory strategy.

However, two considerations pose a significant challenge to the original theory: (1) the predator firm is typically of a larger size than its intended victims, and has more to lose during the period of predation than the rivals; (2) the monopoly profits expected after the rivals' exits are not secure as the price increase is likely to invite new entries. McGee (1958) concluded that predatory pricing is rarely a rational strategy and the observed pricing simply reflects efficiency differences between these firms. But recent game-theoretic analyses have shown that, under certain conditions, predatory pricing can be an equilibrium strategy (see Ordover and Saloner 1989, for a survey). For instance, firms operating in multiple markets may have an incentive to create the reputation of being 'aggressive' in one market, if the reputation thus created can be carried over to other markets. Another case is when the predator firm has better access to financial resources than its victims. Yet another possibility is that the leading firm may strategically use its price as a signal of its lower cost, when the rival firms are uncertain about the cost level of the predator. This may convince rivals to alter their competitive behaviour or simply move to another market.

Exclusionary practices are used by a dominant firm or a monopolist incumbent to exclude firms from entering its market. Tactics include tie-in sales or bundling, exclusive dealings, long-term contracts with buyers and ownership of essential inputs. Many of these exclusionary tactics are used in the context of vertically related markets – that is, output markets and input markets – where market power in one market can be leveraged to obtain market power in another. Although there is a large volume of recent literature supporting the rationality of such behaviour, the debate on this issue remains inconclusive.

Enforcement against predatory and exclusionary practices has been difficult for the court because of the lack of success in developing general operational rules for distinguishing predation from competition. Instead, the court has relied more on the quantitative rule based simply on prices and costs, as suggested by Areeda and Turner (1975). The impact of economic theory has, thus far, been limited.

Mergers

In the US, proposed mergers, acquisitions and joint ventures are reviewed by the DOJ and FTC. Under the Hart–Scott–Rodino Act of 1976, parties to a merger or acquisition, meeting certain dollar thresholds, are required to file pre-merger notification reports with both of the agencies. These filings are followed by a prescribed waiting period before the transaction is consummated. The review process allows DOJ/FTC to evaluate the merits of the proposed deals and challenge them if necessary.

The analytical framework and the specific standards used to review the merger proposals are described in the Horizontal Merger Guidelines issued jointly by DOJ and FTC. The main purpose of the review process is to examine whether a proposed transaction will confer market power upon the newly merged entity. This requires properly defining the product/geographic market for the merger participants and any relevant competitors. The competitiveness of the market, both pre- and post-merger, is measured by a concentration index. Agencies use the Herfindahl–Hirschman Index (H-index) as the concentration measure, where it is defined as the sum of the squared market shares of all firms in the given market. Whether a proposed merger is challenged or not depends on both the pre-merger concentration as well as the 'increase' in the concentration that the proposed merger will induce. Weighed against the potential inefficiency from increased market power is the economic gain attained through economies of scale and scope (which may ultimately benefit consumers). Finally, the Guidelines recognize the possibility that relatively easy entry may quickly erode any market power that the merged firm may have captured.

Due to the complexity in the jurisdictional division between the EC and the member states, merger control in the EU starts with a pre-notification consultation in which the EC's jurisdiction over the proposed merger is examined. Upon the confirmation of its authority, the EC commences the merger review process. The legal and analytical framework surrounding the review process is provided in the EC Merger Regulation No. 139.2004 of 2004. The economic analysis driving the review is similar to that described in the US Merger Guidelines.

Conclusion

Over the last 30–40 years, the traditional, politically motivated antitrust policy has been replaced by a competition policy motivated by rigorous economic and strategic analyses of firm behaviour. The advances made in game-theoretic modelling have contributed to understanding the strategic behaviour of firms and have led to the reformulation of competition policy. The impact of these advances on the decisions of the court has been limited because of the difficulty in transforming theoretical insights into a set of quantifiable rules that can guide the court's decisions.

See Also

- ▶ Competition
- ▶ Industrial Organization
- ▶ Market Power

References

Areeda, P., and D. Turner. 1975. Predatory pricing and related practices under section 2 of the Sherman Act. *Harvard Law Review* 88: 697–733.

Chang, M-H., and J.E. Harrington. 2010. *The impact of a corporate leniency program on antitrust enforcement and cartelization*. Working paper.

Connor, J.M. 2008. *Global price fixing*. Berlin: Springer.

Harrington, J.E. 2006. How do cartels operate? Foundations and trends. *Microeconomics* 2: 1–105.

Harrington, J.E., and M.-H. Chang. 2009. Modeling the birth and death of cartels with an application to evaluating competition policy. *Journal of the European Economic Association* 7: 1400–1435.

Levenstein, M., and V. Suslow. 2006. What determines cartel success? *Journal of Economic Literature* 44: 43–95.

McGee, J. 1958. Predatory price cutting: The standard Oil (N.J.) case. *Journal of Law and Economics* 1: 137–169.

Ordover, J.A., and G. Saloner. 1989. Predation, monopolization, and antitrust. In *Handbook of industrial organization*, ed. R. Schmalensee and R.D. Willig. Amsterdam: North-Holland.

Competitive Advantage

Catherine A. Maritan[1] and Margaret Peteraf[2]
[1]Syracuse University, Whitman School of Management, Syracuse, NY, USA
[2]Dartmouth College, Department of Management, Tuck School of Business, Hanover, NH, USA

Abstract

We define competitive advantage as *the degree to which a firm creates more economic value than rival firms in a given product market*, noting that economic ▶ value is the difference between the value perceived by the firm's customers, measured by their willingness to pay, and the firm's costs. This definition is applicable whether the advantage is sustainable (Barney, J Manag 17:99–120, 1991) or whether it is merely a temporary advantage (D'Aveni, R.A. *Hypercompetition: Managing the dynamics of strategic maneuvering*. New York: Free Press, 1994). As we explain, it is consistent with the ▶ resource-based view of the firm (Wernerfelt, Strat Manag J 5:171–180, 1984) and the ▶ positioning view (Porter, M.E. *Competitive strategy*. New York: Free Press, 1980), as well as the bargaining perspective (Brandenburger and Stuart, J Econ Manag Strat 5:5–24, 1996).

Definition Competitive advantage is the degree to which a firm creates more economic value than rival firms in a given product market.

Competitive advantage is a business-level construct that has been defined in several different ways. It has been used to signify superior financial performance at the market level, thus connoting a business-level performance advantage. It has also been used to refer to superior resources and distinctive competencies, thereby connoting a resource and capability advantage. Competitive advantage is best understood, however, neither as a financial outcome nor as an underlying firm attribute, but, rather, as an intermediate performance outcome, indicative of differences in the degree to which different competitors create ▶ value (e.g., Porter 1985; Barney 2007; Walker 2009; Ghemawat 2010; Grant 2010; Thompson et al. 2011).

Accordingly, we define competitive advantage as *the degree to which a firm creates more economic value than rival firms in a given product market* (Peteraf and Barney 2003). This definition is applicable whether the advantage is sustainable (Barney 1991) or whether it is merely a temporary advantage (D'Aveni 1994). It is consistent with a ▶ resource-based view of the firm (Wernerfelt 1984) as well as with a ▶ positioning view (Porter 1980) and a bargaining perspective (Brandenburger and Stuart 1996).

To facilitate a deeper understanding of this definition and enable the construct's measurement for empirical purposes, it is necessary to give precise meaning to the term 'economic value'. Businesses create economic value when they transform inputs (of a certain value or cost to the firm) into more highly valued goods or services. Thus, the economic value of a firm's goods or services is measured by the difference between the value perceived by the firm's customers, as measured by their willingness to pay, and the firm's costs. This suggests that there are two basic ways in which a firm can create more economic value than its market rivals (and thus gain a competitive advantage over them). It can pursue a differentiation advantage to increase customer willingness to pay or it becomes

relatively more efficient with lower costs (or it can do both).

Competitive Advantage and the Positioning Perspective

At this point, the link between a value-based definition of competitive advantage and the positioning perspective should be apparent. Any one of the basic ▶ generic strategy (Porter 1980) may provide a route to competitive advantage, and managers may utilize a selection of cost drivers and/or value drivers (differentiation drivers) to improve their relative position. At a more fine-grained level, a firm's competitive advantage depends on how well it performs specific activities along its value chain in comparison with its rivals (Porter 1985). At this level, it is possible to pinpoint both the sources and size of a firm's competitive advantage (or disadvantage). See, for example, Ghemawat and Rivkin 2010.

Competitive Advantage and the Resource-Based View

Because a firm's resources and capabilities may constitute the ultimate source of a firm's competitive advantage, there is also a clear link between competitive advantage, defined in terms of economic value created, and the resource-based view (Peteraf and Barney 2003). While economic value is created in the course of performing various activities along the firm's value chain, the firm's resources and capabilities are deployed in the process, and enable the firm to perform its activities at a certain level of efficiency and effectiveness. Thus, the relationship between a firm's value-chain activities and its resource and capability set is a close one.

Competitive Advantage and the Bargaining Perspective

There is also a clear connection between competitive advantage as defined above and the bargaining perspective, given the concern of this perspective with the creation and distribution of value within an industry (Brandenburger and Stuart 1996; Hoopes et al. 2003; Lippman and Rumelt 2003). Brandenburger and Stuart (1996) introduce the notion of a firm's 'added value', which is the amount by which the economic value created by the firm exceeds that of its closest market rival. This is a measure of the firm's competitive advantage, as defined above, although it brings into question the issue of which rival (or rivals) should be used as the comparison point in determining a firm's competitive advantage. The comparison may be made in various ways, including the strongest rival, the average rival in the market or the marginal competitor.

Competitive Advantage and Profitability

It is important to recognize that a value-based definition of competitive advantage is not synonymous with superior financial performance. Indeed, a benefit of conceptualizing competitive advantage from a value-based perspective is that it provides some needed separation between firm characteristics (e.g., resources, capabilities, activities, bargaining position) and profitability outcomes. As Coff (1999) has explained, neither superior capabilities nor superior value creation guarantee that a firm can turn its value advantage into greater profitability, since other stakeholders may claim a share of the value generated. Brandenburger and Stuart (1996) show the same is true when competitive advantage is defined in terms of added value, since the distribution of the value created depends upon the relative bargaining power of various claimants. In the simplest case, there may be a link between competitive advantage and superior profitability, but this link is obscured because the firm simply may forgo current period profits to invest in research and development, market share or customer loyalty in anticipation of reaping future benefits.

See Also

- ▶ Competitive Strategy
- ▶ Competitiveness
- ▶ Generic Strategy
- ▶ Positioning
- ▶ Resource-Based View
- ▶ Sustainable Competitive Advantage
- ▶ Value

References

Barney, J.B. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17: 99–120.

Barney, J.B. 2007. *Gaining and sustaining competitive advantage*. Upper Saddle River: Pearson Education.

Brandenburger, A.M., and H.W. Stuart Jr. 1996. Value-based business strategy. *Journal of Economics & Management Strategy* 5: 5–24.

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

D'Aveni, R.A. 1994. *Hypercompetition: Managing the dynamics of strategic maneuvering*. New York: Free Press.

Ghemawat, P. 2010. *Strategy and the business landscape*. Upper Saddle River: Prentice Hall.

Ghemawat, P., and J.W. Rivkin. 2010. Creating competitive advantage. In *Strategy and the business landscape*, ed. P. Ghemawat. Upper Saddle River: Prentice Hall.

Grant, R.M. 2010. *Contemporary strategy analysis*. Chichester: Wiley.

Hoopes, D.G., T.L. Madsen, and G. Walker. 2003. Guest editors' introduction to the special issue: Why is there a resource-based view? Toward a theory of competitive heterogeneity. *Strategic Management Journal* 24: 889–902.

Lippman, S.A., and R.P. Rumelt. 2003. A bargaining perspective on resource advantage. *Strategic Management Journal* 24: 1069–1086.

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

Porter, M.E. 1980. *Competitive strategy*. New York: Free Press.

Porter, M.E. 1985. *Competitive advantage*. New York: Free Press.

Thompson, A.A., M.A. Peteraf, J.E. Gamble, and A.J. Strickland. 2011. *Crafting and executing strategy: The quest for competitive advantage*. New York: McGraw-Hill.

Walker, G. 2009. *Modern competitive strategy*. New York: McGraw-Hill.

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

Competitive Heterogeneity

Tammy L. Madsen[1], David G. Hoopes[2] and Gordon Walker[3]

[1]Santa Clara University, Leavey School of Business, Santa Clara, CA, USA

[2]California State University Dominguez Hills, Department of Management, College of Business Administration and Public Policy, Carson, CA, USA

[3]Southern Methodist University, The Edwin L Cox School of Business, Dallas, TX, USA

Definition Competitive heterogeneity refers to *enduring* and *systematic* (superior) *differences* in strategic positioning, and, in turn, performance, among relatively close rivals where a firm that produces the largest gap between the value (V) of a good or service to a buyer and the cost (C) of producing that value holds an advantage, or superior position, relative to that of rivals.

Strategy scholars tend to agree that positioning for ▶ competitive advantage relates to how value is created and captured. Despite extensive theoretical work on competitive advantage, empirical work has not evolved in a similar cumulative way. In response, the concept of competitive heterogeneity emerged with the intent of providing a more complete picture of the emergence and persistence of performance heterogeneity among close competitors (Hoopes et al. 2003). Similar to the ▶ resource-based view (RBV), the competitive heterogeneity logic links resources and capabilities to competitive advantage. However, it broadens the theoretical scope regarding sources of sustainable heterogeneity. Indeed, if resources or capabilities associated with performance heterogeneity are not protectable, then the persistence of an advantage must be associated with something other than costly imitation. In other words, barriers to imitation are one type of isolating mechanism; other isolating mechanisms do not depend on a firm's resources or capabilities. It follows, then, that important sources of value and cost differences among firms may lie outside the RBV's boundaries. As such, work on

competitive heterogeneity has broader theoretical roots than the RBV (see Hoopes et al. 2003).

Competitive Heterogeneity: The Value-Price-Cost (VPC) Framework

To define superior performance differences independently of resources and capabilities, work on competitive heterogeneity adopts a bargaining model. In this framework, a buyer and supplier bargain over the price (P) for a good that contributes a value (V) or benefit to the buyer and costs the supplier some amount (C) to produce. Value is the price a buyer is *willing to pay* for a good in the absence of competing products or services yet within budget constraints and having considered other purchasing opportunities. The good's market price lies between value and cost. So, the buyer receives a surplus of value minus the price (V−P), and the supplier captures a profit of price minus cost (P−C). The supplier's resources and capabilities, in turn, influence the value of the good to the buyer and the cost of producing it. The gap between value and cost is referred to as a firm's VC profile. The firm that produces the largest difference between value and cost has an advantage over rivals. It can either attract buyers due to the better surplus its product offers (V−P), make a higher profit (P−C), or both (strategy texts frequently use the VPC framework to define competitive advantage: see Hoopes et al. 2003, for references). Competitive heterogeneity arises when a firm is able to sustain a larger VC profile relative to that of rivals.

The VPC framework illustrates that simply having costly to imitate resources and capabilities does not necessarily produce a competitive advantage. Consider two firms that have different resources and capabilities and hence offer different value at different costs. Yet, V minus C is the same for both firms. Next, assume industry-wide constraints on bargaining over price. Then the dissimilar resources and capabilities of these firms produce the same economic return (P−C). This means that a resource or capability is valuable only when it increases the difference between a firm's value and cost (V−C) compared with that of its rivals. That is, a valuable resource makes a firm more productive in the context of competition.

Implications

Employing the VPC framework offers several benefits for understanding the sources of competitive heterogeneity. For one, models of sustainable heterogeneity typically focus on differences among firms in efficiency, rarely on differences in customer benefit, and even more rarely on differences in both efficiency and customer benefit simultaneously. Using the VPC framework, however, involves considering both the demand side (V) and the supply side (C) of a transaction. The framework also allows one to separate competitive heterogeneity (and competitive advantage) from a firm's resources and capabilities. For instance, the value (V), a buyer's willingness to pay, is not part of a firm's income statement. Thus, the framework avoids issues of tautology and avoids equating competitive advantage with standard performance metrics. The latter point is critical since extant work often blurs distinctions between value capture and competitive advantage. Although standard and absolute performance metrics might be related to value capture, they are not necessarily related to competitive advantage. Last, the framework allows for a more integrated analysis by explicitly accounting for dynamic sources of heterogeneity. For instance, given competition and bargaining, the value captured by a firm may strengthen or erode over time due to shifts in competition or due to changes in consumers' valuations of the firm's offerings. Thus, competitive heterogeneity is rooted in evolving differences in the resources and capabilities leveraged by close competitors, and in consumer heterogeneity.

See Also

- ▶ Capturing Value from Advantages
- ▶ Competitive Advantage
- ▶ Resource-Based View

References

Hoopes, D., T.L. Madsen, and G. Walker. 2003. Why is there a resource-based view? Toward a theory of competitive heterogeneity. *Strategic Management Journal* 24: 889–902.

Competitive Strategy

J. Myles Shaver
University of Minnesota, Carlson School of Management, Minneapolis, MIN, USA

Abstract

Competitive strategy examines the actions that firms undertake – and the performance outcomes associated with those actions – recognizing the existence of direct competitors and other firms that might appropriate profits. This entry traces the evolution of approaches to this issue from the five forces to resource or capability views. It also highlights approaches in the field to assess the competitive interactions inherent in the study and application of competitive strategy.

Definition Competitive strategy examines the actions that firms undertake – and the performance outcomes associated with those actions – recognizing the existence of direct competitors and other firms that might appropriate profits.

Five Forces Framework

The foundational framework to the study of the competitive strategy in management is the five forces model introduced by ▶ Michael Porter (1979). The framework, which is based on the Bain–Mason or ▶ structure–conduct–performance view from ▶ industrial organization economics, assesses how industry profitability is affected by five competitive forces: bargaining power of buyers, bargaining power of suppliers, threat of substitutes, threat of entry and rivalry among competitors.

This framework has proven very popular because of its ability to systematically provide insight into the sources of competitive pressures on the firms and its ability to identify attractive industries. This framework is taught in almost all introductory strategy courses, in addition to many other courses.

Resource and Capability Views

The insight that differences in competitiveness in the product market, as described in the five forces model, could, in parallel, be considered as representing differences in the input market or resources of firms (Wernerfelt 1984), gives rise to the other influential approach in assessing competitive strategy, the resource-based and capabilities-based views of the firm.

The resource-based or capabilities view of the firm highlight the fact that when ▶ firm resources or capabilities are uniquely possessed (Barney 1991; Rumelt 1984), or when the ability to generate such resources or capabilities is uniquely possessed (also known as the ▶ dynamic capabilities approach – Teece et al. 1997), firms will have advantageous competitive outcomes vis-à-vis rivals.

The focus from this approach directs attention towards resources and capabilities internal to the firm, compared with the five forces approach, which focuses on the structure of the industry. The strategic implication from the resource and capability views is that competiveness stems from developing and deploying valuable firm resources while potentially deterring competitors from doing the same (Clarkson and Toh 2010).

This view is central to the current academic study and teaching of competitive strategy.

Strategic Interactions

Recognizing that the success of a chosen strategy will be contingent on the choices of other firms has led to the application of ▶ game theory as a tool to study competitive strategy. The initial application of game theory to these issues focused

on the application of non-cooperative game theory. This approach, while central in industrial organization economics, also had applications within the strategy literature (e.g., Amit et al. 1990; Saloner 1991).

More recently, cooperative game theory has been used to provide insight into strategic interactions (e.g., Brandenberger and Stewart 1996; MacDonald and Ryall 2004). The benefit of this approach is that it provides insight into competitive interaction without requiring the same assumptions on the underlying structure of interaction as do non-cooperative game theory approaches.

In addition to the theoretical insight from game theoretical approaches, a literature that empirically identifies the actions of competitors and their relationships also exists (e.g., Smith et al. 1991; Ferrier 2001). This literature is often referred to as the competitive dynamics literature.

See Also

- ▶ Cooperative and Non-cooperative Game Theory
- ▶ Dynamic Capabilities
- ▶ Firm Resources
- ▶ Five Forces Framework
- ▶ Game Theory
- ▶ Industrial Organization
- ▶ Mason, Edward Sagendorph (1899–1992) and Bain, Joe Staten Jr (1912–1991)
- ▶ Porter, Michael E. (Born 1947)
- ▶ Resource-Based View
- ▶ Structure–Conduct–Performance

References

Amit, R., L. Glosten, and E. Muller. 1990. Entrepreneurial ability, venture investments, and risk sharing. *Management Science* 36: 1233–1246.

Barney, J. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17: 99–120.

Brandenberger, A.M., and H.W. Stewart. 1996. Value based strategy. *Journal of Economics & Management Strategy* 5: 5–24.

Clarkson, G., and P.K. Toh. 2010. 'Keep out' signs: The role of deterrence in the competition for resources. *Strategic Management Journal* 31: 1202–1225.

Ferrier, W.J. 2001. Navigating the competitive landscape: The drivers and consequences of competitive aggressiveness. *Academy of Management Journal* 44: 858–877.

MacDonald, G., and M. Ryall. 2004. How do value creation and competition determine whether a firm appropriates value? *Management Science* 50: 1319–1333.

Porter, M.E. 1979. How competitive forces shape strategy. *Harvard Business Review* 57: 137–145.

Rumelt, R.P. 1984. Towards a strategic theory of the firm. In *Competitive strategic management*, ed. R.B. Lamb. Englewood Cliffs: Prentice Hall.

Saloner, G. 1991. Modeling, game theory, and strategic management. *Strategic Management Journal* 12- (Special issue: fundamental research issues in strategy and economics): 119–136.

Smith, K.G., C.M. Grimm, M.J. Gannon, and M.-J. Chen. 1991. Organizational information processing, competitive responses, and performance in the U.S. domestic airline industry. *Academy of Management Journal* 34: 1–26.

Teece, D.J., G. Pisano, and A. Shuen. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* 18: 509–531.

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

Competitiveness

Aneel Karnani
University of Michigan, Stephen M Ross School of Business, Ann Arbor, MI, USA

Abstract

Competitiveness defines how a firm outperforms its competitors and creates more value for its owners. Competitiveness consists of a set of integrated choices: the domain in which the firm competes, its competitive position – which in turn consists of the value proposition it offers to customers, and its cost structure – and the ▶ organizational design to execute its strategy. To gain a ▶ competitive advantage, a firm must act differently, making choices and choosing alternatives that are distinct from those of its competitors. The very essence of competitiveness is controversial choices and trade-offs.

Definition Competitiveness defines how a firm outperforms its competitors and creates more value for its owners. Competitiveness consists of a set of integrated choices: the domain in which the firm competes, its competitive position and the organizational design to execute its strategy.

Competitiveness defines how a firm outperforms its competitors and creates more value for its owners. Competitiveness consists of a set of integrated choices: the domain in which the firm competes, its competitive position – which in turn consists of the value proposition it offers to customers, and its cost structure – and the ▶ organizational design to execute its strategy. In other words, a firm's competitiveness is determined by its choices on where to compete, how to compete and strategy execution.

Controversial Choices

To gain a ▶ competitive advantage, a firm must act differently, making choices and choosing alternatives that are distinct from those of its competitors. The very essence of competitiveness is controversial choices and trade-offs; without these factors, every firm would choose the same alternatives and there would be no difference among competitors (Karnani 2008). Einstein is said to have defined insanity as doing the same thing and expecting a different result.

Strategic choices are complicated and controversial. Analyses alone do not yield the right answers; managers have to make difficult judgements, often in the context of considerable uncertainty. Managers who are equally intelligent and competent can have very disparate views on the best strategy for the company. That is why making the choices to achieve competitiveness is the central challenge for top managers.

A Vision Is Not a Strategy

In the lobby of many companies you will find a beautifully framed vision statement. However, if you take that vision statement and hang it in the lobby of a different company, most people would not notice the difference. Most vision statements are 'motherhood and apple pie' statements, platitudes about being the best in terms of quality, service, growth, leadership, innovation, customers, employees and/or shareholders. These statements are often trite, generic and exchangeable, not controversial – and hence they do not lead to competitiveness.

Vision statements are useful for energizing people in a company and providing a common purpose and cohesive values. But they provide very little, if any, guidance for making complicated strategic choices. There is much more to formulating a strategy for competitiveness than devising a vision.

Competitive Domain

Companies, even large companies, do not compete in an industry as a whole. A firm must choose the domain in which it competes. This choice involves four dimensions: markets, segments, vertical scope and geographic scope. An industry can be divided into several markets and the firm can decide to what extent it participates in each market. For example, a home appliance company can choose to what extent to emphasize the markets for refrigerators, freezers, washing machines, dryers, ovens, dishwashers and air conditioners. A market can be further divided into segments, and the firm must decide to what extent to participate in each segment. For example, the market for refrigerators can be segmented in several ways: new home construction versus retail, do-it-yourself versus contractor, and by size. A firm also has to decide whether to compete broadly in several markets and segments, or narrowly in a niche.

Vertical scope determines what activities the firm performs and which ones it outsources. A refrigerator company has a choice to manufacture compressors or to buy them from a supplier. Finally, the firm must decide which countries and geographic regions it will participate in and to what extent. For example, in home appliances, Whirlpool and General Electric are both major players in North America, but while Whirlpool is a significant competitor in Europe, General Electric is not.

Some industries are structurally more attractive than other industries (Porter 1979). Similarly, some parts of an industry are structurally more attractive than other parts. Firms try to choose the competitive domain that is more attractive, based not only on the current structure, but also their projection of future structure.

Value Proposition

Having decided where to compete, the firm next needs to decide how to compete. Its competitive position involves choices on two major dimensions: the value proposition it offers to customers and its cost structure. As firms compete with each other, the customer has a critical say in determining which firms win in this competition. Each firm offers a value proposition, that is, a bundle of product (the term 'product' is used here generically to include services), price, place and promotion – the four 'P's of marketing (Borden 1964).

Product decisions include various aspects such as functionality, quality, performance, features, reliability, conformance, durability, safety, serviceability, aesthetics, convenience and customization. Pricing decisions include terms used to determine the price charged to a particular customer, such as volume discounts, early payment discounts, seasonal pricing, bundling and price discrimination. A good example of the complexity in pricing is the airline industry. Distribution is about how the firm gets the product to the customer and involves decisions on market coverage, order processing, inventory management, logistics and delivery process. Promotion is about how the firm communicates with customers to generate a positive response, and includes choices on brand image, advertising and public relations.

Each customer places a different weighting on the four elements of the marketing mix based on his or her particular preferences and trade-offs. Customers with similar weightings can be grouped together to form a segment. Companies compete by trying to offer a superior, and preferably a unique, value proposition to their targeted customers. Intimate knowledge of customers is needed to design such a value proposition.

Cost Structure

A superior value proposition is essential but not sufficient for a firm to outperform its competitors. A firm also needs a cost structure such that it can 'produce' this value proposition profitably. After all, the firm's objective is to create more value for its owners, or shareholder value. Accounting profits are defined as revenues minus total costs. Shareholder value is related not to accounting profits, but, rather, to economic profits, which are equal to accounting profits minus the opportunity cost of the capital used by the firm. Put differently, competitiveness is defined by the firm's ability to earn higher economic profits than its competitors, driven by the appropriate combination of its value proposition and cost structure.

A firm's cost structure is defined by a variety of factors such as economies of scale, technology, automation, location of facilities, supply chain management, outsourcing and supplier relations, operational excellence and employee productivity.

Organizational Design

It is not enough to have a good strategy; the firm also has to implement this strategy well to be successful. It is necessary to design the organization appropriately to execute the strategy (Kates and Galbraith 2007). Top managers control four elements: organizational structure, managerial processes, corporate culture and management of people. Organizational structure determines who reports to whom, who has authority over what decisions, and how information and power flow in the organization. Various managerial processes bring this skeleton to life, such as performance measurement and rewards, budgeting and capital allocation, strategic planning and information systems. Corporate culture defines the values and norms the people in an organization share, and how they interact with each other and the environment. Managing people involves decisions on recruiting, training and development, promotions and succession planning.

Fit

Competitiveness is not determined by decisions on competitive domain, competitive position and execution in isolation. Rather, there is much interaction among all these choices, and what matters is how these decisions fit together – internal consistency. All these choices collectively have also to be congruent with the external environment (such as customers, technology, government regulation and social trends). Both internal and external consistency drive the firm's overall competitiveness.

See Also

- ▶ Competitive Advantage
- ▶ Competitive Strategy
- ▶ Cost Leadership
- ▶ Organizational Design

References

Borden, N.H. 1964. The concept of the marketing mix. In *Science in marketing*, ed. G. Schwartz. New York: Wiley.

Karnani, A. 2008. Controversy: The essence of strategy. *Business Strategy Review* 19: 28–34.

Kates, A., and J.R. Galbraith. 2007. *Designing your organization: Using the star model to solve five critical challenges*. San Francisco: Jossey-Bass.

Porter, M. 1979. How competitive forces shape strategy. *Harvard Business Review* 57: 137–145.

Competitors

John Horn
University of Michigan, Ann Arbor, MI, USA

Abstract

It is critical to understand competitors when you are developing a strategy. They serve not only as a benchmark against which you can set and measure goals, but they often stand in your way of achieving your strategic objectives. Understanding competitors is challenging because they may also be collaborators, suppliers, or customers, but, more importantly, because it is hard to predict what they will do. They pursue their own goals and have different assets, resources, and capabilities from those of your own organization. Competitors are rarely irrational, so systematic analysis can help create insight into their mindset and probable strategic choices.

Definition Competitors are organizations pursuing (at least) a portion of the same markets and opportunities that your own organization is targeting. They often actively try to prevent your organization from achieving its objectives.

Why Competitors Matter

At its core, strategy is about creating plans in order to achieve an objective. For any business or organization, there are always others who will try to help or hinder the achievement of that objective: customers, distributors, suppliers, regulators and public-interest stakeholders. But prime among these in the business environment are competitors – those who are trying to capture at least a portion of the market the organization is pursuing. Markets can grow, and markets evolve, but because we live in a world with limited resources, eventually strategy will run up against competitors that are trying to sell to the same customers, or purchase from a key supplier, or offer a similar product.

Competitors are important for strategy because they are a measure of your relative success. When an organization tracks earnings, it usually compares itself with the profits generated by its competitors. (And if the organization itself does not, you can be sure that investors do!) Gains and losses in market share, a common metric that companies use to measure strategic effectiveness, is only relevant when compared with competitors. (Apple does not gain market share at the expense of Pfizer.) Top-line revenue growth is also best understood relative to the particular competitors in a company's market. And marketing strategy is

predominantly about comparing your price, product portfolio and attributes, placement and promotional activities (both amount of spend and messaging) with those same characteristics of your competitors' products and services. Successful marketing strategy is about beating your ► competition on one or more of these elements in order to improve market share and/or profitability performance.

Competitors are not just a benchmark, however. They often actively try to prevent your organization from achieving those desired levels of success. For example, cut-throat pricing in a market can erode your profitability; entry by others reduces the market share available for your firm; and marketing efforts are typically undertaken not merely to grow the overall size of the market but to grow the competitor's share of the market, thereby slowing your growth. These competitor actions will almost certainly put stress on your organization. If the choices competitors made had no influence on the strategic objectives of other organizations they would not be considered to be competition.

Why Understanding Competitors Is Challenging

Understanding competitors is not a straightforward effort, however. First, there is not always a black-and-white distinction between who is and who is not a competitor. Another organization can be both a competitor and a collaborator. For example, in the US there are two main companies that build ships for the US Navy. These companies compete with each other to secure contracts from the Department of Defense. But they are also collaborators on a submarine programme. Each company builds approximately half of each vessel, the two halves of which are then combined at one of the two company's shipyards. British Petroleum and Shell compete with each other in many different markets of the oil industry, but they are also joint owners of a refinery in South Africa. Are BP and Shell competitors? Yes. But there are also times when they must cooperate.

Second, a competitor may also be a supplier or customer. The major car manufacturers compete with each other in virtually every geography and segment of the car market. But they also sell technologies to each other. Nissan licenses Toyota's hybrid drive system, and Toyota at one time bought batteries and an all electric power train from Tesla for use in Toyota vehicles. Is Toyota competing or cooperating with Nissan and Tesla? It depends.

Third, and most important, it is very hard to understand and predict what a competitor will do. After you've worked out that they are not a collaborator, customer or supplier, it is still very hard to know what your competitor is going to do. If you cannot predict what your competitor is going to do, or understand what moves it might make, it becomes very difficult to craft an effective strategy, since the competitor's moves may end up nullifying your expected gains.

Competitors are hard to understand because they are trying to do what is best for them, not what is best for you. If competitors were doing what was best for you they would not stay in business very long. First and foremost, competitors may have objectives and goals different from those of your organization. We usually assume all companies are trying to maximize long-term profits, but sometimes firms try to grow revenue, or gain market share, or prevent the further loss of market share, or boost short-term earnings. In the long term, we would expect these companies to go out of business if they do not eventually focus on generating profits, but in the shorter term, their strategic objective may look rather different from any theoretical prediction. (If you doubt this, ask yourself if *every* decision your company makes is with the goal of explicitly maximizing long-term profits.)

A second reason why it can be hard to understand competitors is that their resources and capabilities differ from those of your organization. Those different endowments mean they can make different choices to generate profits, and they might be able to make choices that are not feasible for your organization. If you cannot fathom choosing to do something because you do not have the necessary resources at hand it is

hard to imagine that anyone else would decide to go ahead with that choice either.

Companies often say that their competitors are irrational, but most of the time those competitors are not irrational – they are simply trying to achieve different objectives from a different starting point. It would be relatively simple if you could directly ask a competitor what their goals and plans were, and what their current resources, capabilities and investments were. But apart from being illegal in many jurisdictions, there is no reason to believe their answers. If the competitor wanted to create an advantage, an easy way to do it would be to tell you a set of false plans that led you to choose a path that was most beneficial to the competitor, not necessarily to you. In other words, if a competitor had the choice of telling you the truth (which would mean you could counter those moves and prevent the achievement of the goal), or telling you a plan that you would counter in such a way that made the competitor's *real* plan easier to execute, which one do you think the competitor would choose?

Gaining Competitive Insight

If competitors are not irrational, and you cannot ask them directly for their plans, how can you gain better insight into what their objectives and strategies might be? The best way is to engage in an ongoing process of observing and analysing the market from their perspective. As described in a *McKinsey Quarterly* article (Courtney et al. 2009), a four-step, continual process would be:

1. **Stop, Look and Listen to the competition**. What are they saying in press releases? What are the actual choices they have been making? What are the latest social media campaigns they have launched? Where have they announced new acquisitions? Or new markets they are entering? Every company has to communicate with shareholders (if publicly held) and other stakeholders (including privately held and other non-public organizations) to some degree. Pay attention to what they are saying to gain clues about their objectives.
2. **Think like a strategist working for the competition**. If you had the competitor's assets, resources, capabilities and competencies what would you do? What market and customer segments would you focus on? What product white spaces would you pursue? Where would you focus on improving costs? Where would you shift your research and development (R&D) spend? Your competitors are probably wrestling with these same questions, using their existing resources and capabilities. Try to craft what you would do in their stead.
3. **Determine who will be making the decision**. When the time comes to pull the trigger, there will be a person, or small group of people, at the competitor organization who will make the decision. Is it the CEO or President of the business unit or the front-line sales staff? Each has a different set of incentives, and different people also have different proclivities and ways of making decisions. A CEO who rose through the ranks as a CFO is likely to make different decisions from someone who started off as a marketer. Neither is right or wrong, but the decision maker's background and incentives will influence their decision-making style.
4. **Make a prediction and follow up**. The first three steps guide you towards predicting what your competitor will do. Will they lower prices? By how much? In which geographies? To which customer segments? Use what they've said, what they have, and the insights about the decision maker to make a prediction. Do not stop there, though. Observe what strategic choice they actually make, and compare it with your prediction. Learn from analysing where you were right, and, as importantly, where you were wrong. In cases where you did not get things exactly right repeat steps 1 through 3 to refine your analysis, make a new prediction and then assess whether or not you have improved.

Competitors are not trying to help you succeed, and in fact are often aiming for the opposite outcome. You should never expect to fully understand your competitors. But if you get better at gaining insight into how their minds work, and what that implies for their strategic

choices, they will not seem as irrational. That demystification will go a long way towards being better prepared to confront them in the marketplace.

See Also

- ▶ Business Strategy
- ▶ Competition
- ▶ Competitive Advantage
- ▶ Competitive Strategy
- ▶ Five Forces Framework
- ▶ Game Theory

References

Courtney, H., J. Horn, and J. Kar. 2009. Getting into your competitor's head. *McKinsey Quarterly.* 128–137.

Complementarities

David J. Teece
Berkeley Research Group, LLC, Emeryville, CA, USA
Haas School of Business, University of California, Berkeley, Berkeley, USA

Abstract

Complementarities – when two or more items are more valuable in combination – can be vertical or horizontal, and they can involve user utility, prices, and/or technologies. They are important for innovation because most technologies require complementary technologies if their full value is to be realized. Complementarities are important for strategy in a number of realms, including the setting of prices, investments, and the boundaries of the firm.

Definition Complementarities occur when two or more items are more valuable in combination than in isolation. They can be vertical or horizontal, and they can involve user utility, prices, and/or technologies.

Classification

Economics and strategy; Foundations; Innovation and strategy

Complementarities generate important sources of value throughout the economic system. The basic notion is that the whole (i.e., two or more complements in combination) is greater than the sum of its parts. For example, the laser and the computer, when combined, enabled, among other things, CD and DVD players. There are, however, many subtle variations on the complementarity theme.

The most commonly studied types of complementarity involve factor prices or value from use (Carlaw and Lipsey 2002). Innovation studies (e.g., Rosenberg and Frischtak 1983) look instead at the impact of new combinations of existing technologies.

There are in fact at least six types of complementarity of economic and strategic importance, presented here in the historical order in which they were identified. While there is inevitably overlap among the concepts, each one has its uses.

Product Pricing (Associated with Cournot 1838)

Cournot complements exist when, for example, two firms hold market power (duopoly) over separate inputs to a common downstream good, they can maximize their joint profits by colluding (where legal) or by tacitly coordinating in order to avoid raising factor costs so high that the downstream product loses too much of its sales.

Consumption Utility (Associated with Edgeworth 1897)

Two goods, X and Y, are Edgeworth complements if a consumer derives greater utility from consuming them together than from consuming them in isolation. The quantity demanded of either good is

affected by a change in the quantity demanded of the other.

Factor Prices (Associated with Hicks 1970)

Factors of production are Hicksian complements when a decrease in the price of one leads to an increase in the quantity used of its complements.

Asset Prices (Associated with Hirshleifer 1971)

If an innovation is likely to move factor demand as described by Hicksian complementarity or by undermining the demand for a substitute factor, then the innovator can potentially use this foreknowledge to profit from the expected future changes in asset prices (e.g., going short in markets that would be negatively impacted).

Technologies (Associated with Teece 1986, 2006)

The realized value of an innovation typically depends on altering the nature of one or more existing (complementary) assets or technologies and/or on creating new ones. For example, reaping the full value from the introduction of electricity required the creation of electric motors that could be attached to machines.

Innovations (Associated with Bresnahan and Trajtenberg 1995)

Improvements in a general-purpose technology will increase the innovation possibilities in downstream sectors. For instance, the improvement of a cellular network opens new technological opportunities for firms providing wireless data devices. This is closely related to technological complementarities, but an improvement in any single downstream sector is less critical to value capture by the focal (general-purpose) technology.

Edgeworth complementarity is primarily important for conceptualizing the fundamentals of why optimal pricing of monopolistically supplied inputs requires coordination. For strategy purposes, Cournot and Hicks complementarities can be analyzed using game theory, although reality often falls short of what the theoretical models predict. Hirshleifer complementarity can also be risky to rely on in practice because it abstracts from the market uncertainty affecting most innovations; however, it provides a useful framework for understanding the asset pricing consequences of innovation in one product area on another and how this can be harnessed to help innovators capture value.

Technological complementarity is pervasive and significant. Absent complementary technologies, many products simply will not get developed and commercialized. This was the case, for example, in the US electrical supply industry at the end of the nineteenth century. The industry had a killer app – lighting – but was mired in a "war of the currents" between alternating and direct current, each of which had certain deficiencies. It was only with the development of rotary converters that one system (alternating current) was able to develop a dominant position and spur rapid deployment (David 1992).

Innovation complementarity is important mainly in that it shines a light on the spillover benefits of general-purpose technologies. This is an important reason that owners of general-purpose technologies are unable to earn more than a fraction of the total social benefits that they help create.

A model for developing strategy around complements with regard to the appropriability potential of particular innovations is Profiting From Innovation (Teece 1986, 2006, forthcoming). It places complementarity in a broader context, such as whether complements are cospecialized from an investment standpoint, and is useful for analyzing the key make-or-buy issues for choosing the degree of vertical integration and other aspects of organizational design.

See Also

- ▶ Appropriability
- ▶ Collusion and Cartels
- ▶ Complementary Asset
- ▶ Duopoly
- ▶ Game Theory
- ▶ General-Purpose Technology
- ▶ Innovation
- ▶ Make-or-Buy Decisions: Applications to Strategy Research
- ▶ Organizational Design
- ▶ Outsourcing
- ▶ Profiting from Innovation
- ▶ Rivalry and Collusion
- ▶ Vertical Integration

References

Bresnahan, T.F., and M. Trajtenberg. 1995. General purpose technologies: 'Engines of growth'? *Journal of Econometrics* 65 (1): 83–108.

Carlaw, K.I., and R.G. Lipsey. 2002. Externalities, technological complementarities and sustained economic growth. *Research Policy* 31 (8): 1305–1315.

Cournot, A.A. 1838. *Recherches sur les Principes Mathematiques de la Theorie des Richesses*. Translated into English by N. Bacon as *Researches in the Mathematical Principles of the Theory of Wealth*. London: Haffner, 1960.

David, P.A. 1992. Heroes, herds and hysteresis in technological history: Thomas Edison and 'The Battle of the Systems' reconsidered. *Industrial and Corporate Change* 1 (1): 129–180.

Edgeworth, F.Y. 1897. Teoria Pura del Monopolio. *Giornale degli Economisti* 15: 13–31. Translated as The pure theory of monopoly in *Papers Relating to Political Economy*, vol. I. London: Macmillan and Co., 111–142, 1925.

Hicks, J. 1970. Elasticity of substitution again: Substitutes and complements. *Oxford Economic Papers* 22 (3): 289–296.

Hirshleifer, J. 1971. The private and social value of information and the reward to inventive activity. *American Economic Review* 61 (4): 561–574.

Rosenberg, N., and C.R. Frischtak. 1983. Long waves and economic growth: A critical appraisal. *American Economic Review* 73 (2): 146–151.

Teece, D.J. 1986. Profiting from technological innovation. *Research Policy* 15 (6): 285–305.

Teece, D.J. 2006. Reflections on profiting from innovation. *Research Policy* 35 (8): 1131–1146.

Teece, D.J. forthcoming. Profiting from innovation in the digital economy: Standards, complementary assets, and business models in the wireless world. *Research Policy.*

Complementary Asset

Marco Ceccagnoli
Georgia Tech, College of Management, Atlanta, GA, USA

C

Abstract

Being first to market or possessing strong intellectual property protection may not be enough to profit from a groundbreaking invention. One must also consider the upstream and downstream complementary assets necessary to successfully commercialize an invention. Particularly important are the specialized and co-specialized assets that are idiosyncratic to the innovation. Since these cannot be readily bought and sold in the market, their ownership provides a source of sustainable competitive advantage. Ultimately, it is a combination of inventions that are hard to imitate and ownership of complementary assets that are hard to acquire that affect the degree to which companies capture rents from their innovations.

Definition Complementary assets are the upstream and downstream assets necessary to successfully commercialize an invention. These can be generic assets that can be used for multiple purposes, or specialized and co-specialized assets that are idiosyncratic to the innovation and don't easily trade on markets. Identification and control of complementary assets that are hard to acquire are crucial to capture rents from innovations.

It is not enough to develop a groundbreaking invention and be first to market to capture the ▶ profiting from innovation (PFI). One must also consider the upstream and downstream complementary assets necessary to successfully commercialize an invention (Teece 1986).

Many of the world's most respected companies are those that have created innovations that have changed the world. There are many companies, however, who also introduced innovations that have had similar impacts on our lives yet were not as successful. Failure often lies in the inability

of the innovator to gain access to complementary assets that are hard to acquire.

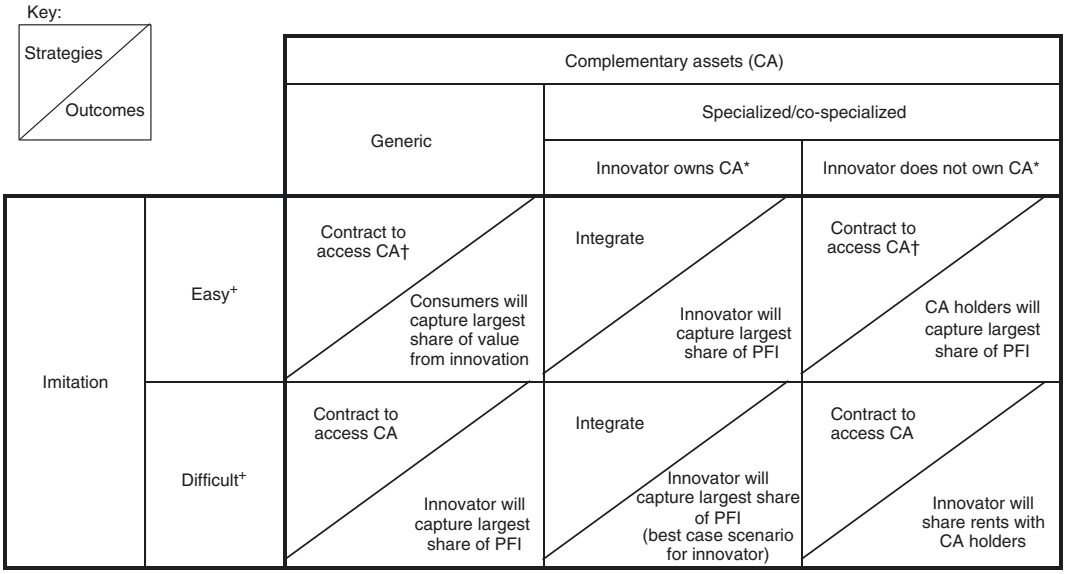
If an invention is one part of a manufactured good, the remaining components are considered complementary assets, as are the capability to manufacture the good, a strong brand associated with that product, and a distribution network for selling and servicing the finished good. Complementary assets can be divided into three groups: generic, specialized and co-specialized, depending on the level of interdependence between the asset and the invention (Teece 1986).

1. Generic assets are essentially commodities that can be used for multiple purposes. Manufacturing facilities used to produce many consumer products, such as shoes or toys, are typically generic.
2. Specialized assets must be tailored to the innovation. The sales force of pharmaceutical companies is a classic example, since it is often extensively trained in a specific class of drugs. When new products developed in those classes critically depend on the specialized sales force, but the latter can be redeployed at low cost in alternative uses, there is unilateral dependence between the innovations and the complementary assets.
3. Co-specialized assets are more valuable when they are used in conjunction with the innovation. With reference to the previous example, if the specialized sales force cannot be redeployed from one therapeutic category to another, there would be complete mutual dependence – or co-specialization – between the innovation and the complementary asset.

Since specialized and co-specialized complementary assets cannot be readily bought and sold in the market, their ownership provides a source of sustainable competitive advantage. Examples abound in high-tech industries. For instance, Apple Inc., which has come to dominate the tablet computer market, did not succeed by being the first to market; in fact, the touch-screen tablet was first commercialized in the 1990s. Apple's strategy for the iPad was to imitate and improve the original tablet technology. Apple achieved a competitive advantage in this market by controlling key specialized assets: a strong brand; several complementary technologies which successfully transferred from its digital music player, the iPod; in-house digital rights management software; tacit technical capabilities that deliver a product with proverbial design; and an interface that is easy to use. Apple also controls key co-specialized assets such as the Apps and iTunes Stores: huge marketplaces owned by Apple that enhance the user experience through the online purchase of functional applications and music. Their use in conjunction with the innovation is value-enhancing. On the one hand, Apple benefits from the virtual stores since they encourage its consumers to remain loyal and enhance its bargaining position; on the other hand, the iPad benefits the virtual stores, since it provides developers and artists with a large installed base of Apple customers. Finally, Apple outsources production and assembly associated with the iPad, since these are generic complementary assets that are available in competitive markets.

It is important to stress that it is the combination of inventions that are hard to imitate and ownership of complementary assets that are hard to acquire that affect the degree to which companies capture the value from innovation (Teece 1986; Arora and Ceccagnoli 2006). Inventions that are hard to imitate are said to benefit from a strong ▶ appropriability regime (Teece 1986). The latter is mainly determined by the ease of replication and how well the inventor can guard his invention from actual or potential competitors through legal means. A summary of typical cases is represented in Fig. 1.

A strong appropriability regime combined with generic complementary assets or specialized/cospecialized assets controlled by the innovator leads to inventors capturing most of the value from innovation. A common situation, however, involves start-ups lacking downstream complementary assets: when these are specialized and the invention is easy to imitate, these companies tend to lose the competitive battle in the marketplace, as owners of complementary assets imitate and exploit their inventions. In this case, vertical integration in the product market is deemed to be a



*: Ownership (or lack of it) may reflect cases whereby the innovator is in a better (or worse) position relative to potential rivals for the acquisition of CA.
†: In this case it may be hard for the innovator to recover the investment required to create the innovation.
+: Ease of imitation is mainly determined by the efficacy of intellectual property rights protection and the replicability of the innovation.

Complementary Asset, Fig. 1 Profits from innovation (*PFI*): Teece's model (Adapted from Teece (1986))

failure. A combination of a strong appropriability regime and lack of specialized complementary assets leads small companies to contract for access to the complementary assets. Contracting can be accomplished through a variety of cooperative strategies that allow the innovator to share the PFI with the complementary assets holders, such as licensing, joint ventures or the sale of the company (Rothaermel 2001; Gans et al. 2002; Arora and Ceccagnoli 2006; Ceccagnoli and Hicks 2013). In such cases, profits captured by the innovator are proportional to the strength of the appropriability regime; profits captured by owners of complementary assets are proportional to the degree to which these are specialized (Ceccagnoli and Rothaermel 2008).

With economic profits in developed nations shifting to knowledge assets, complementary assets are the crucial link in transferring the knowledge asset into successful products or services. Liberalization of markets means that competitive advantage is not likely to occur from access to a market and must come from ownership of assets which do not trade on markets (Teece 1998). Even as global intellectual property rights protection has strengthened in recent years, it is still relatively difficult to protect one's invention via the courts. Identifying and controlling key complementary assets may be crucial to capture the lion's share of the PFI and achieve an innovation-based sustainable competitive advantage.

See Also

- ▶ Appropriability
- ▶ Cooperation and Competition
- ▶ First-Mover Advantage
- ▶ Markets for Technology
- ▶ Profiting from Innovation
- ▶ Technology Strategy

References

Arora, A., and M. Ceccagnoli. 2006. Patent protection, complementary assets, and firms' incentives for technology licensing. *Management Science* 52: 293–308.

Ceccagnoli, M., and D. Hicks. 2013. Complementary assets and the choice of organizational governance: Empirical evidence from a large sample of US technology-based firms. *IEEE Transactions in Engineering Management* 60: 99–112.
Ceccagnoli, M., and F.T. Rothaermel. 2008. Appropriating the returns from innovation. In *Technological innovation: Generating economic results*, vol. 18, ed. G.D. Libecap and M.C. Thursby. London: JAI/Elsevier.
Gans, J.S., D.H. Hsu, and S. Stern. 2002. When does start-up innovation spur the gale of creative destruction. *RAND Journal of Economics* 33: 571–586.
Rothaermel, F.T. 2001. Incumbent's advantage through exploiting complementary assets via interfirm cooperation. *Strategic Management Journal* 22: 687–699.
Teece, D.J. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy* 15: 285–305.
Teece, D.J. 1998. Capturing value from knowledge assets: The new economy, markets for know-how, and intangible assets. *California Management Review* 40: 55–79.

Computational Simulation

Michael J. Prietula and Abhishek Kathuria
Emory University, Goizueta Business School, Atlanta, GA, USA

Abstract

A significant and growing set of approaches in strategic management research are centred on the use of computational models realized as simulations. We provide a characterization of what constitutes a computational simulation and enumerate the possible roles computational simulations can play in strategic management research. By exploring the broad fundamentals and issues underlying the use and contribution of computational modelling, we hope to help facilitate the use of simulation in providing insight into key issues of strategic management. We provide a brief examination of the history, benefits, uses and forms of computational simulations, and explicate the concerns and issues that lie at the core of any simulation development effort.

Definition A computational simulation is a dynamic, process-oriented model instantiated on a computer. These can range from traditional economic models (expressed as equations) to more abstract constructs and processes (expressed as objects, agents, operators and algorithms).

A significant and growing suite of tools and approaches in strategic management research are centred on the use of computational models realized as simulations. However, as the study of strategic management itself is a moving target, it is important to understand the broad fundamentals and issues underlying the computational simulation approaches in order to intelligently discern how they can contribute to strategic management research.

Brief Historical Overview

The computer has been an adjunct for research in strategic management for decades – consider the early work of Jay Forrester, incorporating systems dynamics simulations as mechanisms to improve a firm's strategic decisions (Forrester 1958). In fact, by 1965 'simulation had become a widely used methodology' in the social sciences (Dutton and Starbuck 1971: 3), including such publications as *Industrial Dynamics* (Forrester 1961), *Simulation in Social Science* (Guetzkow 1962) and *A Behavioral Theory of the Firm* (Cyert and March 1963). The garbage can model (Cohen et al. 1972), models of adaptive search (Levinthal and March 1981), models of cooperation (Axelrod 1984) and the exploration-exploitation model (March 1991) all followed. Movements towards the consideration of the microphenomenon underlying macro-behaviour (e.g., Schelling 1978) afforded legitimacy to reductionism in examining strategic constructs, but the use of simulation methods in mainstream management journals remained infrequent (Berends and Romme 1999). Over time, interdisciplinary organizations emerged to accommodate intellectual exchanges (e.g., the World Congress on Social Simulation), the National Science Foundation provided support for research and summer

schools educating Ph.D. students and faculty, and key conferences such as that held by the Academy of Management offered pre-meeting professional development workshops and support for computational simulations topics on management and strategy. Thus, the past several years have witnessed a distinct increase in the use and sophistication of computational simulations in management research. In other disciplines, computational simulation has become a 'third branch' of science (Pool 1992), with that trend growing in studies of organization science (Carley 2002) and economics (LeBaron and Tesfatsion 2008; Tesfatsion and Judd 2006). This trend is now gaining strength in the study of strategic management, which often bridges both disciplines. Furthermore, the exact nature of what type of simulation to engage depends on the particular investigative or theoretical context – for example, is it 'the strategy' (economic, macro-organizational) or the 'the strategist' (cognitive science, micro-organizational)? Accordingly, it is important to understand what constitutes a computational simulation and the possible roles computational simulations can play in strategic management research.

What Is a Computational Simulation?

For our purposes a computational simulation is simply a dynamic, process-oriented model instantiated on a computer. These can range from traditional economic models expressed as equations (e.g., Berry and Pakes 1993; Werden et al. 1991) to models representing individual agents to societies (e.g., North and Macal 2007; Zacharias et al. 2008). In actuality, this is quite a powerful statement given the broad nature of the *types* and *extent* of models that can be realized on computer, and the fundamental robustness of computational formalisms as suggested by the Church-Turing thesis (Harel 1987). Consequently, as one moves away from a more standard form (e.g., a system of linear equations, an LP model) to more flexible ones (e.g., using an agent-based package, programming your own model), it is essential that the assumptions, both explicit and implicit, are known and articulated. As these assumptions are embedded in any model that is created, they must be understood in order to appreciate their contribution to the analysis of model behaviour. For example, Davis et al. (2007) provide a comparison of five approaches to simulation, focusing on several dimensions: focus, common research question, key assumptions, theoretical logic and common experiments. Following that lead, we focus on computational simulations and examine the benefits and uses of simulation for strategic management research.

Some Benefits of Computational Simulation

As computational simulations are dynamic instantiations of models, incorporating incremental and causal changes in parameters over a precedent-temporal (i.e., event-driven or continuous) interval, they embody specification of processes. The value of this resides in the ability to specify the model at multiple and varying levels of complexity and formality, and to include the complex dynamics inherent in business, market or sector activity unfolding over time (Prietula 2011). Consider four general benefits of computational simulations for strategic management research.

First, computational simulations focus on organizational phenomena that go together in a theoretically meaningful way. A set of constructs are defined that operate dynamically, evidencing individual and collective behaviour over time, possibly addressing multiple levels of abstraction. For example, Lant and Mezias (1990) define and manipulate levels of entrepreneurial activity, as well as types of entrepreneurial strategy, under a learning model in order to examine the consequences (as performance, resources and bankruptcies) over extended time periods. Aggarwal et al. (2011) developed an agent-based model to examine the performance impacts of governance modes used to make decisions in inter-organizational alliances. Their model showcased the interactions between different governance structures, patterns of interdependencies and levels of organizational search capabilities.

Second, computational simulations describe both the process and the product of behaviour. That is, they allow a reductionist argument and permit traces of behaviour over time that can be associated with theoretical constructs in the model. There is a specific distinction between behaviour and performance. The oligopoly model described in Cyert and March (1963) included many process components. Prietula and Watson (2000, 2008) replicated that model and examined how the model component processes contributed to outcome product, but also demonstrated how its routines account for economic behaviour. By incorporating changes in risk preference at extremes of performance and alternative reference group strategies, Hu et al. (2011) extend the prior process descriptive simulations of adaptive aspirations by March (1988) and March and Shapira (1987, 1992). Miller et al. (2009) used a simulation model to study participation by incorporating a new risk preference function and examining alternative strategies for setting reference groups of organizations in online communities as a means to use social learning processes to shape demand for products. The authors model product demand as a function of interpersonal communication and firm strategy, and find key contingencies that can inform firm strategies in this context. Markle (2011) uses experimental data and setting from a gift exchange game published in a prior study as a basis for a computer simulation to examine the judgement of firms regarding employee reciprocity. The simulation supports the dysfunctional learning process suggested in prior theory by demonstrating systematic bias towards an overemphasis on employee self-interest, and subsequent wage choice inefficiencies.

Third, computational simulations enforce both the formalism and uniformity describing organizational phenomena. Unlike verbal or 'pictographic' articulations of theory, computational simulations are more demanding of exact specifications of objects and processes of the model. If it is asserted in the model, it is represented in the code (at some level of abstraction), and the implications of the code are unambiguous in execution. But not all the components of the simulation embody components of the model, as ancillary elements may be needed to hold together and enable components of the model so that it can be realized in the simulation. One solution to this is to provide the code itself (Cohen et al. 1972; Cyert and March 1963; Levinthal and March 1981). Nevertheless, as models become more complex the visibility of the constructs and their dynamics becomes somewhat obscured, thus obfuscating the model itself. Therefore, model complexity and ancillary elements (which may invade the definitions and behaviours of the model constructs) can lead to model obfuscation. Two basic methods to mitigate obfuscation risk are model assembly and code assembly.

Model assembly refers to borrowing components that are well defined in another model and manipulating them to fit the context of interest. For example, Gavetti et al. (2005), Siggelkow and Levinthal (2005), Levinthal and Posen (2007), Ethiraj et al. (2008), and Aggarwal et al. (2011) all construct strategic simulations based on modifying the NK model developed in the context of biology and borrowed from physics (Kauffman 1993), while Carley and Svoboda (1996) incorporate an optimization procedure (simulated annealing) to model organizational adaptation. Systems dynamics has been well represented in model assembly approaches. Repenning (2002) examined the dynamics of implementation using systems dynamics, Black et al. (2004) developed a model grounded in ethnographic data from a prior study, and in his classic textbook Sterman (2000) has a rich set of examples. Other popular model assemblies one may encounter include genetic algorithms (e.g., Goldberg 1989) for evolutionary contexts (e.g., Lee et al. 2002) and cellular automata (Wolfram 2002). *Code assembly* refers to the use of an existing computational model or modelling environment (see subsequent section on "Forms of Computational Simulations"). Code assembly includes not only elements of a model (static and dynamic components) but the underlying mechanisms of running it directly. As noted, it is important to be aware of the assumptions and underlying mechanisms of model and code assembly, and how those relate to the model under construction.

Finally, another benefit of computational simulations is that postulated constructs can be manipulated explicitly. In the real world it may be difficult to find cases that give sufficient coverage of a parameter space to test hypotheses or theories. Accordingly, computational simulation can instantiate a theoretical model and examine its behaviour under varying conditions. For example, Gary (2005) manipulated seven diversification strategies in a computational simulation, which led to insights and refinements to extant theoretical descriptions.

How Computational Simulations Are Used

The previous section discussed some of the key advantages of computational simulations, but it is also important to consider what role these simulations may play from a researcher's perspective. Consider a sampling of purposes for computational simulations in social science relevant to strategic management research (Burton and Obel 1980; Carley and Prietula 1994; Axtell et al. 1996; Axelrod 1997; Carley 1999; Davis et al. 2007; Harrison et al. 2007):

- Prediction (e.g., of consequences of alternative policy decisions)
- Proof (e.g., existence or sufficiency to demonstrate or account for phenomena)
- Discovery (e.g., of new effects of hypothesized mechanisms)
- Replication (e.g., of other theoretical or computational models to test reproducibility of results)
- Explanation (e.g., of what processes underlie the presence of a phenomena)
- Critique (e.g., to seek more parsimonious explanations for hypothesized phenomena)
- Prescription (e.g., for generating evidence to select a better policy, organizational design or strategy)
- Empirical guidance (e.g., suggesting further research areas to pursue)
- Theory development (e.g., refine and modify theoretical constructs or the conditions under which they apply)
- Hypothesis generation (e.g., run a series of simulations from a theoretical model to produce derivative hypotheses to be tested)
- Hypothesis testing (e.g., run a series of simulations to test hypotheses proposed by a theory)
- Instantiation (e.g., move from a verbal theory to a computational form)
- Docking or alignment (e.g., determining how/if two or more models that embody distinctively different mechanisms are equivalent or one can be subsumed within the other in explaining or predicting similar phenomena).

Forms of Computational Simulations

It is important to understand that the model, per se, is not the focus; rather, it is what the model *represents* – that is, the theoretical justification and specification of the constructs and form selected for the simulation. The substance of the theoretical rhetoric and reasoning cannot be lost in the implementation. In any such model, one must be able to easily and unambiguously identify not only the theoretically relevant components, but how these components contribute to process behaviours.

Several general software options are available in helping to construct a computational simulation (recall previous discussion on *code assembly*), and each has its particular benefits and risks. These can be classified into three general (not necessarily independent) categories: free code, mathematical software and pre-structured environments. *Freecode* involves building a model from a general purpose programming language such as Basic, C++ or Java (including some statistical packages that allow general programming). The notion is that this approach is free of any bias towards particular organizational assumptions or design constraints. *Mathematical* software packages, such as Matlab, Maple, Magma, Mathematica and Sage, afford more constrained approaches that exploit underlying representations of mathematical objects and process. Finally, *pre-structured environments* are

those that impose some type of design constraint on the specifications of the model, and reflect a broad range of environments. At lower levels of detail, Java packages and C++ libraries are available as pre-written components of simulations that may be reused. This makes the components transparent, but sufficient understanding of those languages is required in order for them to be exploited to the full. At higher levels of detail, programming environments dedicated to general types of computational and representational perspectives are available, including various systems dynamics environments or versions of agent-based frameworks as in MASON, Netlogo, Swarm, Ascape or RePast. Researchers also offer simulations that are more specific as to their underlying assumptions and approaches to representing organizations (Carley 2002; Ren et al. 2006), types of markets (Somani and Tesfatsion 2008) or societies (Epstein 2006).

Concerns and Issues

In general, computational simulations require both *validation* (Did you build the *correct* model?) and *verification* (Did you build the model *correctly*?), which are issues at the core of any simulation development effort (e.g., Conway 1963; Kleindorfer et al. 1998; Naylor and Finger 1967; Van Horn 1971). Most concerns regarding computational simulation can be stated in these terms. In strategic management simulations the issue of validation has tended to receive much more attention. Verification is problematic in any software endeavour but in computational simulations it is mostly absent, except through exogenous review by releasing code, algorithms, or appealing to trusted sources (e.g., pre-structured environments). On the other hand, validation is where discussions specific to computational simulations in management have emerged (e.g., Burton 2003; Burton and Obel 1995; Miller 1998; Thomsen et al. 1999). Transparency is lost as the complexity of the model increases. Once constructed, there is a fundamental difference between the following two questions (which must always be asked): What in the *model* is accounting for these phenomena? What in the *simulation* is accounting for these phenomena? For example, there are many alternative ways to *implement* NK, genetic algorithm and agent-based models that can result in subtle, but substantial, variations in behaviour.

Another issue centres on the plasticity of the computational environment, where models may simply be 'mere' *Gedenken* (thought) experiments. In fact, both science and philosophy have a rich history of thought experiments and their contributions (Horowitz and Massey 1991; Sorensen 1992). In fact, once a simulation is realized it ceases to be a thought experiment and becomes instead a substantive artefact of – and for – research. The thoughts are in the code. In such a context it may be more appropriate to consider the philosophical concerns of computational social science in general (e.g., Henrickson and McKelvey 2002).

Perhaps a good place to begin in understanding the role of computational simulations and what they can contribute to strategic management research (and researchers) is John Sterman's Jay Wright Forrester Prize Lecture (2002), in which he notes that bounded rationality limits our ability to understand even the simplest dynamical systems. Through simulations, the complexity of our own theories may be explicated by creating and witnessing and sharing the behaviours of our creations.

See Also

- ▶ Behavioural Strategy
- ▶ Local Search
- ▶ Simulation Modelling and Business Strategy Research

References

Axelrod, R. 1984. *The evolution of cooperation*. New York: Basic Books.

Axelrod, R. 1997. Advancing the art of simulation in the social sciences. In *Simulating social phenomena*, ed. R. Conte, R. Hegselmann, and P. Terna. Berlin: Springer.

Axtell, R.L., R. Axelrod, J.M. Epstein, and M.D. Cohen. 1996. Aligning simulation models: A case study and results. *Computational and Mathematical Organization Theory* 1: 123–141.

Berends, P.A.J., and A.G.L. Romme. 1999. Simulation as a research tool in management studies. *European Management Journal* 17: 576–583.

Berry, S., and A. Pakes. 1993. Some applications and limitations of recent advances in empirical industrial organization: Merger analysis. *American Economic Review* 83: 247–252.

Black, L.J., P.R. Carlile, and N.P. Repenning. 2004. A dynamic theory of expertise and occupational boundaries in new technology implementation: Building on Barley's study of CT scanning. *Administrative Science Quarterly* 49: 572–607.

Burton, R.M. 2003. Computational laboratories for organization science: Questions, validity and docking. *Computational and Mathematical Organization Theory* 9: 91–108.

Burton, R.M., and B. Obel. 1980. A computer simulation test of the M-Form hypothesis. *Administrative Science Quarterly* 25: 457–466.

Burton, R.M., and B. Obel. 1995. The validity of computational models in organization science: From model realism to purpose of the model. *Computational and Mathematical Organization Theory* 1: 57–71.

Carley, K.M. 1999. On generating hypotheses using computer simulations. *Systems Engineering* 2: 69–77.

Carley, K.M. 2002. Computational organization science: A new frontier. *Proceedings of the National Academy of Sciences* 99: 7257–7262.

Carley, K.M., and K.M. Prietula. 1994. Computational organization theory: An introduction. In *Computational organization theory*, ed. K.M. Carley and M.J. Prietula. Hillsdale: Erlbaum.

Carley, K.M., and D.M. Svoboda. 1996. Modeling organizational adaptation as a simulated annealing process. *Sociological Methods & Research* 25: 138–168.

Cohen, M.D., J.G. March, and J.P. Olsen. 1972. A garbage can model of organizational choice. *Administrative Science Review* 17: 1–25.

Conway, R.W. 1963. Some tactical problems in digital simulations. *Management Science* 10: 47–61.

Cyert, R.M., and J.G. March. 1963. *A behavioral theory of the firm*. Englewood Cliffs: Prentice-Hall.

Davis, J.P., K.M. Eisenhardt, and C.B. Bingham. 2007. Developing theory through simulation methods. *Academy of Management Review* 32: 480–499.

Dutton, J.W., and W.H. Starbuck. 1971. *Computer simulation of human behavior*. New York: Wiley.

Epstein, J.M. (ed.). 2006. *Generative social science*. Princeton: Princeton University Press.

Ethiraj, S.K., D.A. Levinthal, and R.R. Roy. 2008. The dual role of modularity: Innovation and imitation. *Management Science* 54: 939–955.

Forrester, J.W. 1958. Industrial dynamics: A major breakthrough for decision makers. *Harvard Business Review* 36: 37–66.

Forrester, J.W. 1961. *Industrial dynamics*. Cambridge, MA: The MIT Press.

Gary, M.S. 2005. Implementation strategy and performance outcomes in related diversification. *Strategic Management Journal* 26: 643–664.

Gavetti, G., D.A. Levinthal, and J.W. Rivkin. 2005. Strategy making in novel and complex worlds: The power of analogy. *Strategic Management Journal* 26: 691–712.

Goldberg, D.E. 1989. *Genetic algorithms in search, optimization & machine learning*. Boston: Addison-Wesley.

Guetzkow, H. (ed.). 1962. *Simulation in social science: Readings*. Englewood Cliffs: Prentice Hall.

Harel, D. 1987. *Algorithmics: The spirit of computing*. Reading: Addison-Wesley.

Harrison, J.R., Z. Lin, G.R. Carroll, and K.M. Carley. 2007. Simulation modeling in organizational and management research. *Academy of Management Review* 32: 1229–1245.

Henrickson, L., and B. McKelvey. 2002. Foundations of a 'new' social science: Institutional legitimacy from philosophy, complexity science, postmodernism, and agent-based modeling. *Proceedings of the National Academy of Sciences* 99: 7288–7295.

Horowitz, T., and G.J. Massey (eds.). 1991. *Thought experiments in science and philosophy*. Savage: Rowman & Littlefield.

Hu, S., D. Blettner, and R.A. Bettis. 2011. Adaptive aspirations: Performance consequences of risk preferences at extremes and alternative reference groups. *Strategic Management Journal* 32: 1426–1436.

Kauffman, S.A. 1993. *The origins of order: Self-organization and selection in evolution*. Oxford: Oxford University Press.

Kleindorfer, G.B., L. O'Neill, and R. Ganeshan. 1998. Validation in simulation: Various positions in the philosophy of science. *Management Science* 44: 1087–1099.

Lant, T.K., and S.J. Mezias. 1990. Managing discontinuous change: A simulation study of organizational learning and entrepreneurship. *Strategic Management Journal* 11: 147–179.

LeBaron, B., and L. Tesfatsion. 2008. Modeling macroeconomies as open-ended dynamic systems of interacting agents. *American Economic Review* 98: 246–250.

Lee, J., K. Lee, and S. Rho. 2002. An evolutionary perspective on strategic group emergence: A genetic algorithm-based model. *Strategic Management Journal* 23: 727–746.

Levinthal, D.A., and J.G. March. 1981. A model of adaptive organizational search. *Journal of Economic Behavior & Organization* 2: 307–333.

Levinthal, D.A., and H.E. Posen. 2007. Myopia of selection: Does organizational adaptation limit the efficacy of population selection? *Administrative Science Quarterly* 52: 586–620.

March, J.G. 1988. Variables risk preferences and adaptive aspirations. *Journal of Economic Behavior & Organization* 9: 5–24.

March, J.G. 1991. Exploration and exploitation in organizational learning. *Organization Science* 2: 71–87.
March, J.G., and Z. Shapira. 1987. Managerial perspectives on risk and risk taking. *Management Science* 33: 1404–1418.
March, J.G., and Z. Shapira. 1992. Variable risk preferences and the focus of attention. *Psychological Review* 99: 172–183.
Markle, A.B.. 2011. Dysfunctional learning in decision processes: The case of employee reciprocity. *Strategic Management Journal* 32: 1411–1425.
Miller, J. 1998. Active nonlinear tests (ANT) of complex simulation models. *Management Science* 44: 820–830.
Miller, K.D., F. Fabian, and S.J. Lin. 2009. Strategies for online communities. *Strategic Management Journal* 30: 305–322.
Naylor, T.H., and J.M. Finger. 1967. Verification of computer simulation models. *Management Science* 14: 92–101.
North, M.J., and C.M. Macal. 2007. *Managing business complexity: Discovering strategic solutions with agent-based modeling and simulation*. Oxford: Oxford University Press.
Pool, R. 1992. The third branch of science debuts. *Science* 256: 44–47.
Prietula, M. 2011. Thoughts on complexity and computational models. In *The SAGE handbook of complexity and management*, ed. P. Allen, S. Maguire, and B. McKelvey. Thousand Oaks: Sage.
Prietula, M.J., and H.S. Watson. 2000. A behavioral theory of the firm: Organizational and economic perspectives. *Organization Science* 11: 565–585.
Prietula, M.J., and H.S. Watson. 2008. When behavior matters: Games and computation in a behavioral theory of the firm. *Journal of Economic Behavior & Organization* 66: 74–94.
Ren, Y., K.M. Carley, and L. Argote. 2006. The contingent effects of transactive memory: What is it more beneficial to know what others know? *Management Science* 52: 671–682.
Repenning, N.P. 2002. A simulation-based approach to understanding dynamics of innovation implementation. *Organization Science* 13: 109–127.
Schelling, T.C. 1978. *Micromotives and macrobehavior*. New York: Norton.
Siggelkow, N., and D.A. Levinthal. 2005. Escaping real (non-benign) competency traps: Linking the dynamics of organizational structure to the dynamics of search. *Strategic Organization* 3: 85–115.
Somani, A., and L. Tesfatsion. 2008. An agent-based test bed study of wholesale power market performance measures. *IEEE Computational Intelligence Magazine* 3: 56–72.
Sorensen, R.A. 1992. *Thought experiments*. Oxford: Oxford University Press.
Sterman, J.D. 2000. *Business dynamics: Systems thinking and modeling for a complex world*. New York: McGraw-Hill.
Tesfatsion, L., and K.L. Judd (eds.). 2006. *Handbook of computational economics, vol. 2: Agent-based computational economics*. Amsterdam: Elsevier.
Thomsen, J., R.E. Levitt, J.C. Kunz, C.I. Nass, and D.B. Fridsma. 1999. A trajectory for validating computational emulation models. *Computational & Mathematical Organization Theory* 5: 385–401.
Van Horn, R.L. 1971. Validation of simulation results. *Management Science* 17: 247–258.
Werden, G., A. Joskow, and R. Johnson. 1991. The effects of mergers on price and output: Two case studies from the airline industry. *Managerial and Decision Economics* 12: 341–352.
Wolfram, S. 2002. *A new kind of science*. Champaign: Wolfram Media.
Zacharias, G.L., J. MacMillan, and S.B. Van Hemel (eds.). 2008. *Behavioral modeling and simulation: From individuals to societies*. Washington, DC: National Academies Press.

Computer Industry

Marco Corsino[1] and Paola Giuri[2]
[1]Universita di Bologna, Bologna, Italy
[2]Facoltà di Economia, Università di Bologna, Bologna, Italy

Abstract

The early history of the computer industry was dominated by a single firm, IBM, and the computing platform it launched, the System/360. In subsequent years IBM's dominance waned, but the emergence of dominant platforms remained the principle shaping competitive dynamics and the appearance of a concentrated industry structure. In such a context two main strategic factors underpin the achievement of a sustainable competitive advantage: the entry mode and the nature of expenditures spurring technological ▶ innovation. The importance of these features reverberates in the unfolding era of cloud computing and mobile devices.

Definition The computer industry comprises all businesses that design and/or produce host computers, multi-user and single user computers, microprocessor-based devices capable of supporting attached peripherals.

The origin of the computer industry dates back to the early 1950s with the appearance of mainframes,

a technological breakthrough that shaped the development of the industry until the late 1970s.

This era was dominated by a single firm, IBM, and its System/360, a single computing platform whose architectural characteristics led to the emergence of a concentrated industry structure (Bresnahan and Greenstein 1999). Along with concentration, another feature that shaped the early development of the industry was platform persistence, that is, a coordinated migration across new versions of the same platform that guaranteed backward compatibility.

Industry concentration and persistence of standards remained two major forces behind industry evolution in the 1980s. Major events of this period, such as the introduction of the IBM PC in 1981 and the appearance of the Apple Macintosh in 1984, witnessed cases of platform shift rather than radical innovations and did not undermine the stability of the computer industry. However, platform concentration and persistence did not characterize the evolution of the minicomputer segment in the 1990s because of the different nature of commercial relationships between buyers and suppliers (Bresnahan and Greenstein 1999).

Dramatic changes in the industry occurred in the early 1990s when the introduction of the 'client/server' platform reshaped the competitive landscape. At the beginning of 2000s, Wi-Fi notebooks and server-PC for e-commerce show the Internet as the main arena in computing technology. In this scenario the market for PCs strengthened its dominance of the computer industry. PC unit shipments soared by 79.6% over the period 2000–2011 and accounted for more than 95% of total computers in use in 2011. The market is structured in three major segments: mobile PCs is the largest one, accounting for 61.2% of total shipments, followed by desktop PCs with a 34.8% share and PC servers with the remaining 4% (Computer Industry Almanac Inc. 2012).

A distinguishing feature in the evolution of the industry until the 1990s was the persistence of international technological and competitive advantage of the United States (Bresnahan and Malerba 1999). In more recent years the US has maintained its prominent role, but the geographical landscape has changed because of the growth of emerging economies.

The distribution of PCs-in-use by geographical area in 2011 reveals that ten countries account for 61.6% of the total PCs in use: the United States alone hold about one-fifth (19.4%) of the total (Computer Industry Almanac Inc. 2012). Four EU countries (Germany, UK, France and Italy) jointly account for 14%. Japan accounts for about 6%, while the remaining countries in the top ten are the BRIC emerging economies (Brazil, Russia, India and China). This group commands a 22.1% share and China alone represents 12.2% of all PCs in use in 2011.

The increasing role of emerging economies in the distribution of PCs in use is also appreciated by looking at the change in the ranking since the mid-1990s. China ranked 12th in 1995 and jumped to second place in 2011. India and Brazil also climbed the ranking from the margins of the top 15 in 1995 to the 5th and 9th largest countries in 2011.

The computer industry has also experienced a remarkable process of market concentration over the period 1995–2010. Specifically, the share of the total market held by the top five vendors surged from 36.5% in 1995, to 42.4% in 2000, to 47.8% in 2005 and topped 60.9% in year 2010 (Gartner 1997, 2002, 2006, 2012). The two top vendors of PCs in 2010 were HP and Acer, which together account for 18.4% and 14.3% of the total market, respectively. The next two companies in the rank, Dell and Lenovo, account for 12.3% and 10.4%. The fifth company, Asus, grabs a share of 5.5%.

Only one company, HP, was among the top five vendors both at the beginning and at the end of the period 1995–2010. This points to a certain degree of turbulence at the top of the PC market. Nevertheless, this turbulence is primarily due to a wave of mergers and acquisitions that involved a few large companies in the market. The most remarkable deal involving two US companies was the acquisition of Compaq by HP in 2002 for US$25 billion. A second deal was the acquisition by Lenovo, a Chinese manufacturer founded in 1984, of IBM's personal computer business for US$1.75 billion in 2005. A third deal involved a Taiwanese producer, Acer Inc.,

which acquired Gateway for US$710 million in 2007.

One important strategic factor driving competitive advantage in the computer industry is the *entry mode*. During the early stage of the industry, under the hegemony of the IBM 360 platform, any direct confrontation with the market leader through the introduction of rival platforms ultimately failed (Bresnahan and Malerba 1999). Entry via a compatibility strategy instead proved more valuable since the 1970s, when pressures from the US antitrust authorities caused IBM to unbundle its hardware and system software. Indeed, vertical disintegration in the industry favoured the entry of smaller entrepreneurial firms producing products and components compatible with existing platforms.

However, the most valuable alternatives were either to avoid a head-to-head confrontation with IBM by uncovering market niches not served by the dominant firm or, since the mid-1980s, to enter the market with a branded clone strategy. New market segments and niches, like the minicomputer and microcomputer segments, were mainly populated by new firms not active in the main existing markets. They were characterized by new platforms and technologies and by demand from new users, or from old users in a radically different way.

In this industry the heterogeneity of buyers played a very important role for the persistence of platforms and emergence of new segments and firms (Bresnahan and Greenstein 1999). The presence of learning by experience with existing platforms and the cost of switching platform for users raised barriers to entry in existing markets and thus supported the persistence of platforms. However, different categories of buyers, such as experienced and technically sophisticated users in large firms and laboratories, professional users in commercial activities and domestic users, valued technological innovations and product attributes differently. This heterogeneity spurred entry and competition in new market segments.

A second important factor concerns R&D investments and the nature of costs spurring continuous innovation. The computer industry demands high R&D intensity, with an average ratio of R&D expenditures over sales of 7.6% in 2010 (European Commission 2012). Nonetheless, the size of the market makes it possible for firms to spread the research costs over a huge amount of shipments. Hence, economies of scale due to high R&D fixed expenditures, albeit significant, are not sufficient to gain a sustained competitive advantage (Booz & Company 2011). To gauge the forces that trigger superior performance and lead to a concentrated market it is necessary to look at the platform level where endogenous sunk costs (Sutton 1991) play a key role. In particular, the following features of such costs are worth considering (Bresnahan and Greenstein 1999): (1) irreversibility, that is, development and coordinating costs for sellers as well as costs linked to in-house software development and training for buyers are sunk; (2) specificity, that is, expenditures to improve a platform bear positive effects on the demand for components of the platform, regardless of who the seller is; (3) unlimited efficacy, that is, the value of platforms rises when rapid technical progress occurs in individual components or the span of compatibility across components gets larger; (4) (near-)unanimity about efficacy, that is, when many potential customers react to expenditures.

Another important factor accounting for the profitability of companies operating in the industry is the pursuit of distinctive innovation goals (i.e., innovative products and services) and the building of distinctive cultural attributes (i.e., openness to new ideas from customers, suppliers, competitors, and other industries) (Booz & Company 2011). For example, Apple's innovation strategy was a key success factor for becoming the second most valuable company in the US by market capitalization in 2010 and the most innovative company of the year.

As for innovation goals, empirical evidence from the computer workstation market points out that, on average, firms with a larger breadth of product portfolio are less likely to exit the market than more focused competitors. Nevertheless, the benefits arising from multiple offerings decrease as the number of rival offerings gets larger. Furthermore, by culling the product line managers

can significantly lower the likelihood of market exit (Sorenson 2000).

Sustained technological innovation continues to be a key factor affecting the evolution of the computer industry. The convergence of computing and communication technologies opens up new roads for the future. Two major developments are worth noting. First, the unfolding of the era of cloud computing where huge data centres hosting vast storage systems and servers provides, through the Internet, the processing power once offered by PCs in homes and offices (The Economist 2009). Second, the market for smartphones and other mobile devices is expanding rapidly and it is alleged to overcome PCs as the prominent market in the future (Gartner 2011). Although triggered by impressive ▶ technological changes, competition in this new market is still guided by the same principle that dominated the computer industry since its inception, namely the emergence of dominant platforms that drive concentration within specific segments (Asymco 2012).

See Also

- ▶ Innovation
- ▶ Software Industry
- ▶ Technological Change

References

Asymco. 2012. The evolution of the computing value chain.

Booz & Company. 2011. The global innovation 1000. Available at http://www.strategy-business.com/article/11404

Bresnahan, T.F., and S. Greenstein. 1999. Technological competition and the structure of the computer industry. *Journal of Industrial Economics* 47: 1–40.

Bresnahan, T.F., and F. Malerba. 1999. Industrial dynamics and the evolution of firms' and nations' competitive capabilities in the world computer industry. In *Sources of industrial leadership: Studies of seven industries*, ed. D.C. Mowery and R.R. Nelson. Cambridge: Cambridge University Press.

Computer Industry Almanac Inc. 2012. Computers-in-use forecast by country.

European Commission. 2012. The 2012 EU Industrial R&D investment scoreboard. Available at http://iri.jrc.ec.europa.eu/scoreboard12.html.

Gartner. 1997. Dataquest reports worldwide PC market thrived in 1996 with 18 percent growth; Compaq is still the market leader, but IBM, HP, and Dell pick up the pace.

Gartner. 2002. Gartner dataquest says 2001 is a year battered PC vendors would rather forget.

Gartner. 2006. Gartner says EMEA region became largest PC market in the world based on unit shipments in 2005.

Gartner. 2011. Gartner says android to command nearly half of worldwide smartphone operating system market by year-end 2012.

Gartner. 2012. Gartner says worldwide PC shipments in fourth quarter of 2011 declined 1.4 percent; year-end shipments increased 0.5 percent.

Sorenson, O. 2000. Letting the market work for you: An evolutionary perspective on product strategy. *Strategic Management Journal* 21: 577–592.

Sutton, J. 1991. *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration*. Cambridge, MA: The MIT Press.

The Economist. 2009. Cloud computing: Clash of the clouds, October. Available at http://www.economist.com/node/14637206

Concept of Strategy and Organizational Evolution

Robert A. Burgelman
Stanford University, Graduate School of Business, Stanford, CA, USA

Abstract

The historical development of the concept of strategy and its role in organizational evolution approached through normative (prescriptive) and positive (descriptive) research lenses is briefly discussed, highlighting the contributions made by scholars rooted in different social science disciplines. A high-level integrative framework – the strategy diamond – is presented. This combines the positional and resource-based approaches to strategy as well as the formulation and implementation aspects of strategy, and proposes that a company's internal selection environment – reflected in its strategic leadership culture – is key in securing alignment and realignment in the face of a highly dynamic external selection environment.

Definition The concept of strategy plays a proactively adaptive role in organizational evolution through the rational determination of the organization's purposes and objectives, its product (and/or service) market-positioning, and the planning of the use of its resources and the deployment of its capabilities to achieve success in competition and/or collaboration in the context of a highly dynamic external selection environment.

On the Nature and Role of Strategy

In the broadest sense, the concept of strategy concerns the rational determination of an organization's purposes and objectives, the strategic positioning of the company in its external environment and the planning of the use of resources and the deployment of capabilities so as to achieve purposes and objectives and achieve success in competitive and/or collaborative endeavours. Strategy is concerned with the external and internal forces that have the potential to materially affect the company's destiny. Destiny is an archaic idea, indicating a fixed and inevitable future. Strategy, by contrast, is a modern idea, indicating an open-ended future that can be determined by strategic choice. It implies degrees of freedom and the possibility, to some extent, to choose one's future (Burgelman 2002a).

Strategy has a strong thinking component. It is forward-looking and concerned with exploring multiple scenarios, alternatives and ► real options. It is externally focused and tries to anticipate states of nature and the behaviour of the relevant actors – competitors and partners – in a situation. Incisive strategic thinking at its best requires considerable intellectual effort. But senior executives sometimes view strategy with scepticism, because great strategies are just that: great strategies, or plans. From the perspective of key players, strategy becomes real when significant resources are committed, when strategy is turned into action. Strategic action is *consequential*: it involves commitments that cannot easily be undone and moves the company in a direction that is not easily reversible. Waiting to act, however, can also have strategic value if it allows useful additional information to become available. This suggests a criterion for distinguishing strategy from tactics: action is tactical if its outcomes do not significantly affect subsequent degrees of freedom to act. In dynamic environments, however, this distinction is often difficult to know in advance. Hence, alertness to the potential consequences of actions taken or not taken is one key attribute of strategic leadership.

Strategy in Structured and Unstructured Situations

Strategy in large, established organizations takes the form of *strategy-making*, a complex process involving the thinking and action of key actors situated throughout the organization. In today's global competitive environment such complex organizations usually face other complex organizations.

The study of strategic interaction between, and strategy-making within, complex organizations involves unstructured situations and is somewhat different from the study of strategic behaviour in well-structured situations. In well-structured situations all the competing players are known and each player is a rational actor whose strategic moves are drawn from a predetermined set. Particular combinations of players' strategic moves have clearly defined, if sometimes probabilistic, payoffs. Such strategic situations lend themselves well to the quantitative methods of decision theory and game theory (Saloner 1994). In the case of unstructured situations, all potential players (organizations) may not be known in advance, the strategies of differently positioned actors within each competing organization may not be well aligned, strategic moves are not limited to a predetermined set of options and payoffs in competitive interaction between players (organizations) are not always clearly defined. The study of strategy-making involving complex organizations is therefore likely to be relatively untidy, and more difficult to capture in relatively simple analytical models.

Strategy and Organizational Evolution

The concept of strategy can be usefully linked with the organizational learning and organizational ecology perspectives of evolutionary organization theory (Hannan and Freeman 1989; Mintzberg et al. 1998). The organizational learning perspective focuses on how organizations search for information to try to adapt; that is, proactively manage their fit with the external selection environment, through internal variation, selection and retention processes. While organizational learning does not necessarily lead to organizational adaptation – organizations can learn the wrong lessons! – this perspective leaves room for cognitive processes and knowledge development that is purposeful, even if only myopically so, in driving organizational change. Strategy-making as adaptive organizational capability is one manifestation of the organizational learning perspective in evolutionary organization theory.

The organizational ecology perspective, on the other hand, suggests that organizational change must be understood at the level of entire populations of similar organizations, and as the result of replacement and selection rather than of adaptation. Incumbent companies fail in the face of environmental change because inertia prevents them from adapting and are replaced with new ones that do different things or the same things differently ('better', in the eyes of the majority of customers). The study of Intel's exit from the DRAM business, for instance, adds some empirical evidence in support of organizational ecology. Organizational ecology, however, leaves little room for adaptation based on strategy. Yet strategy-making processes clearly helped Intel transform itself from a memory company into a microprocessor company, thereby preventing its demise (Burgelman 1991, 1994). Hence, organizational ecology does not always provide a complete explanation of organizational change.

Established organizations continue to remain subject to the selection force of the external environment. Many do in fact succumb to it in the long run (Burgelman and Grove 2007). But established organizations have also gained the opportunity to substitute, to some extent, internal selection for external selection. This is the central idea of the internal ecology model of strategy-making (see further below). An established company can be viewed as an ecological system in its own right, and its survival and continued success depend on the functioning of this internal ecology. While ecological processes at the level of organizational populations (industries) involve organizational founding and disbanding rates, the internal ecology of strategy-making involves entering new businesses – or other types of organizational activities – and exiting from failing ones over time. Different parts of the internal ecology of strategy-making can be linked to different forms of adaptation, and this helps reconcile opposing ideas about various consequences of strategic change.

Analysis based on the internal ecology model of strategy-making asks questions, such as: how does an organization's strategy come about and how does it evolve? What is the link between strategy-making and inertia? Which sorts of strategy-making processes lead to major strategic change that is survival enhancing?

A Brief History of Strategy

This author's interest in the concept of strategy arose from studying the issue of 'optimal firm size' through the lens of business economics (Burgelman 1969). In the course of that study it became clear that optimal size was a highly static concept, and also an ephemeral one that could be thought of but not demonstrated given internal organizational and external environmental dynamics. Firm size could be more usefully viewed as the by-product at a particular moment in time of firm growth (Burgelman 1969: 93). And, to the extent that an optimum size could in fact be determined, it would have to be a temporary one. This drove attention away from the extensive economics-based literature on firm size to the newly emerging literature at the time about the role of strategy in the development and growth of firms (Penrose 1959; Chandler 1962; Ansoff 1965). In light of the novel insights produced by this new literature, the problem of optimum firm

size could be redefined in terms of an evolving process of developing an optimal strategy and the ongoing optimal adjustment of the organization's structure (Burgelman 1969: 149).

Selected Normative (Prescriptive) Foundations

Chandler's pathbreaking historical study *Strategy and Structure* (1962) offered 'strategy' as the unifying theoretical concept for studying the managerial actions that guide a company's development: 'Strategy can be defined as the determination of the basic long-term goals and objectives of an enterprise, and the adoption of courses of action and the allocation of resources necessary for carrying out these goals' (Chandler 1962: 15–16).

Chandler offered insight into the process through which companies develop internal competencies and capabilities in response to exogenously arising external growth opportunities. He argued that companies then seek to deliberately exploit these further through diversification and that they develop new structural and administrative arrangements to support their diversification strategy. Chandler's historical study of major US-based companies spawned a series of empirical studies in various developed countries to test his core proposition that 'structure follows strategy'. This line of research culminated in Richard P. Rumelt's *Strategy, Structure and Economic Performance* (1974). Almost contemporaneously, industrial economist Edith T. Penrose's *The Theory of the Growth of the Firm* (1959) provided another path-breaking study that elucidated the internal dynamics of a company's growth. Penrose offered novel insights the 'entrepreneurial services' that constitute a company's internal impulse to grow and in the nature of the limitations of a company's management team that constrain the rate of growth at any moment in time. Chandler was apparently unaware of Penrose's work in preparing his manuscript, but in his endnotes (Chandler 1962: 453) recognizes the important complementarities between her findings and his own.

H. Igor Ansoff's *Corporate Strategy* (1965) drew on Chandler's pioneering study to distinguish strategic from administrative and operating decisions. He viewed strategic decisions as primarily focused on how the firm positions itself in the external environment and as different from administrative and operating decisions. He considered Chandler's study to be especially relevant for administrative decisions and Richard Cyert and James. G. March's *A Behavioral Theory of the Firm* (1963) for operational ones. Building on a series of earlier, more fragmentary contributions, Ansoff went on to construct the first comprehensive normative analytical framework for strategically managing an enterprise in the competitive environment. His framework comprises a series a concepts such as 'objectives', 'strategy' (including product-market scope, growth vector, synergy, competitive advantage), 'capabilities', 'strengths and weaknesses' and 'gap analysis' that are still of great relevance in today's research, teaching and practice of strategic management.

Kenneth B. Andrews' *The Concept of Corporate Strategy* (1971) built on a long tradition of case-based research and teaching of 'Business Policy' at Harvard Business School and extended the normative strategy framework. This extended framework encompasses not only the rational analytical foundation of strategy formulation (in line with Ansoff) but also key administrative issues related to strategy implementation (in line with Chandler), as well as the corporate value system that forms the foundation for acknowledging the corporation's societal obligations other than to its stockholders, and the personal values and aspirations of the top executives that shape the relative uniqueness of a corporation. In part inspired by Philip Selznick's concept of 'distinctive competence' in *Leadership in Administration* (1957), Andrews views the role of the CEO in terms of enlightened balancing of internal capabilities (in terms of Strengths and Weaknesses) and external demands (in terms of Opportunities and Threats), and of societal and corporate interests. The SWOT analysis associated with these balancing acts has remained a useful strategic management tool.

Related, but somewhat separately, field studies of strategy in the multinational sphere, such as

John Stopford and Louis T. Wells in *Managing the Multinational Enterprise* (1972) and Christopher A. Bartlett and Sumantra Ghoshal's *Managing Across Borders* (1989) provided important insight in the complex matrix-type relationships involving functional, product and geographical dimensions of organizational structures and their interplays with strategy-making processes.

Outside academia, management consulting firms developed normative frameworks with an emphasis on portfolio management in multi-business strategy. The Boston Consulting Group, for instance, proposed the 'growth-share matrix' – a tool based on the implications of learning curve theory – to more effectively guide financial resource allocation among a company's different businesses as they move through their lifecycle. These sorts of efforts on the part of the many large and small players in the strategy consulting industry are ongoing. The emergence of a 'strategy-as practice' research stream during the 2000s motivated by a renewed interest in the contributions of strategy professionals such as strategic planners and strategy consultants to strategy is noteworthy (e.g., Golsorkhi et al. 2010).

In parallel to the development of normatively oriented strategy frameworks in business schools and consulting firms, from the 1950s to the 1970s industrial organization (IO) researchers and game theorists in university economics departments had been moving beyond the classical model of perfect competition (and perfect information) to develop powerful mathematical and econometric tools for analysing the implications of information asymmetries between economic actors that made the role of strategy in modern economic competition meaningful. Michael Porter's *Competitive Strategy* (1980) and *Competitive Advantage* (1985) summarized the key insights of this intellectual tradition in a set of relatively simple frameworks that helped business executives address more sharply and systematically two important questions: (1) What determines industry attractiveness?, and (2) What determines competitive advantage? Porter's answers to these questions highlight the importance of top management seeking to understand the so-called 'five forces'; that is, the bargaining power of customers and suppliers, the nature of the rivalry among competitors (e.g., country-club like or bare-knuckle), the threat of potential new entrants and substitution. In view of this, strategy involves seeking a favourable strategic position for the firm in the industry by choosing a product-market scope (industry-wide or focused), adopting a generic strategy focused on differentiation *or* on cost leadership to provide unrivalled customer value, and carefully managing the value chain associated with the chosen generic strategy. Porter's seminal contribution spawned the so-called 'positional school' in the strategic management literature.

Porter in 'Toward a dynamic theory of strategy' (1994) further discusses the important but limited role that a wide range of situation-specific and relatively simple mathematical models – mostly gametheoretic – can play in furthering the development of a comprehensive theory of strategy and argues in favour of more complex 'frameworks' that encompass many variables and seek to capture much of the complexity of actual strategic competition. He suggests that the development of such frameworks will require detailed longitudinal case studies of companies, industries and nations.

Rooted in Penrose's theory of the growth of the firm, Birger Wernerfelt in 'A resource-based view of the firm' (1984) proposed an alternative to the position-based view of competitive advantage. Resources in the context of the resource-based view (RBV) framework are broadly defined to encompass technical, commercial as well as knowledge, administrative and cultural factors. Competitive advantages may derive from resource heterogeneity among firms and from the sustainability of resource-based advantages (e.g., Barney 1986, 1991; Dierickx and Cool 1989). C. K. Prahalad and Gary Hamel's 'The core competence of the corporation' (1990) provided a contribution to the RBV that resonated strongly among practitioners – so strongly that Michael Porter with 'What is strategy?' (1996) came back to argue in favour of strategic positioning implemented in terms of a relatively unique set of tightly integrated value activities as the sustainable basis for a company's competitive advantage. Nevertheless, the RBV continued to spawn a large

body of research and theorizing (e.g., Peteraf 2005; Kraaijenbrink et al. 2010) with special interest in the competitive role of so-called 'dynamic capabilities' (e.g., Eisenhardt and Martin 2000; Teece 2007).

Selected Positive (Descriptive) Foundations

While many original treatises of strategy were normatively oriented and focused strongly on the singular role of the CEO as enlightened 'rational actor', the empirical study of strategy did not quite support this somewhat idealized and heroic view. Based on systematic field research in a variety of organizations Brian Quinn in *Strategies for Change* (1980) realized the limitations of the top executive in leading strategic change and introduced the concept of 'logical incrementalism' as a more realistic approach. Quinn, however, continued to emphasize the dominant role of top management in the strategy-making process.

Joseph L. Bower in *Managing the Resource Allocation Process* (1970) and Henry Mintzberg in 'Patterns in strategy formation' (1978) were among the first academic researchers to pay systematic attention to strategy-making as an organizational process. Bower identified the role of 'structural context' in shaping the behaviour of managers at different levels in the organization involved in strategic capital investment projects, which suggested – in contrast to Chandler – that strategy may also follow structure. Bower and Yves L. Doz in 'Strategy formulation: a social and political process' (1978), and Bower and Clark G. Gilbert in *From Resource Allocation to Strategy* (2005), summarized a long series of academic field studies of various substantive areas related to strategy, mostly carried out at the Harvard Business School, that highlighted the cognitive and political complexities of the strategy-making process. Mintzberg defined strategy as a 'pattern' – consistency of behaviour – over time, and identified four types of patterns in strategy formation: 'deliberate', 'emergent', 'realized', and 'non-realized'. Mintzberg and Waters (1985) further explored the patterns associated with deliberate and emergent strategy.

Robert A. Burgelman in A model of the interaction of strategic behavior, corporate context and the concept of strategy' (1983a) integrated some of Chandler's, Penrose's and Bower's insights with findings about the role of multiple levels of management in internal corporate venturing into a framework that distinguishes induced strategic behaviour (driven by the current strategy) and autonomous strategic behaviour (indeterminate relative to the current strategy) in the strategy-making process. In this framework, the concept of strategy is viewed as representing the more or less explicit articulation of the firm's theory about its past and current achievements, which provides a basis for maintaining its identity and for securing continuity in its strategic activities through the induced strategy process. This view of the concept of strategy is somewhat related to C. K. Prahalad and Richard A. Bettis' 'The dominant logic: a new linkage between diversity and performance' (1986), which emphasizes the importance of 'unlearning' what has made the company previously successful in order to be able to change strategic action. Internal entrepreneurial initiatives associated with the autonomous strategy process are an alternative source of strategic renewal (Burgelman 1983b).

Further efforts to link the induced/autonomous strategy processes framework to the variation-selection-retention processes of evolutionary organization theory suggested that strategy-making could be viewed as an intra-organizational ecological process nested in higher-level (organization, population, community) ecological systems (Burgelman 1991). This also offered insight into novel sources of strategic inertia, notably the phenomenon of "co-evolutionary lock-in" (Burgelman 2002b), and additional insight into the role of path dependence in organizational strategy. These insights are somewhat related to Pankaj Ghemawat's *Commitment*: *The Dynamic of Strategy* (1991), which focuses on the relatively rare major strategic actions that are difficult to reverse (a somewhat similar view as the one discussed earlier) but does not examine how these might be potential causes of path dependence. Burgelman and Andrew S. Grove in 'Let chaos reign, then rein in chaos – repeatedly: managing strategic dynamics for corporate longevity' (2007) indicate how

strategy-making can be related to complexity and chaos theories and may help companies cope with various types of nonlinear strategic dynamics by effectively balancing concerns of maintaining 'fit' with the dynamics of the current environment (through the induced strategy process) and 'evolvability' by seeking out new environments (through the autonomous strategy process).

Throughout the late 1990s and 2000s important research streams sought to combine ideas from complexity and chaos theories (e.g., Brown and Eisenhardt 1997), exploration and exploitation in organizational learning (March 1991) and the concept of 'organizational ambidexterity' (e.g., O'Reilly and Tushman 2008) to further illuminate the strategy conundrums related to innovation, entrepreneurship, renewal, transformation and the like. These and related research streams used qualitative methods (e.g., Martin and Eisenhardt 2010), computational models (e.g., Levinthal and Posen 2007) and a very wide variety of large-sample empirical studies.

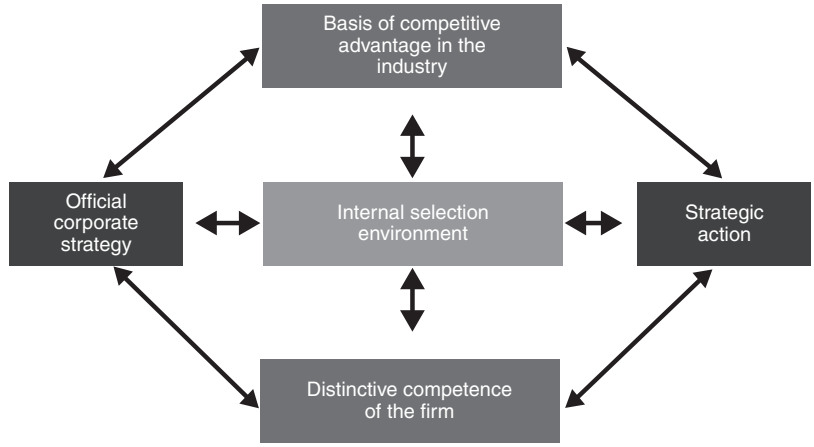
An Integrative Framework of Strategy

In the 50 years since the publication of *Strategy and Structure*, the field of strategic management has blossomed into many different theoretical views. Henry Mintzberg, Bruce Ahlstrand and Joseph Lampel in *Strategy Safari* (1998), a synthesis of the field of strategic management, identified five different meanings of strategy, and ten different schools of thought. This shows the fertility of the strategy field but can be mind boggling for academics and practitioners alike. The remainder of this essay presents a simple framework – the 'strategy diamond' – to integrate some of the major ideas from the strategic management literature related to strategy.

The strategy diamond framework encompasses five dynamic forces: (1) *official corporate strategy*, which defines the nature and the scope of the business(es) a company wants to be a winner in and its intended competitive advantage relative to the other players in the industry with respect to value creation for customers; (2) *basis of competitive advantage in the industry* associated with the company's chosen product-market position in the industry, which is determined by the industry forces; (3) *distinctive competencies*, which encompass the technical, commercial, and administrative and managerial competencies in which the company excels and which, by combining them in capabilities, serve to create customer value relative to competitors; (4) *strategic actions*, which refer to the consequential actions; that is, actions that involve binding trade-offs, commit a company in a strategic direction and are not easily reversed, and through which it actually uses its product-market position and distinctive competence to achieve competitive advantage; and (5) *internal selection environment*, which can be viewed as reflecting the company's strategic leadership culture and comprises the organizational contextual elements that help maintain alignment of the other four forces (e.g., personnel selection, organization structure, planning and control systems, resource allocation, measurement and reward systems, corporate values and norms). Figure 1 shows the strategy diamond framework.

As shown in Fig. 1, the strategy diamond combines linkages between strategic position (emphasized in the positional school) and distinctive competence (emphasized in the resource-based view), and between official corporate strategy (strategy formulation) and strategic action (strategy implementation) into one framework. In addition, a company's internal selection environment (reflecting the company's strategic leadership culture) serves to maintain alignment in the face of the dynamics of internal and/or external forces. Such dynamics may cause pressures on the company's profitable growth performance and misalignment in the linkages between the key forces in the framework. The associated tension creates 'strategic dissonance' in the organization, which signals that the company may be facing a 'strategic inflection point'. Resolving strategic dissonance and realigning the key forces in novel ways to take advantage of strategic inflection points is a key strategic leadership task (Burgelman and Grove 1996).

Integrated frameworks such as the strategy diamond may help guide the strategic



Concept of Strategy and Organizational Evolution, Fig. 1 The strategy diamond: dynamic forces driving company evolution (Source: Adapted from Burgelman 1994, 2002)

management field toward a dynamic theory of strategy (Porter 1994) and serve to examine the micro-foundations of dynamic capabilities (Teece 2007). First, the scalability of the strategy diamond framework – it can be scaled up and down: functional, single business, multi-business – may help facilitate integration across levels of strategy in the organization. Also, the strategy diamond framework alerts researchers and managers to two potential strategic traps. Companies that rely heavily on positional advantages shield themselves from competitive pressures but face a potential 'position trap': the security of their positional advantage may relax their diligence in continuing to hone and develop their distinctive competencies. As a result, their existing competencies may lose some of their efficiency or strength, which may make them potentially vulnerable to new, fitter competitors attacking their strategic position. On the other hand, companies that rely heavily on distinctive competence to compete vigorously with similar others may be able to sharply hone these competencies and become best in class. However, such efforts potentially create a 'competence trap': the relentless efforts to hone existing distinctive competence may make the company vulnerable to new competitors with different distinctive competencies as competitive dynamics in the industry change (e.g., Levitt and March 1988; Leonard-Barton 1992; Barnett 1997; Siggelkow and Levinthal 2005), or they may simply fail to appreciate the competitive importance of achieving a dominant strategic position (e.g., in the face of increasing returns to adoption). The strategy diamond framework may help prepare top management to better face the transient nature of all sources of competitive advantage in dynamic environments (e.g., D'Aveni 1994).

Finally, explicitly drawing attention to both official corporate strategy and strategic action highlights the reality that the effectiveness of strategic position and dynamic capabilities ultimately depends on human actors engaging in strategic action. Favourable competitive positions must be recognized and acted upon by strategic actors; and capabilities in and of themselves only constitute a potential until deployed by strategic actors. The strategy diamond may thus help researchers and managers appreciate better that in the end strategy is only as good as strategic action (execution); and that while strategy without capabilities is powerless, capabilities without strategy are aimless.

See Also

- ▶ Acquisition Strategy
- ▶ Business Strategy
- ▶ Competitive Strategy
- ▶ Real Options
- ▶ Strategic Groups
- ▶ Strategic Decision-Making

References

Andrews, K.B. 1971. *The concept of corporate strategy*. Homewood: Dow Jones-Irving.

Ansoff, H.I. 1965. *Corporate strategy: An analytical approach to business policy for growth and expansion*. New York: McGraw-Hill.

Barnett, W.P. 1997. The dynamics of competitive intensity. *Administrative Science Quarterly* 42: 128–160.

Barney, J.B. 1986. Strategic factor markets: Expectations, luck, and business strategy. *Management Science* 32: 1231–1241.

Barney, J.B. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17: 99–120.

Bartlett, C.A., and S. Ghoshal. 1989. *Managing across borders*. Boston: Harvard Business School Press.

Bower, J.L. 1970. *Managing the resource allocation process*. Boston: Division of Research, Harvard Business School Press.

Bower, J.L., and Y.L. Doz. 1978. Strategy formulation: A social and political process. In *Strategic management*, ed. C. Hofer and D. Schendel. Englewood Cliffs: Prentice Hall.

Bower, J.L., and C.G. Gilbert (eds.). 2005. *From resource allocation to strategy*. Oxford: Oxford University Press.

Brown, S.L., and K.M. Eisenhardt. 1997. The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly* 42: 1–34.

Burgelman, R.A. 1969. Optimale grootte in bedrijfseconomisch perspectief. Unpublished Licenciate dissertation, Antwerp University, Faculty of Applied Economics (UFSIA).

Burgelman, R.A. 1983a. A model of the interaction of strategic behavior, corporate context, and the concept of strategy. *Academy of Management Review* 8: 61–70.

Burgelman, R.A. 1983b. Corporate entrepreneurship and strategic management: Insights from a process study. *Management Science* 29: 1349–1364.

Burgelman, R.A. 1991. Intraorganizational ecology of strategy making and organizational adaptation: Theory and field research. *Organization Science* 2: 239–262.

Burgelman, R.A. 1994. Fading memories: A process theory of strategic business exit in dynamic environments. *Administrative Science Quarterly* 39: 24–56.

Burgelman, R.A. 2002a. *Strategy is destiny: How strategy-making shapes a company's future*. New York: Free Press.

Burgelman, R.A. 2002b. Strategy as vector and the inertia of coevolutionary lock-in. *Administrative Science Quarterly* 47: 325–357.

Burgelman, R.A., and A.S. Grove. 1996. Strategic dissonance. *California Management Review* 38: 3–23.

Burgelman, R.A., and A.S. Grove. 2007. Let chaos reign, then rein in chaos – Repeatedly: Managing strategic dynamics for corporate longevity. *Strategic Management Journal* 28: 965–979.

Chandler Jr., A.D. 1962. *Strategy and structure: Chapters in the history of the American Industrial Enterprise*. Cambridge, MA: The MIT Press.

Cyert, R.M., and J.G. March. 1963. *A behavioral theory of the firm*. Englewood Cliffs: Prentice Hall.

D'Aveni, R. 1994. *Hypercompetition: Managing the dynamics of strategic maneuvering*. New York: Free Press.

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

Eisenhardt, K.M., and J. Martin. 2000. Dynamic capabilities: What are they? *Strategic Management Journal* 21: 1105–1121.

Ghemawat, P. 1991. *Commitment: The dynamic of strategy*. New York: Free Press.

Golsorkhi, D.E., L. Rouleau, D. Seidel, and E. Vaara (eds.). 2010. *Strategy as practice*. Cambridge: Cambridge University Press.

Hannan, M.T., and J. Freeman. 1989. *Organizational ecology*. Cambridge, MA: Harvard University Press.

Kraaijenbrink, J., J.-C. Spender, and A.J. Groen. 2010. The resource-based view: A review and assessment of its critiques. *Journal of Management* 36: 349–372.

Leonard-Barton, D. 1992. Core capabilities and core rigidities: A paradox in new product development. *Strategic Management Journal* 13: 248–266.

Levinthal, D.A., and H.E. Posen. 2007. Myopia of selection: Does organizational adaptation limit the efficacy of population selection? *Administrative Science Quarterly* 52: 586–620.

Levitt, B., and J.G. March. 1988. Organizational learning. In *Annual review of sociology*, vol. 14, ed. W.R. Scott. Palo Alto: Annual Reviews.

March, J.G. 1991. Exploration and exploitation in organizational learning. *Organization Science* 2: 71–87.

Martin, J., and K.M. Eisenhardt. 2010. Rewiring: Cross-business unit collaboration in multibusiness organizations. *Academy of Management Journal* 53: 265–301.

Mintzberg, H. 1978. Patterns in strategy formation. *Management Science* 24: 934–948.

Mintzberg, H., and J. Waters. 1985. Of strategies, deliberate and emergent. *Strategic Management Journal* 6: 257–272.

Mintzberg, H., B. Ahlstrand, and J. Lampel. 1998. *Strategy safari*. New York: Free Press.

O'Reilly, C.A., and M.L. Tushman. 2008. Ambidexterity as dynamic capability: Resolving the innovator's dilemma. *Research in Organizational Behavior* 28: 185–206.

Penrose, E.T. 1959. *The theory of the growth of the firm*. Oxford: Blackwell.

Peteraf, M.A. 2005. Research complementarities: A resource-based view of the resource allocation process model (and vice versa). In *From resource allocation to strategy*, ed. J.L. Bower and C.G. Gilbert. Oxford: Oxford University Press.

Porter, M.E. 1980. *Competitive strategy*. New York: Free Press.
Porter, M.E. 1985. *Competitive advantage*. New York: Free Press.
Porter, M.E. 1994. Toward a dynamic theory of strategy. *Strategic Management Journal* 12: 95–112.
Porter, M.E. 1996. What is strategy? *Harvard Business Review* 74: 61–78.
Prahalad, C.K., and R.A. Bettis. 1986. The dominant logic: A new linkage between diversity and performance. *Strategic Management Journal* 7: 485–501.
Prahalad, C.K., and G. Hamel. 1990. The core competence of the corporation. *Harvard Business Review* 68: 79–91.
Quinn, J.B. 1980. *Strategies for change: Logical incrementalism*. Homewood: Dow Jones-Irwin.
Rumelt, R.P. 1974. *Strategy, structure and economic performance*. Boston: Division of Research, Harvard Business School Press.
Saloner, G. 1994. Game theory and strategic management: Contributions, applications, and limitations. In *Fundamental issues in strategy*, ed. R.P. Rumelt, D.E. Schendel, and D.J. Teece. Boston: Harvard Business School Press.
Selznick, P. 1957. *Leadership in administration*. New York: Harper & Row.
Siggelkow, N., and D.A. Levinthal. 2005. Escaping real (non-benign) competency traps: Linking the dynamics of organizational structure to the dynamics of search. *Strategic Organization* 3: 85–115.
Stopford, J., and L.T. Wells Jr. 1972. *Managing the multinational enterprise*. New York: Basic Books.
Teece, D.J. 2007. Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal* 28: 1319–1350.
Wernerfelt, B. 1984. A resource-based view of the firm. *Strategic Management Journal* 5: 171–180.

Conglomerates

Greg Linden[1] and David J. Teece[1,2]
[1]Berkeley Research Group, LLC, Emeryville, CA, USA
[2]Haas School of Business, University of California, Berkeley, Berkeley, CA, USA

Abstract

Conglomerates are multiproduct companies in which the requirements and output of at least one product division in terms of physical capital and technical skills are very different from the others. In the United States, this organizational form was more common in the middle of the twentieth century than it is today. In emerging economies, conglomerates remain more important and may be well suited to the business environment. Competing theories attribute conglomeration either to economically efficient shareholder value enhancement or to attempts by managers to secure personal gain. Empirical studies show that, on average, conglomerates destroy value compared with similar single-product firms. However, numerous cases of successful conglomerates exist, suggesting that management matters considerably.

Definition Conglomerates are multiproduct companies in which the requirements and output of at least one product division in terms of physical capital and technical skills are very different from the others. The business logic behind conglomerates relies on the astute allocation of capital by management to the highest-yielding uses.

Conglomerates are ▶ multiproduct companies in which the requirements and output of the product divisions in terms of physical capital and technical skills are very different from each other. Put another way, a conglomerate is a manifestation of unrelated ▶ diversification.

The business units of a conglomerate may have no commonalities apart from a shared administrative structure and financial oversight by a head office, which supervises strategy and may or may not involve itself in operational matters. A conglomerate thus lies somewhere between fully integrated M-form firms and hands-off investors such as holding companies, in which the business units are granted a high degree of freedom with respect to their management approaches (Davidson 1984).

In practice, multiproduct companies may have a combination of related and unrelated divisions. An unrelated division might have been acquired for purely financial reasons, as in a leveraged buyout, or it might be a remnant of a group of activities that were inter-related but have mostly been sold or became obsolete. As long as there is one entirely unrelated division, the company as a whole will be considered a conglomerate by most researchers.

Rise and Fall of an Organizational Form

The assembly of large, complex conglomerates reached its peak in the United States in the 1960s and 1970s, when companies such as International Telephone & Telegraph (ITT) diversified from telecommunications into a broad range of sectors. After Harold Geneen became CEO in 1959, ITT used leveraged buyouts to expand into hospitality, food, automotive parts, energy, books, cosmetics and other industries, building a vast commercial empire that wasn't unwound until after his departure from the company in the early 1980s. Other notable US-based conglomerates of that era (and their acquisitive CEOs) include Gulf +Western (under Charles Bluhdorn), Litton (under Tex Thornton), Textron (under Rupert Thompson) and Transamerica (under John R. Beckett). The choice of conglomerate diversification by these and similar firms may have resulted from the relatively strong antitrust enforcement by the US government at that time, which posed barriers to attempts at related diversification by firms that were already large in their initial industry (Williamson 1992).

The trend gave rise to research (e.g., Mueller 1969) suggesting that conglomerate ▶ growth through acquisitions (often highly leveraged) reflected managerial self-aggrandizement. In more formal terms, these studies claimed that conglomerates result from an agency problem in which executives seek to make their firm large in pursuit of prestige or personal financial rewards at the expense of shareholders.

A more positive view later emerged. Williamson (1975), for example, claimed that the conglomerate form could be economically efficient because headquarters executives are able to allocate cash among divisions according to expected rates of return. Under the assumption that sharing information inside the boundaries of a firm is relatively easy, an internal capital market could be better at identifying and funding high-yield investments than financial markets allocating resources among separate firms. Similarly, Teece (1982) hypothesized that conglomerates may benefit from an accumulation of industrial experience that allows them to identify and vet promising acquisitions in ways that competing acquirers might lack.

Research has confirmed that the unrelated mergers of the 1960s allowed the acquiring firms to apply their capital budgeting expertise to the conglomerate, while the internal capital market of the merged firm benefited from freer information flows than the public capital markets (Hubbard and Palia 1999). At the time, budgeting expertise and market information were both more scarce than they are today.

Despite speculation that conglomerates might be an efficient form of organization, most of them were disassembled during the 1980s (Davis et al. 1994). The fad for building a portfolio of businesses much like a portfolio of stocks had given way to newer concepts such as ▶ core competence (Prahalad and Hamel 1990), which holds that a firm's businesses should be linked to underlying technologies around which the firm has built up specialized capabilities.

Conglomerates, often with significant ownership by a single family, remain more common in less developed economies. However, the Asian financial crisis of 1997 induced divestments by – even dissolution of – many large business groups in industrializing Asia.

In emerging economies, the conglomerate form can be a source of competitive advantage, provided that local institutions are strong enough to prevent the corruption that often comes with majority ownership by a well-connected family (Peng and Jiang 2010). Amsden and Hikino (1994) review the evidence from late industrializing countries such as India, Mexico, South Korea and South Africa and conclude that the skills for licensing and applying foreign technology seem to be transferable across unrelated industries, which provides a competitive advantage to large business groups. An alternative explanation for at least some cases is offered by Kock and Guillen (2001), who note that personal contacts with officials can be more important than industrial capabilities, especially where trade barriers reduce competitive pressure. Conglomerates trade at a premium relative to single-segment companies in Latin America and Japan (Khorana et al. 2011; Fig. 3).

The motivations for unrelated diversification have continued to be a source of ongoing debate as to whether managers making unrelated investment are pursuing promising opportunities or merely feathering their own nest at the expense of shareholders (see Maksimovic and Phillips 2013, for a review). Chatterjee and Wernerfelt (1991) analysed the high and low performers among unrelated diversifiers and found evidence that both profitable and wasteful conglomeration take place. The low performers seemed to be motivated by risk diversification, which Chatterjee and Wernerfelt interpret as evidence of investments that provide stability for managers but not shareholders because shareholders can manage risk as much or as little as they choose by diversifying their individual portfolios. The high performers did not share the same initial risk profile, which suggests that the managers in these firms were truly able to identify (unrelated) opportunities that were in the long-run interests of shareholders.

Empirical Assessments

The empirical studies of the performance of conglomerates are intertwined with the broader literature on the performance and valuation effects of diversification in general. Some diversification studies include the relatedness of business segments as a variable, but many do not.

A common style of diversification research involves checking whether diversified firms generate more value (either in profits or in stock valuation) than equivalent collections of unaffiliated firms. While there is an ongoing debate about whether diversification results in a stock price discount relative to single-segment peers, the reality is that some conglomerates trade persistently at a premium while others trade at a discount (Khorana et al. 2011: 93). This suggests that management matters.

In the empirical literature, the term 'conglomerate' is applied inconsistently. It is often used for any firm operating in more than a single industry classification of the industry census. This is not, however, always a reliable measure of relatedness. Alternative methods have been used, such as the input-output approach of Anjos and Fracassi (2015), but they are relatively rare.

In the case of the North American Industrial Classification System (NAICS) used in the United States, the level of analysis could be two-digit, three-digit or four-digit classifications, with two-digit groupings being more broad. To demonstrate the practical difficulty of distinguishing relatedness using these official classifications, consider the case of Fujifilm. As Fujifilm's primary photographic film business was eroded by the rise of digital photography, it diversified into a line of cosmetics (NAICS code 3256). While this would appear at first glance to be unrelated to film manufacturing (NAICS 3259), the two businesses were connected by collagen, a substance used in photographic film to prevent fading and in cosmetics to help maintain the elasticity of skin. Using a four-digit filter would designate this as a case of unrelated diversification. Using a three-digit filter would place these in a single industry, NAICS 325, chemical manufacturing. However, without prior knowledge of the underlying technological linkage, the correct choice would not be obvious. Hoberg and Phillips (2010: 3792) show that even products in different two-digit NAICS codes can have meaningful similarities.

A further complication is that the business segment data used in these studies are self-reported by firms based on how they choose to organize their business units for internal organizational purposes. Each segment can only be assigned one NAICS code for statistical purposes, even though business units operating within a segment might belong to different codes (Villalonga 2004).

These problems may help explain why studies of diversification and conglomeration have arrived at conflicting results about the impact on firm value. Nevertheless, the findings are worth reviewing.

Theoretically, conglomerates are able to reap the same benefits as firms diversified only in related product lines. For example, they can reallocate capital and human resources to units as needed to take advantage of positive demand shocks

(Maksimovic and Phillips 2002). But this requires that they avoid the pitfall of internal politics that can distort resource allocations in favour of the least efficient segments (Rajan et al. 2000).

Empirically, a negative outcome seems to dominate. Berger and Ofek (1995: 54) found that, while firms with more than one segment had lower profitability on average than equivalent groups of single segment firms, firms that had one or more unrelated segments performed significantly worse. Villalonga (2004) and Ammann et al. (2012: 274) also found a negative effect using different data.

A study by Ozbas and Scharfstein (2010) helps to explain why conglomerates may be inefficient. They find that multiproduct firms invest less in their unrelated segments relative to market expectations of profit potential (proxied by tobin's q) than do single-segment firms in the same industries. They suggest that ▶ agency problems between managers and shareholders may be causing the underinvestment because firms in which managers own more of the company are less likely to underinvest in promising business divisions. Their finding is supported by the results of Hoechle et al. (2012: 49), who showed that the gap between the discounts for related and unrelated diversification was reduced when controls for the quality of corporate governance were included in the analysis. In other words, agency problems, as proxied by weaker governance, account for a larger share of the diversification discount in cases of unrelated diversification.

Nevertheless, some conglomerates are able to build distinctive advantages. Anjos and Fracassi (2015) see the benefits of conglomerates as flowing from their access to information from across a variety of industries that leads to cross-industry insights not available to specialized firms. They find that conglomerates with 'high centrality' in buyer-supplier networks produce 'more and better patents'.

The wisdom or folly of conglomerate diversification clearly varies on a case-by-case basis. Good governance and wise management can turn some cases of unrelated diversification into above-normal profits.

See Also

- ▶ Agency Problems
- ▶ Core Competence
- ▶ Diversification
- ▶ Growth Through Acquisitions
- ▶ Managerial Discretion
- ▶ Multiproduct Companies

References

Ammann, M., D. Hoechle, and M. Schmid. 2012. Is there really no conglomerate discount? *Journal of Business Finance & Accounting* 39: 264–288.

Amsden, A.H., and T. Hikino. 1994. Project execution capability, organizational know-how and conglomerate corporate growth in late industrialization. *Industrial and Corporate Change* 3: 111–147.

Anjos, F., and C. Fracassi. 2015. Shopping for information? Diversification and the network of industries. *Management Science* 61: 161–183.

Berger, P.G., and E. Ofek. 1995. Diversification's effect on firm value. *Journal of Financial Economics* 37: 39–65.

Chatterjee, S., and B. Wernerfelt. 1991. The link between resources and type of diversification: Theory and evidence. *Strategic Management Journal* 12: 33–48.

Davidson, W.H. 1984. Administrative orientation and international performance. *Journal of International Business Studies* 15: 11–23.

Davis, G.F., K.A. Diekmann, and C.H. Tinsley. 1994. The decline and fall of the conglomerate firm in the 1980s: The deinstitutionalization of an organizational form. *American Sociological Review* 59: 547–570.

Hoberg, G., and G. Phillips. 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies* 23: 3773–3811.

Hoechle, D., M. Schmid, I. Walter, and D. Yermack. 2012. How much of the diversification discount can be explained by poor corporate governance? *Journal of Financial Economics* 103: 41–60.

Hubbard, R.G., and D. Palia. 1999. A reexamination of the conglomerate merger wave in the 1960s: An internal capital markets view. *Journal of Finance* 54: 1131–1152.

Khorana, A., A. Shivdasani, C. Stendevad, and S. Sanzhar. 2011. Spin-offs: Tackling the conglomerate discount. *Journal of Applied Corporate Finance* 23: 90–101.

Kock, C.J., and M.F. Guillén. 2001. Strategy and structure in developing countries: Business groups as an evolutionary response to opportunities for unrelated diversification. *Industrial and Corporate Change* 10: 77–113.

Maksimovic, V., and G. Phillips. 2002. Do conglomerate firms allocate resources inefficiently across industries? Theory and evidence. *Journal of Finance* 57: 721–767.

Maksimovic, V., and G. Phillips. 2013. Conglomerate firms, internal capital markets, and the theory of the firm. *Annual Review of Financial Economics* 5: 225–244.
Mueller, D.C. 1969. A theory of conglomerate mergers. *Quarterly Journal of Economics* 83: 643–659.
Ozbas, O., and D.S. Scharfstein. 2010. Evidence on the dark side of internal capital markets. *Review of Financial Studies* 23: 581–599.
Peng, M.W., and Y. Jiang. 2010. Institutions behind family ownership and control in large firms. *Journal of Management Studies* 47: 253–273.
Prahalad, C.K., and G. Hamel. 1990. The core competence of the corporation. *Harvard Business Review* 68: 79–91.
Rajan, R., H. Servaes, and L. Zingales. 2000. The cost of diversity: The diversification discount and inefficient investment. *Journal of Finance* 55: 35–80.
Teece, D.J. 1982. Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior and Organization* 3: 39–63.
Villalonga, B. 2004. Diversification discount or premium? New evidence from the business information tracking series. *Journal of Finance* 59: 479–506.
Williamson, O.E. 1975. *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.
Williamson, O.E. 1992. Markets, hierarchies, and the modern corporation: An unfolding perspective. *Journal of Economic Behavior and Organization* 17: 335–352.

Consulting Firms

Anne Layne-Farrar
LECG, Chicago, IL, USA

Definition A consulting firm is a firm of experts (consultants) that provide professional advice for a fee. The primary purpose of a consulting firm is to provide access to industry-specific specialists and subject matter expertise. Expertise covers a wide range – any subject in which individuals earn a specialized degree or accumulate specialized knowledge.

While business advisory services have existed in some form for over 100 years, it wasn't until a 1930 *Business Week* article reported the emergence of a new professional service dubbed 'management consulting' that the services gained their current name. Broadly speaking, the consulting industry provides firms with ready access to specialized business knowledge and processes, organizational expertise and managerial techniques. Because it is dependent on client profitability, consulting is generally a cyclical industry. The industry is expected to generate $391 billion in revenues globally in 2012, about 40% of which comes from the United States.

The Early Years

Business consulting first emerged in the US, where the early industry was driven largely by the federal government – directly and indirectly. The Sherman Act of 1890 prohibited firms from sharing information, which created a role for consultants to provide an appropriate legal channel for reporting business knowledge and best practices. More narrowly, in banking, the Glass–Steagall Act of 1933 created the need for independent consultants to examine due diligence issues, strategic planning and accounting. Finally, the Hoover Commission of the 1940s, charged with restructuring the federal bureaucracy, fostered considerable growth of management consulting by contracting with outside experts, which gave consulting firms national exposure and a reputation for implementing successful organizational models.

Growth and Specialization

Since the 1950s and 1960s, consulting has grown more multifaceted and has moved from a mostly American phenomenon to an international one. Important types of consulting now include management and strategic consulting, operations management, business advisory services, IT and computer consulting, human resource consulting and economic/antitrust consulting. While there are significant overlaps between some of these categories, the types of engagement within each segment, along with the primary suppliers, differ, as discussed below.

Consulting Firms, Table 1 Top 10 global consulting firms by estimated revenue (2009)

Firm	Major segment	Revenues ($ million)	Consultants
Accenture	IT and computers	25,300	186,000
Capgemini	IT and computers	12,279	83,500
KPMG	Tax and audit	7,270	25,000
PricewaterhouseCoopers	Tax and audit	6,900	19,000
Deloitte	Tax and audit	6,300	15,000
McKinsey & Co.	Management	5,300	9,000
Booz Allen & Hamilton	Management (government)	4,100	13,000
Boston Consulting Group	Management	2,400	6,000
Bain & Co.	Management	1,600	4,200
Booz & Co.	Management	1,000	3,200

Segments and Firms

Traditionally, management and strategic consulting has been the largest and most prevalent type of consulting. Firms in this segment of the industry assist business and government clients in improving performance, streamlining organizational structure and attaining strategic objectives. Prominent firms include Bain & Co., Boston Consulting Group, Booz Allen Hamilton and McKinsey & Co.

Operations management consulting provides advice regarding logistical challenges and the deployment of resources. While closely linked to management consulting, consultants in this sector help companies adjust their business models, business processes and organizational structure in order to streamline supply chains, manufacturing strategies, marketing and sales, and structural aspects arising from mergers and acquisitions. Important firms in this segment include A.T. Kearney, Oliver Wyman Group and PRTM.

Business advisory consulting typically provides organizations with audit, tax and risk management services. Firms in this sector undertake various reporting tasks to maintain their clients' regulatory compliance; they also provide financial accounting services, internal and external audits and actuarial services. Major firms in this segment include Deloitte, Ernst & Young, KPMG and PricewaterhouseCoopers.

IT and computer consulting is a relative latecomer but now holds a critical role in the industry. Consulting firms in this segment provide businesses with advice on how to incorporate information technology and computing to meet their business and strategic objectives. Often, consultants will also implement, deploy and manage IT platforms on behalf of their clients. Prominent firms in this segment include Accenture, Capgemini, IBM Global Services, Infosys, TCS and Wipro.

▶ Human resources consulting is a growing segment. These firms advise businesses on personnel and hiring strategies and legal compliance, salaries and benefits packages, outsourcing and training. These firms may also manage payroll and benefits plans for their clients. Important firms in this sector include Aon Consulting, Mercer and Towers Watson.

Other niches within the broad consulting industry include health care consulting, legally oriented financial consulting, antitrust and economic consulting, regulation and international arbitration. Prominent firms serving these niches include Charles River Associates, Compass-Lexecon and NERA.

Table 1 above reports statistics on the top ten global consulting firms.

See Also

- ▶ Acquisition Strategy
- ▶ Business Strategy
- ▶ Competition Policy
- ▶ Human Resources
- ▶ Information Technology and Strategy
- ▶ Market Entry Strategies
- ▶ Operations Management and Strategy

Consumer Surplus

Ishtiaq Pasha Mahmood
Department of Strategy and Policy, IMD, Singapore, Singapore

Definition Consumer surplus is defined in economics as the level of benefit (utility) that consumers derive by being able to purchase a product for a price that is less than the maximum they would be willing to pay.

Consumer surplus is defined in economics as the level of benefit (utility) that consumers derive by being able to purchase a product for a price that is less than the maximum they would be willing to pay. Gain in consumer utility is given by a function of consumer utility gained from the good purchased less the opportunity cost of foregoing the purchase of alternate goods with the same money. Consumer surplus is maximized in a theoretical world of perfect competition, in which producer surplus, or value capture, is equal to zero for all firms in the market (Varian 2005).

Business strategy endows the traditional economic definition with an additional crucial dimension when taken from the perspective of the firm: value. Firms face a trade-off when allocating resources towards creating value (i.e., innovating, producing) and appropriating value (i.e., extracting profits in the marketplace). Firms require both a focus on value creation and value appropriation to thrive, and precisely how much effort should be allocated to one versus the other varies widely under different industries, technological environments and targeted consumer groups (Mizik and Jacobson 2003).

In order to obtain greater profits, firms have traditionally attempted to appropriate as much consumer surplus as possible. Firms may seek to 'capture' consumer surplus through various pricing methods, such as skim and penetration pricing (Noble and Gruca 1999). In the case of the former, the goal of the firm is to ride down the demand curve and appropriate as much consumer surplus over time as possible. In contrast, penetration pricing entails creating the highest degree of market share possible while focusing less on immediate profits, sometimes even temporarily at their own expense (Gaimon 2008). This holds true in the case of many products which exhibit positive network externalities and create ▶ value for the firm indirectly as they become widespread (e.g., Adobe Reader and other freeware; video game systems).

Firms can also practise price differentiation among customers with different wants or information (Varian 1980), and the rise of the Internet has led to a marked increase in dynamic pricing strategies available to firms. Under dynamic pricing, firms can actively vary prices in response to changes in consumer demand, in an attempt to garner as much surplus as possible by getting consumers to pay different prices at different times, depending upon waiting costs or benefits (Su 2007). This can be readily observed in the case of online ticket sales, which are cheaper when purchased on certain dates and vary dynamically along with consumer demand. While highly effective in capturing consumer surplus, dynamic pricing is often considered to be a form of unfair price discrimination and is not without controversy. When Amazon incorporated this value appropriation strategy, it used customers' purchase history and charged different prices for the same products based upon perceived willingness to pay by individual customer profiles, thereby generating significant negative publicity and ethical questions.

Although firms traditionally focus on appropriating consumer surplus, alternate strategies with a focus on creating value for non-traditional market segments represent a potentially lucrative approach to consumer surplus and value creation. Indeed, value creation is a precondition for value capture, and it is inimical to long-run survival to focus solely on capture without adding economic value (Priem 2007). By targeting such groups as BoP (base of the pyramid) consumers that have been ignored by others, for instance, firms can create genuine value for markets where it did not exist before, while generating profit in the process (Prahalad and Hammond 2002; Prahalad and Brugmann 2007). As an example, cheap solar kits sold in Africa help to create power and enterprise in areas power grids do not reach. In India, mobile banking has enabled the urban poor to send money to family in villages easily and instantly. By concentrating on

a bottom-up focus on product innovation and value creation, firms can profit from large non-traditional markets while creating immense consumer value where it did not previously exist. With such abundant opportunities to pursue profits while simultaneously increasing social welfare and creating new value for end users, consumer surplus need not be viewed as a zero-sum game.

See Also

- ► Profit
- ► Value

References

Gaimon, C. 2008. The management of technology: A production and operations management perspective. *Production and Operations Management* 17: 1–11.

Mizik, N., and R. Jacobson. 2003. Trading off between value creation and value appropriation: The financial implications of shifts in strategic emphasis. *Journal of Marketing* 67: 63–76.

Noble, P., and T. Gruca. 1999. Industrial pricing: Theory and managerial practice. *Marketing Science* 18: 435–454.

Prahalad, C.K., and J. Brugmann. 2007. Co-creating business's new social impact. *Harvard Business Review* 85: 80–90.

Prahalad, C.K., and A. Hammond. 2002. Serving the world's poor, profitably. *Harvard Business Review* 80: 48–57.

Priem, R. 2007. A consumer perspective on value creation. *Academy of Management Review* 32: 219–235.

Su, X. 2007. Intertemporal pricing with consumer strategic behavior. *Management Science* 53: 726–741.

Varian, H. 1980. A model of sale. *American Economic Review* 70: 651–659.

Varian, H. 2005. *Intermediate microeconomics: A modern approach*, 7th ed. New York: W.W. Norton & Company.

Contestability

Kira Rudie Fabrizio
Duke University, The Fuqua School of Business, Durham, NC, USA

Definition Contestability refers to the degree to which the actual and potential entry of new competitors restricts the market price to the competitive level, thereby eliminating the potential for firms in the industry to earn positive economic rents (i.e., profits).

Contestability refers to the degree to which the entry and potential entry of new competitors restricts the ► market price to the competitive level, thereby eliminating the potential for firms in the industry to earn positive economic rents.

The determining characteristic of the contestability of a given market is the ability of potential competitors, with the same cost functions as incumbent firms, to enter and exit the market without loss of capital investments. This turns directly on the existence of barriers to entry and exit; when there are no barriers to entry and exit, the market is contestable, and the presence of actual and/or potential entrants results in a competitive market price. If incumbent firms were to raise prices above this level, entry would occur immediately and quickly reduce the market price. Thus, contestable markets are said to be subject to 'hit-and-run' entry, wherein entrants can easily and quickly enter, compete down prices and freely exit the market.

In contestable markets, entrants are assumed to be able to compete at no disadvantage to incumbents, and enter and exit without loss of capital. Barriers to entry and exit decrease the contestability of the market, and thereby provide an opportunity for incumbents to raise prices above competitive levels and earn profits. These barriers can be usefully characterized as sunk costs, scale economies, incumbent cost and other advantages, legal and regulatory restrictions, retaliation and exit costs.

Sunk costs play a particularly important role in the theory of contestable markets. Sunk costs are costs that are not recoverable if a firm exits the market. Examples include investments in assets that are specialized to that market, such as physical manufacturing assets, research knowledge, brand, or any other investments that are worth less in their next best use. When large sunk cost investments are required for entry into a market, it is harder for firms to enter (because of the capital requirements) and to exit (because of the lost asset value). The presence of sunk costs thereby protects incumbents from the threat of hit-and-run entry.

The condition of a perfectly contestable market, like that of perfect ▶ competition in economics, is more useful as a foil than as a description of reality. The conditions necessary for a market to be perfectly contestable are very rare, if not non-existent. However, the theory of contestability is valuable as a way to consider where the assumptions fail, or, in other words, why a particular market is not contestable. By understanding what prevents a market from being contestable, firms are able to understand the structural forces that allow for the potential to earn rents in the market. Such an understanding allows managers to anticipate the expected potential for profitability in a given market and pursue opportunities accordingly.

Even more proactively, incumbent firms can take steps to enhance the profitability of a market by decreasing the contestability of the market. This could take the form of creating entry barriers, for example by trying to induce the establishment of regulatory hurdles for new firms to enter. Firms could also pursue a strategy to increase the extent to which investments required to enter and compete are specialized to the industry, and thus sunk in nature. Strategies for doing so include investments in advertising, research and development, and specialized equipment. Such sunk costs create a barrier to entry and exit and protect the firm from profit destroying hit-and-run entry. Finally, to the extent that the market is (or can be) characterized by scale, scope and learning economies, incumbent firms can craft strategies to exploit these cost-reducing opportunities as a means to create cost advantage relative to potential entrants, thereby reducing the contestability of the market and improving the potential to earn profits.

See Also

- ▶ Competition
- ▶ Exit Barriers
- ▶ Market Price

Contracting Out

Kulwant Singh
National University of Singapore, Business School, Singapore, Singapore

Definition Contracting out is the process by which a firm formally agrees to repeated purchases of goods or services from an external provider instead of producing these internally.

Contracting out occurs when a firm enters into a contractual agreement to purchase products or services that it could potentially have produced internally. This reflects the managerial choice, in a ▶ 'make-or-buy' decision, not to undertake an activity within a firm but to purchase a product or service from the market. To the extent that no firm produces all the components of its products or services internally, all firms undertake some degree of contracting out.

Economic perspectives attribute contracting out to efficiency considerations, and to external purchasing potentially offering lower total costs than internal production (Coase 1937; Williamson 1985). Resource-based perspectives explain contracting out as resulting from firms choosing to purchase lower value-added activities from the market, to free scarce resources for activities that generate greater value (Barney 1991). Strategy research conditions these arguments, indicating that firms may undertake internal production even if economically inefficient or when resources are inadequate, for strategic reasons such as retaining the option to develop important capabilities in the future or to gain from spillover benefits (Leiblein and Miller 2003).

▶ Transaction cost economics (Coase 1937) provides an analytical framework that embeds economic and resource-based arguments for contracting out. This perspective proposes that contracting out results from comparing the total costs of internal production and of administrating the production process, against the purchase price

of the product and the transactions costs of purchasing from the market (Williamson 1985). In the transaction cost perspective, contracting out is evidence that the total costs of internal production exceed the total costs of purchasing from the market.

The seemingly simple decision to contract out has major implications for firm strategy, structure and performance. Firms may contract out to reduce the scope of their activities and to focus resources and capabilities on greater value-creating activities, in order to improve firm performance. This will affect the structure and management of the firm, its ▶ vertical integration and boundaries, and the resources and capabilities it has to develop. Hence, contracting out offers the advantages of greater specialization and simplified internal structure and administration. However, increased contracting out is associated with greater costs to manage the agreement and enforcement of contracts, and to manage the quality and integration of externally produced products.

Contracting out replaces the flexibility of control over internal production with the limited price-based coordination offered in market exchange. Contracting out is undertaken on terms agreed through formal arm's length contracting between a firm and external organizations. Parties to these contracts typically attempt to specify all relevant terms *ex ante* so that contracts are legally enforceable *ex post*. However, it is not possible to write contracts that fully specify all contingencies, except under conditions of perfect competition. As a result, contracting out is vulnerable to the hazards of market exchange, which may be particularly high for exchange under conditions of complexity, uncertainty and significant investments in transaction-specific assets (Williamson 1985).

To overcome the difficulties of costly contracting in non-perfectly competitive markets, firms often convert arm's length contracting into related-party exchanges. These relational contracts are an intermediate arrangement between arm's length contracting and internal production. Relational contracting assumes incomplete contracts, encourages information sharing and coordination, and allows *ex post* interpretation of terms by the parties. This provides the parties with some flexibility in adapting the terms of contracting out to meet contingencies, helping address the disadvantages of arm's length contracting.

Contracting out has traditionally focused on basic components or support services. However, improvements in information and communications technology, logistics and quality management systems, and the expansion of supplier networks have broadened the scope of contracting out. Examples of the expanded scope of contracting out include firms purchasing final products from suppliers or assigning the entire manufacturing process to other firms; the ▶ outsourcing of financial, administrative and other non-core services to external organizations; and government organizations contracting with private organizations for the delivery of public products and services. As a result, the scope and importance of contracting out is increasing significantly. Collectively, these decisions on contracting out affect the structure and organization of industries and the efficiency of markets.

See Also

- ▶ Make-or-Buy Decisions: Applications to Strategy Research
- ▶ Outsourcing
- ▶ Transaction Cost Economics
- ▶ Vertical Integration

References

Barney, J.B. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17: 99–120.

Coase, R. 1937. The nature of the firm. *Economica* 4: 386–405.

Leiblein, M.J., and D.J. Miller. 2003. An empirical examination of transaction and firm-level influences on the vertical boundaries of the firm. *Strategic Management Journal* 24: 839–859.

Williamson, O. 1985. *The economic institutions of capitalism*. New York: Free Press.

Cooper, Arnold C. (Born 1933)

Charlotte R. Ren[1] and Timothy B. Folta[2]
[1]University of Pennsylvania, Philadelphia, PA, USA
[2]University of Connecticut, Storrs, CT, USA

Abstract

Arnold C. Cooper is a pioneer and a world-renowned authority on entrepreneurship. He has made key contributions to the fields of entrepreneurship and strategic management in diagnosing the 'small firm effect', and a more complex understanding of both the process of new firm formation and predictors of new venture success.

Born in Chicago in 1933, Arnold C. Cooper was brought up in New Castle, Indiana. He attended Purdue University and earned his Bachelor of Science degree in chemical engineering in 1955 and his Master's degree in industrial administration in 1957. After a year of working for Procter & Gamble, Cooper earned a doctorate in business administration in 1962 at Harvard University, where he also served as an assistant professor. He returned to Purdue University in 1963 as an associate professor in the Krannert School of Management, becoming professor in 1970 and the Louis A. Weil, Jr. Professor in 1984, which role he retained until his retirement in 2005. He was a visiting professor or scholar at Stanford University, Manchester Business School in England, IMD in Switzerland and the Wharton School at the University of Pennsylvania.

Cooper is the author or co-author of seven books and has written extensively on entrepreneurship, strategic management and the management of technology. His work has yielded a number of honours and awards, including the 1997 'Global Award for Entrepreneurship Research', a lifetime contribution award; the 1999 'Richard D. Irwin Outstanding Educator Award' from the Division of Business Policy and Strategy of the Academy of Management; the 1993 'Coleman Entrepreneurship Mentor Award' from the Entrepreneurship Division of the Academy of Management; the 2009 'Entrepreneurship Educator of the Year' from the United States Association for Small Business and Entrepreneurship; and the 2012 'IDEA Foundational Paper Award', given by the Entrepreneurship Division of the Academy of Management for a paper that significantly changed the conversion in the field of entrepreneurship for at least a decade. He is a Fellow of the Academy of Management and the Strategic Management Society. He has served as major professor for five doctoral students who have won prizes for their dissertations. Purdue University awarded him an Honorary Doctor of Management degree in 2005. In addition to his scholarly contributions to the field of strategic management, Cooper played an integral role in the field's early development. He organized the first conference in entrepreneurial research in 1970 at Purdue University, and was Chairman of the Division of Business Policy and Strategy in the Academy of Management.

Cooper's early work examined product development in small firms; one finding suggested that small firms could develop new products at lower cost than large firms (Cooper 1964). He later examined the processes of new firm formation. He may have been the first to diagnose what is currently known as the 'small firm effect', the tendency for smaller firms to spawn more entrepreneurial ventures. In examining the founding of 250 technically oriented firms in Silicon Valley, he found that the spin-off rate in smaller firms was approximately ten times greater than in large companies (Cooper 1971). Later, he was involved in examining the responses of established firms to major technological threats (Cooper and Schendel 1976).

A key contribution to the early development of the scholarly study of entrepreneurship was Cooper's pioneering effort in developing, with colleagues, a large-scale, longitudinal data set. They tracked, over 3 years, 2994 entrepreneurs who were members of a trade association (National Federation of Independent Business). This paved the way for the rigorous study of two important

streams in entrepreneurship research. The first stream centred on processes of new firm formation. One aspect of this looked at entrepreneurs' perceived chances of success. Cooper and colleagues found that about 95% of entrepreneurs perceived their own odds for success better than others' odds, and this optimism was not influenced by objective predictors such as education, experience and capital (Cooper et al. 1988). This study suggested the salient existence of what psychologists would term 'post-decisional bolstering' among entrepreneurs, which has immense theoretical and practical implications for understanding entrepreneurs' decision-making processes and risk-taking propensity. The second stream was concerned with predictors of new venture success, one of the most important areas of research in entrepreneurship. Among the long list of predictors examined were human capital, financial capital, parents' entrepreneurial experience, prior working experience and the similarity of the new business to the entrepreneur's prior organizations (Cooper et al. 1994). One interesting conclusion was that these determinants often differentially affected failure (discontinuance), marginal survival and high growth. This conclusion inspired another study, which found support for the idea that different entrepreneurs have different performance thresholds – defined as the level of performance under which the firm will exit (Cooper et al. 1997). This study explains why financial performance does not exclusively influence entrepreneurial firm survival, and was awarded the 2012 'IDEA Foundational Paper Award'.

Cooper's research contributions continue to serve as a platform for many in the field of strategic management and entrepreneurship.

See Also

- ▶ Business Policy and Strategy
- ▶ Corporate Venturing
- ▶ Innovation
- ▶ Intrapreneurship
- ▶ Technological Change

Selected Works

Cooper, Arnold C. 1964. R&D is more efficient in small companies. *Harvard Business Review* 42: 75–83.
Cooper, Arnold C. 1971. Spin-offs and technical entrepreneurship. *IEEE Transactions on Engineering Management* 19: 2–6.
Cooper, Arnold C., and D. Schendel. 1976. Strategic responses to technological threats. *Business Horizons* 19: 61–69.
Cooper, Arnold C., C.Y. Woo, and W.C. Dunkelberg. 1988. Entrepreneurs' perceived chances for success. *Journal of Business Venturing* 3: 97–108.
Cooper, Arnold C., J. Gimeno, and C.Y. Woo. 1994. Initial human and financial capital as predictors of new venture performance. *Journal of Business Venturing* 9: 371–395.
Cooper, Arnold C., J. Gimeno, and C.Y. Woo. 1997. Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms. *Administrative Science Quarterly* 42: 750–783.

Cooperation and Competition

David J. Teece
Berkeley Research Group, LLC, Emeryville, CA, USA
Haas School of Business, University of California, Berkeley, Berkeley, CA, USA

Definition Cooperation and competition may occur simultaneously between firms for strategic, non-contradictory reasons.

Cooperation and competition are potentially complementary modes of engagement among firms, especially in high-tech industries. A single set of firms may, for example, collaborate on pre-competitive standards or process technology, then compete fiercely against each other in product markets. Or Company X may cooperate with one division of Company Z while engaging in fierce rivalry with another of Z's business units.

Cooperation often takes place within ▶ strategic alliances, which are constellations of bilateral and possibly multilateral contracts and understandings among firms. These complex linkages have become more possible (and necessary) than ever due to the global dispersion of technical know-how, the liberalization of trade, and the

reduction of communication costs. Finding the most suitable collaborators has also become imperative for success because the same factors have made competition more global and intense than ever before.

Strategic alliances have become quite common, particularly in high-technology industries. Joint R&D, know-how, manufacturing and marketing agreements are potentially more valuable than standard, market-based ties because they can be used to access complementary technologies and assets without direct ownership. As compared with the price system of market exchange, alliance networks enable investment plans for complementary assets to be coordinated more concisely; as compared with hierarchy (internalization), incentives are not dulled through bureaucratic decision-making.

In some cases, these complex relationships can be better protected from contractual hazards than equivalent discrete contracts. This can be true when, in addition to the normal transaction-specific safeguards, a complex alliance encompasses broader relational and multi-transactional structures (de Figueiredo and Teece 1996).

To profit in this type of complicated ecosystem, firms need to coordinate with their partners not only operationally but strategically (Teece 1992). Operational coordination involves the everyday management of the collaborative activity. Strategic coordination refers to activities that affect the distribution of returns to collaboration by shaping access to complementary goods, appropriability from the firm's own intangible assets, and industry structure (Teece 1986; Pisano and Teece 2007). A game-theoretic approach has been used extensively to evaluate strategies for '► co-opetition' (Brandenburger and Nalebuff 1996).

Although antitrust authorities sometimes look askance at cooperation, complex forms of cooperation are often necessary to promote healthy competition. This can be true in industries such as semi-conductors, where the development of next-generation technologies require investments beyond the capabilities of most firms in the industry. Even if the cost of R&D is affordable, collaboration can reduce some wasteful duplication of effort and/or help an industry with weak intellectual property rights to break out of a low R&D equilibrium induced by a 'free rider' problem. Horizontal linkages can also assist in the definition of technical standards.

To preserve rivalry, research collaboration should not expand to encompass all firms in an industry. Were that to occur, the benefits of pluralism in research approaches would not be realized. The smartphone industry, for example, benefits from having networks of firms contributing to the development of different platforms (e.g., iOS, Android etc.) that compete against each other (as well as, in many cases, competing within a network).

Another important motive for cooperation is access to complementary assets. The profitable commercialization of technology requires timely access to complementary assets on competitive, if not preferential, terms. Thus, an innovating firm or consortium that has developed the core technology needed for a new product or process with good prospects has taken only the first step. It must next secure access to complementary technologies and complementary assets on favourable terms in order to successfully commercialize the product or process. Apple's reliance on Samsung for the primary microchips in its early iPhone and iPad models, despite Samsung's role as a competitor in mobile electronics, provides a clear example.

In setting limits for cooperation, competition policy must balance, at the margin, the partially exclusive values of competition and pluralism. A sensible balance is vital to a well-functioning national system of innovation.

See Also

- ► Co-opetition
- ► Inter-firm Cooperation
- ► Research and Development (R&D) Alliances
- ► Strategic Groups

References

Brandenburger, A.M., and B.J. Nalebuff. 1996. *Co-opetition*. New York: Doubleday.

De Figueiredo, J.M., and D.J. Teece. 1996. Mitigating procurement hazards in the context of innovation. *Industrial and Corporate Change* 5: 537–559.

Pisano, G., and D.J. Teece. 2007. How to capture value from innovation: Shaping intellectual property and industry architecture. *California Management Review* 50: 278–296.

Teece, D.J. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy* 15: 285–305.

Teece, D.J. 1992. Competition, cooperation, and innovation: Organizational arrangements for regimes of rapid technological progress. *Journal of Economic Behavior & Organization* 18: 1–25.

Cooperative and Non-cooperative Game Theory

Olivier Chatain
University of Pennsylvania, Wharton School, Philadelphia, PA, USA

Abstract

This article outlines the differences between cooperative and non-cooperative ▶ game theory. It introduces some of the main concepts of cooperative game theory as they apply to strategic management research.

Definition Cooperative game theory focuses on how much players can appropriate given the value each coalition of player can create, while non-cooperative game theory focuses on which moves players should rationally make.

Game theory comprises two branches: cooperative ▶ game theory (CGT) and non-cooperative game theory (NCGT). CGT models how agents compete and cooperate as coalitions in unstructured interactions to create and capture ▶ value. NCGT models the actions of agents, maximizing their utility in a defined procedure, relying on a detailed description of the moves and information available to each agent. CGT abstracts from these details and focuses on how the value-creation abilities of each coalition of agents can bear on the agents' ability to capture value. CGT can be thus called *coalitional*, while NCGT is *procedural*. Note that 'cooperative' and 'non-cooperative' are technical terms and are not an assessment of the degree of cooperation among agents in the model: a cooperative game can as much model extreme competition as a non-cooperative game can model cooperation.

CGT analyses situations where agents can cooperate to create value by joining coalitions, but also where agents compete to capture value. Its formalism comprises a set of agents and a function that returns the value each subset of agent (i.e., each *coalition*) can create on its own (the *characteristic function*). The characteristic function is the input to a *solution concept*, which returns the value captured by each agent (their *imputation*).

Because no structure is put on the bargaining procedure, the solution concept models the outcome of the negotiations occurring among the agents, accounting for the value each coalition can create. We concentrate on the solution concepts of the core and the Shapley value, but a number of others have been developed. For textbook expositions, see, for instance, Osborne and Rubinstein (1994) and Owen (1995).

The core is the set of imputations such that each coalition receives at least as much as it can create on its own. This concept formalizes a notion of individual and coalitional self-interest. The core is thus appropriate to model the competitive nature of free-form competitive interactions. However, the core may not exist and is not necessarily unique. Non-existence of the core signals the intrinsic instability of a particular interaction. Non-uniqueness of the core implies that competitive forces alone are not enough to determine each agent's value capture and that negotiating abilities also come into play. The core is an attractive concept when building formal models for strategic management, as it focuses on the big picture, providing logically consistent bounds to value capture under competition, without delving into specific tactical moves.

An agent's added value (or marginal contribution) is the value lost to the grand coalition (the coalition of all agents) if the agent is not included. The added-value principle, implied by the core, states that under competitive free-form interaction an agent can capture no more than her added value. Otherwise, the other agents would be better

off to exclude her and trade only among themselves.

The Shapley value is the average added value of an agent to all possible coalitions. It always exists and is unique. It emphasizes the fairness of the division of value among agents. It has been used to model the allocation of costs among users of a common facility and to measure political power. It can also be seen as the expected outcome of a non-cooperative negotiation procedure, for instance to model negotiation in a supply chain (de Fontenay and Gans 2005).

Biform games are a blend of CGT and NCGT. They model the intuitive distinction between shaping the game and playing the game. These comprise two phases. In the first phase, modelled and solved non-cooperatively, agents independently take actions that determine the value they can create as coalitions (i.e., the characteristic function). In the second phase, modelled and solved as a cooperative game using the core, agents create and capture value. Biform games are well suited to model business strategy where decisions are about a firm's ability to create value (e.g., by building capabilities) and to influence the environment to improve value capture, while deemphasizing tactical decisions (e.g., price-setting).

CGT was introduced to strategic management research by Brandenburger and Stuart (1996), who proposed added value as a key analytical concept. Lippman and Rumelt (2003) argued for using CGT in strategy research. MacDonald and Ryall (2004) characterized when an agent is assured to capture value in the core. Brandenburger and Stuart (2007) introduced biform games, used by Stuart (2004, 2005), Adner and Zemsky (2006), and Chatain and Zemsky (2007, 2011) in applied theory. Adegbesan (2009) uses CGT to analyse strategic factor markets. Chatain (2011) applies CGT ideas in a large sample empirical study by relating proxies of added value to firm performance.

See Also

- ▶ Co-opetition
- ▶ Game Theory
- ▶ Value

References

Adegbesan, T. 2009. On the origins of competitive advantage: Strategic factor markets and heterogeneous resource complementarity. *Academy of Management Review* 34: 463–475.

Adner, R., and P. Zemsky. 2006. A demand-based perspective on sustainable competitive advantage. *Strategic Management Journal* 27: 215–239.

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

Brandenburger, A., and H.W. Stuart. 2007. Biform games. *Management Science* 53: 537–549.

Chatain, O. 2011. Value creation, competition and performance in buyer-supplier relationships. *Strategic Management Journal* 32: 76–102.

Chatain, O., and P. Zemsky. 2007. The horizontal scope of the firm: Organizational tradeoffs vs. buyer-supplier relationships. *Management Science* 53: 550–565.

Chatain, O., and P. Zemsky. 2011. Value creation and value capture with frictions. *Strategic Management Journal* 32: 1206–1231.

de Fontenay, C.C., and J.S. Gans. 2005. Vertical integration in the presence of upstream competition. *RAND Journal of Economics* 36: 544–572.

Lippman, S.A., and R.P. Rumelt. 2003. A bargaining perspective on resource advantage. *Strategic Management Journal* 24: 1069–1086.

MacDonald, G., and M.D. Ryall. 2004. How do value creation and competition determine whether a firm appropriates value? *Management Science* 50: 1319–1333.

Osborne, M.J., and A. Rubinstein. 1994. *A course in game theory*. Cambridge, MA: The MIT Press.

Owen, G. 1995. *Game theory*, 3rd ed. San Diego: Academic.

Stuart, H.W. 2004. Efficient spatial competition. *Games and Economic Behavior* 49: 345–362.

Stuart, H.W. 2005. Biform analysis of inventory competition. *Manufacturing & Service Operations Management* 7: 347–359.

Co-opetition

Tomi Laamanen
University of St Gallen, St Gallen, Switzerland

Definition Co-opetition can be defined as simultaneous competition and collaboration between two or more organizations.

The concept of 'co-opetition' was first coined by Ray Noorda, founder of Novell, and popularized in the strategic management field by Adam

M. Brandenburger and Barry J. Nalebuff in their book *Co-opetition* (Brandenburg and Nalebuff 1996), where they used the concept to extend the use of game theory in strategic management and to provide an analysis tool called the 'value net' that incorporated customers, complementors, competitors and suppliers into an analysis of cooperative and competitive dynamics. In strategy research, the concept has become part of the standard vocabulary and a label for situations where firms simultaneously cooperate with and compete against each other. To date, the occurrence and implications of co-opetition have been studied in multiple industries including the steel industry (Gnyawali et al. 2006), breweries (Bengtsson and Kock 2000) and health care (Gee 2000).

Even though co-opetition commonly involves an explicit decision or agreement by a firm to cooperate with its competitor, mutual forbearance can also be considered as a form of implicit co-opetition. When mutual forbearance equilibrium emerges, firms continue to compete with each other, but they also implicitly collude by restraining from engaging in competitive moves that could attract retaliation from their competitor. Gnyawali et al. (2006) apply this logic to cooperative networks. They find that the positioning of firms in relation to others in a network of alliance partners affects the way firms engage in and respond to competitive actions.

So far the strategic management research has had a tendency to view co-opetition predominantly as a one-dimensional construct indicating the tension or balance that ▶ cooperation and competition engender between two or more firms. The fact that firms are complex entities that tend to engage in multiple activities simultaneously, however, makes it possible to develop the concept further. One way to increase the granularity of the concept is to divide it into *reciprocal*, *parallel* and *sequential co-opetition*.

Reciprocal co-opetition refers to co-opetition where a firm collaborates and competes with its competitors in its core product market. Vertically integrated firms that also supply their competitors are a good example of this (e.g., Figueiredo and Teece 1996). Another example is the global airline alliances. In reciprocal co-opetition there is a constant need for monitoring and adjusting how much to invest in the alliance in order to safeguard one's own future. Khanna et al. (1998) model how learning races take place in alliances. Reciprocal co-opetition can also take place in industry consortia aimed at advancing the development of a specific technology domain (e.g., Teece 1996). For example, Motorola, Nokia, Qualcomm and Samsung participate in multiple patent pools where firms cross-license some of their patents to each other. All the consortium members can use the patents in the joint pool, but they have to pay for their use. As the consortium then divides the money earned according to the share of 'essential' patents provided by each firm to the consortium, each firm has to explicitly decide which patents to share and which to keep proprietary. Thus, in addition to competing in the product markets and collaborating in the patent pool, firms also compete for the revenues of the patent pool.

Parallel co-opetition refers to co-opetition that takes place across two firms when they collaborate with each other in some business areas or regions while competing in others. In this case, simultaneous ▶ competition and cooperation can be sustained as they are contained within specific areas of a firm's operations. Probably the best-known example of this is the NUMMI alliance of Toyota and GM to produce cars in their California site while engaging in vigorous competition worldwide. There are also many examples of competing firms joining forces to enter new markets, while continuing to compete simultaneously in other markets. A good example of this is the internet TV alliance of Fox and NBC.

Finally, *sequential co-opetition* refers to co-opetition that takes place across two firms when their competition and collaboration is punctuated over time. For example, two competing software firms may decide to join forces to make an offer to a large client project for which they both alone would be too small. Yet they can continue to compete in any subsequent projects or choose to cooperate depending on the project. While these three subcategories of co-opetition share similarities in terms of balancing the amount of cooperation and competition for the benefit of a firm's overall

performance, they differ in terms of their strategic and organizational implications, stability and the safeguards needed.

See Also

- ▶ Competition
- ▶ Complementarities
- ▶ Cooperation and Competition
- ▶ Cooperative and Non-cooperative Game Theory

References

Bengtsson, M., and S. Kock. 2000. 'Coopetition' in business networks: To cooperate and compete simultaneously. *Industrial Marketing Management* 29: 411–426.

Brandenburg, A.M., and B.J. Nalebuff. 1996. *Co-opetition*. New York: Doubleday.

Figueiredo, J.M.D., and D.J. Teece. 1996. Mitigating procurement hazards in the context of innovation. *Industrial and Corporate Change* 5: 537–559.

Gee, E.P. 2000. Co-opetition: The new market milieu. *Journal of Healthcare Management* 45: 359–363.

Gnyawali, D.R., J.Y. He, and R. Madhavan. 2006. Impact of co-opetition on firm competitive behavior: An empirical examination. *Journal of Management* 32: 507–530.

Khanna, T., R. Gulati, and N. Nohria. 1998. The dynamics of learning alliances: Competition, cooperation, and relative scope. *Strategic Management Journal* 19: 193–210.

Teece, D.J. 1996. Firm organization, industrial structure, and technological innovation. *Journal of Economic Behavior & Organization* 31: 193–224.

Core Business

Irene M. Duhaime[1] and J. L. Stimpert[2]
[1]Georgia State University, J Mack Robinson College of Business, Atlanta, GA, USA
[2]Michigan State University, Eli Broad College of Business, East Lansing, MI, USA

Definition

A "core business" is the business that is the primary focus of a company's business operations. That primary focus may be the business on which the company was founded, or a business to which the company's business operations have evolved over time.

Referring to the *primary* focus of a company's business operations, the concept of core business necessarily suggests that firms will have other businesses that are less core or central. Thus, the concept of core business is closely associated with several streams of strategic management research on multibusiness or multiproduct firms, including research on the relationship between diversification strategy and performance; research on firm focus, divestment, and the performance of divested units; and research on core competence and capabilities.

Diversification or Expansion Beyond the Core Business

Chandler's epic *Strategy and Structure* (1962) describes the way four companies – DuPont, General Motors, Standard Oil and Sears, Roebuck – grew and evolved. Chandler explores how geographical expansion, increasingly broad product lines, diversification and organizational complexity led to an evolution in organizational structure, and specifically, the evolution from a reliance on centralized, functional structures to multidivisional structures.

Building on Chandler's work, Rumelt (1974) develops a typology of diversified business firms, and then investigates how the choice of diversification strategy influences firm performance levels. Rumelt's typology includes four diversification strategies, including single business firms (that have not diversified beyond their core business), dominant strategies, related strategies, and unrelated strategies. Rumelt concludes that firms pursuing a strategy of related diversification, in which firms' businesses are related to their core business, enjoy higher performance than firms pursuing other strategies. A great deal of strategic management research replicates and extends Rumelt's research and findings. This very large body of research suggests few definitive conclusions about the relationship between diversification strategy and firm performance (Bausch and Pils 2009).

Firm Focus, De-diversification and Divestment

The 1970s saw a wave of merger and acquisition activity that included many acquisitions of unrelated businesses and the rise of conglomerate firms. This merger wave in the 1970s was then followed by an emphasis on firm focus and considerable divestment activity. Peters and Waterman's (1982) bestselling book *In Search of Excellence*, profiling a number of companies that had enjoyed a long history of outstanding performance, includes a chapter with the title "Stick to the Knitting". The chapter claims that diversifying away from a core business will distract management attention and dissipate firm resources, leading to lower firm performance. The chapter advises managers and firms to focus on their core business and to avoid acquiring unrelated businesses and far-flung diversification.

The de-diversification wave of the late 1970s and early 1980s prompted researchers to take up the scholarly study of divestment. In early studies of divestment, Duhaime and Grant (1984) and Montgomery and Thomas (1988) sought to understand the factors associated with the divestment of business units. These studies conclude that low affinity or relationship to a firm's core business is a strong predictor of divestment, a finding that is consistently verified in subsequent studies examining the factors influencing divestment decisions and by Markides' (1995) study of corporate refocusing. Other factors associated with divestment include unfavorable industry characteristics and poor business unit performance.

Another line of divestment research examines competing theories about the performance of divested units. One perspective predicts that divested units will face considerable challenges and have unfavourable prospects for success, while a competing view suggests that divested units, once freed from the constraints of operating within a larger, more bureaucratic firm, will enjoy an improvement in performance. Neither perspective has yielded unequivocal support in research studies examining the performance of divested business units.

Core Competence/Core Capabilities

Closely related to the concept of core business are the concepts of core competence (Prahalad and Hamel 1990) and core capabilities (Leonard-Barton 1992). Prahalad and Hamel describe core competence as central or key capabilities that can be employed to enhance the success of other lines of business beyond the core business in multibusiness or multiproduct firms. Rarely based on a single asset or resource, core competences are instead rich combinations of collective learning and capabilities.

Leonard-Barton's work is important because it demonstrates that managers and their firms can become so wedded to existing core competences that they can fail to see or appreciate evolutions in consumer wants and needs and product and process technologies that can render core competences and capabilities obsolete. Thus, core competences and capabilities that provide advantage can become core rigidities that limit adaptability and the development of new competences.

See Also

- ▶ Core Competence
- ▶ Corporate Strategy
- ▶ Diversification
- ▶ Exit Barriers
- ▶ Growth Through Acquisitions
- ▶ Resource Redeployment
- ▶ Strategies for Firm Growth

References

Bausch, A., and F. Pils. 2009. Product diversification strategy and financial performance: Meta-analytic evidence on causality and construct multidimensionality. *Review of Managerial Science* 3: 157–190.

Chandler, A.D. 1962. *Strategy and structure: Chapters in the history of the American industrial enterprise*. Cambridge, MA: The MIT Press.

Duhaime, I.M., and J.H. Grant. 1984. Factors influencing divestment decision-making: Evidence from a field study. *Strategic Management Journal* 5: 301–318.

Leonard-Barton, D. 1992. Core capabilities and core rigidities: A paradox in managing new product development. *Strategic Management Journal* 13: 111–125.
Markides, C.C. 1995. *Diversification, refocusing and economic performance*. Cambridge, MA: The MIT Press.
Montgomery, C.A., and A.R. Thomas. 1988. Divestment: Motives and gains. *Strategic Management Journal* 9: 93–97.
Peters, T.J., and R.H. Waterman Jr. 1982. *In search of excellence: Lessons from America's best-run companies*. New York: Harper & Row.
Prahalad, C.K., and G. Hamel. 1990. The core competence of the corporation. *Harvard Business Review* 68: 79–91.
Rumelt, R.P. 1974. *Strategy, structure, and economic performance*. Boston: Harvard Business School Press.

Core Competence

Douglas Miller[1] and Chih Liu[2]
[1]University of Illinois, College of Business Administration, Champaign, IL, USA
[2]University Of Illinois at Urbana-Champaign, Urbana, USA

Definition A 'core competence' (or competency) is a corporation's learned ability to coordinate technologies and production processes across boundaries in the organization. As defined by C. K. Prahalad and Gary Hamel, a bundle of capabilities can be considered a core competence if it creates value in core products, is applicable across a wide range of markets and is not easily imitated.

As introduced by C. K. Prahalad and Gary Hamel, the term 'core competence' (or competency) refers to a corporation's learned ability to coordinate technologies and production processes across boundaries in the organization. Prahalad and Hamel's 1990 article explains how some ▶ multiproduct companies are able to sustain leadership by creating previously unanticipated products. In their definition, a core competence creates value in core products, is applicable across a wide range of markets and is not easily imitated. As such, core competences are the collective learning that integrates and harmonizes multiple streams of technologies.

Prahalad and Hamel submit that, in the era of intense global competition, price or performance advantage can only provide a company with short-run ▶ competitiveness. In the long run, competitiveness comes from an ability to cost-effectively build the core competences that spawn unanticipated products. Consolidating corporate-wide technologies and production skills into competences enables individual businesses to adapt to rapidly changing conditions. Thus, the identification and retention of existing competences, along with facilitating the acquisition of new competences, becomes the critical task of top management. Unlike physical assets, which deteriorate over time, a core competence does not diminish with use. Instead, competences are enhanced as they are applied and shared, and fade if they are not used. Thus, Prahalad and Hamel call for managers to think of the corporation as a portfolio of competences instead of a portfolio of businesses.

A 'core product' is a component that contributes to the main perceived value of multiple end products. For instance, Honda's main value proposition to customers is a combination of reliability and fuel efficiency. Honda's engines are core products that create this value in cars, motorcycles and lawn mowers. Core products are the physical embodiments of core competence; thus, according to Prahalad and Hamel, companies should seek to maximize their world manufacturing share in core products, not just the market share of end products. Manufacturing for a wide variety of customers provides feedback to the corporation to further enhance its core competence.

The core competence approach to managing the multibusiness firm is similar to the ▶ resource-based view or capabilities view of corporate diversification. Traditional approaches to corporate portfolio management emphasized the evaluation of market opportunities, and then the accumulation of skills and assets to exploit those opportunities. In the 1980s, management scholars began to put greater emphasis on internal analysis of a firm's tangible and intangible resources. Prahalad and Hamel's article was a strong statement about how a large corporation could use all its strengths – not just high R&D spending – to actually create new markets.

In practice, managers and consultants may begin a core competence analysis by identifying core products. Then, they evaluate which technologies and production experience combine to create those core products. An alternative approach is to list all the capabilities of business units, assess their strength and see how they combine. A group of related capabilities is a core competence if it creates value for customers, is superior to the capabilities of competitors, is difficult to imitate and can be applied to multiple markets. A successful corporation will not just have one core competence; however, a list of more than several competences probably includes some that are not truly core. Once core competences are identified, top management must implement an incentive system that treats core competences as corporate resources. For example, career paths should free competence-carrying employees from the confines of individual product lines. Building a strategy around core competences requires communication, involvement and a deep commitment to working across divisional boundaries.

While the idea of core competence has persisted, research published soon after Prahalad and Hamel's article clarified certain points about corporate-level capabilities. First, Robert Grant (1996) (and others) developed a knowledge-based view of the firm, arguing that corporations not only use assets they own, but also technology and information they gather from outside. Second, there are multiple barriers to imitation. Since core competences are complex combinations of technology and experience, they are difficult for competitors to observe and understand, let alone copy. However, less complex resources can also be difficult to imitate, and thus a potential source of competitive advantage, because they are acquired under unique historical conditions, evolve over time through many small decisions, or are embedded in a social structure or corporate culture that is itself costly to imitate (Barney 1991). Third, Dorothy Leonard-Barton (1992) described how a firm's commitment to its core competences can prevent it from recognizing and incorporating new technologies. Thus, a 'core capability' can become a 'core rigidity'. Each of these subsequent insights helps managers see core competences in a more dynamic way.

See Also

- ▶ Competitiveness
- ▶ Management of Technology
- ▶ Resource-Based View

References

Barney, J.B. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17: 99–120.

Grant, R.M. 1996. Prospering in dynamically-competitive environments: Organizational capability as knowledge integration. *Organization Science* 7: 375–387.

Leonard-Barton, D. 1992. Core capabilities and core rigidities: A paradox in managing new product development. *Strategic Management Journal* 13: 111–125.

Prahalad, C.K., and G. Hamel. 1990. The core competence of the corporation. *Harvard Business Review* 68: 79–91.

Corporate Social Responsibility

Alison Mackey[1] and Tyson B. Mackey[2]
[1]University of Utah, Salt Lake City, UT, USA
[2]Clarkson University, New York, NY, USA

Abstract

Corporate social responsibility (CSR) research is the subject of great debate within the management literature. At the heart of the debate is a tension between two opposing views – the neoclassical economies view in which managers must maximize shareholder value, and the ▶ stakeholder view in which managers have responsibilities to society (and to other firm stakeholders) that go beyond simply maximizing shareholder wealth. A large empirical literature has arisen aimed at resolving this tension by demonstrating that CSR can maximize shareholder wealth and therefore no tension exists. These efforts may, ironically,

compromise the unique contribution of CSR in strategic management scholarship, reducing it to simply another value-enhancing business strategy.

Definition Corporate social responsibility is the subset of a firm's responsibilities that are aimed at directly benefiting society. These firm activities are typically voluntary or discretionary (e.g., community involvement, philanthropy, volunteerism).

Corporate social responsibility is the subject of great debate both within practitioner and academic circles. The literature has struggled to clearly define the nature and type of responsibilities that society can legitimately impose on business firms. Many have looked to Carroll's (1979) framework, which posited four core social responsibilities: economic, legal, ethical and discretionary; or to Carroll and Buchholtz (2003), who consolidated the original framework into three levels of social expectations: (1) required (pertaining to economic and legal responsibilities), (2) expected (pertaining to ethical responsibilities) and (3) desired (pertaining to corporate social responsibilities) for guidance in defining corporate social responsibility. However, currently there is no clear consensus as to the definition of corporate social responsibility (Crane et al. 2008).

The central point of tension within this debate rests on the theory and purposes of the firm. Traditional neoclassical economic arguments point to the responsibility managers have to maximize the wealth of the firm's shareholders (Friedman 1963, 1970). To the extent that corporate social responsibility is inconsistent with efforts to maximize the wealth of the firm's shareholders, traditional economic logic suggests that these socially responsible activities should be avoided. Business and society arguments capture the other side of the debate – that managers have responsibilities to society that go above and beyond simply maximizing the wealth of the firm's shareholders. Utilizing a ▶ stakeholder theory of the firm (Freeman 1984), such arguments suggest that the traditional neoclassical economic theory of the firm causes managers to ignore the interests of other important stakeholders (e.g., employees, suppliers, customers, and society at large) and that the interests of these other stakeholders may at times supersede the interests of a firm's shareholders.

Much of the academic literature on corporate social responsibility attempts to resolve this debate by empirically demonstrating that corporate social performance (CSP) (i.e., the actual implementation and practice of corporate social responsibility) is positively associated with corporate financial performance (CFP) – that is, these efforts attempt to resolve the debate by demonstrating that there is no conflict between a firm's corporate social responsibility and the neoclassical theory of the firm. Bragdon and Marlin's (1972) early examination of the correlation between corporate social performance and corporate financial performance launched a multi-decade search for the so-called 'business case' for corporate social responsibility. Empirical results to date have been mixed and inconsistent (Margolis and Walsh 2003) although some meta-analyses point to evidence suggesting a slight positive correlation (see, especially, Orlitzky et al. 2003). Critiques of this empirical literature focus on the well-documented methodological problems within this body of work (e.g., sampling issues, reliability and validity concerns of CSP and CFP measures, lack of casual theory linking CSP to CFP) as well as the more general critiques about the appropriateness of the search for a link between CSP and CFP. To some, the entire effort to demonstrate the business case for corporate social responsibility is problematic because it ignores the essential point made by many business and society scholars, that there are circumstances in which the interests of the firm's shareholders will need to be set aside in favour of the interests of the firm's other stakeholders. In sum, the firm may need to 'do good' even if 'doing good' does not help it to 'do well' financially. Attempting to resolve the debate on the merits of corporate social responsibility purely by focusing on actions that will maximize shareholders' value fails to address the central and unique theoretical question of the corporate social responsibility

literature (Mackey et al. 2007). Further, some note concern that efforts to establish the business case for corporate social responsibility merely reduce the construct to just another form of good management or value-enhancing business strategy (e.g., product differentiation).

While this literature has, to some, evolved into a normative field of study with an activist agenda aimed at increasing the supply of firms engaging in forms of corporate social responsibility, the prominence of this topic and the demands facing corporations today (i.e., corporate social responsibility and even broader expectations of corporative citizenship as emerging business imperatives) cannot be ignored. Surprisingly, the lack of consistency in terminology and conceptualization, combined with the immense proliferation of terminology (e.g., ▶ business ethics, corporate citizenship, corporate social performance, corporate responsibility, stakeholder theory, stakeholder management, sustainability (social and environmental)), has not stopped the CSR literature from growing into a major area of research. It is unclear, though, as the literature matures, whether the problems plaguing this literature (e.g., lack of a common definition or set of core principles, lack of consensus as to whether or not CSR is good for business) will eventually lead to its downfall or whether the literature can continue to evolve and expand despite the struggles related to this concept.

See Also

- ▶ Business Ethics
- ▶ Measurement of Social Value, the
- ▶ Reputation
- ▶ Stakeholder

References

Bragdon Jr., J.H., and J.A.T. Marlin. 1972. Is pollution profitable? *Risk Management* 19: 9–18.

Carroll, A.B.. 1979. A three-dimensional conceptual model of corporate performance. *Academy of Management Journal* 4: 497–505.

Carroll, A.B.., and A.K. Buchholtz. 2003. *Business and society: Ethics and stakeholder management*, 5th ed. Mason: South-Western Cengage Learning.

Crane, A., A. McWilliams, D. Matten, J. Moon, and D. Siegel. 2008. *The Oxford handbook of corporate social responsibility*. Oxford: Oxford University Press.

Freeman, R.E. 1984. *Strategic management: A stakeholder approach*. Boston: Pitman/Ballinger.

Friedman, M. 1963. *Capitalism and freedom*. Chicago: University of Chicago Press.

Friedman, M. 1970. The social responsibility of business is to increase its profits. *New York Times Magazine* 32: 122–126.

Mackey, A., T.B. Mackey, and J.B. Barney. 2007. Corporate social responsibility and firm performance: Investor preferences and corporate strategies. *Academy of Management Review* 32: 817–835.

Margolis, J.D., and J.P. Walsh. 2003. Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly* 48: 268–305.

Orlitzky, M., F.L. Schmidt, and S. Rynes. 2003. Corporate social and financial performance: A meta-analysis. *Organization Studies* 24: 403–411.

Corporate Strategy

Margarethe F. Wiersema[1] and Joshua Hernsberger[2]

[1]University of California Irvine, Paul Merage School of Business, Irvine, CA, USA

[2]University of California, Irvine, CA, USA

Abstract

Corporate strategy refers to the decisions of a firm's top management concerning the scope of the firm, in terms of its geographic and product markets, as well as the degree of ▶ vertical integration. Corporate strategy defines the firm in terms of the extent of its international activities, the degree of vertical integration, and the diversity of the product markets in which it competes. Acquisitions, divestments, foreign direct investments and internal capital investments all constitute decisions that are likely to impact on the strategic scope of the firm and thus its corporate strategy.

Definition Corporate strategy defines the firm in terms of its strategic scope, which includes decisions on the extent of its international activities,

the degree of vertical integration, and the diversity of the product markets in which it competes.

Corporate strategy refers to the decisions of a firm's top management concerning the scope of the firm and the businesses in which the firm competes. The concept of corporate strategy should be distinguished from that of business strategy, or competitive strategy, which refers to the decisions a firm's management makes on how to compete in a given business (Andrews 1980). There are several dimensions to corporate strategy including geographic scope, ▶ vertical integration and product market scope (Chandler 1962).

Corporate strategy includes decisions regarding the firm's vertical scope, or the degree of vertical integration. Vertical integration refers to the degree to which the firm internally produces or controls the steps required to bring a product or service to market. The production of a good can be conceptualized as a series of discrete stages in terms of a value chain that the firm undertakes, from raw materials to the product or service consumed by the end user (Porter 1985). The set of functional activities constitute the steps in a value chain. A firm's corporate strategy in terms of its vertical scope includes decisions on which steps of the value chain to participate in. The more steps in the value chain in which the firm operates, the greater its degree of vertical integration. Thus, corporate strategy entails decisions regarding the degree of the firm's vertical integration.

Corporate strategy also includes decisions regarding the firm's geographic scope or the extent of the firm's international activities. A firm's geographic scope, or its international ▶ diversification, includes decisions as to which foreign markets the firm competes in as well as the geographic location of the steps in the firm's value chain. A firm that operates beyond its domestic borders is considered internationally diversified. Geographic scope includes decisions that encompass all foreign aspects of a firm, from the geographic markets where it sells its products to the global locations where it produces its products and/or where its capabilities reside (Wiersema and Bowen 2011). Thus, corporate strategy entails decisions regarding the extent of the firm's international activities.

A firm's corporate strategy also encompasses decisions regarding the product market scope or the diversity of the product markets in which the firm competes. A firm can compete in either a single business operating in one industry or manage a portfolio of businesses operating in several industries. A firm that competes in more than one product market is considered to be diversified, or a multi-business firm. The extent of a firm's product market diversification is determined by both the number and the relatedness of the businesses within the firm's portfolio. Relatedness refers to whether the businesses in the firm's business portfolio share underlying resources or capabilities. Thus, corporate strategy entails decisions regarding the extent of the firm's diversification in terms of the product markets it competes in and the relatedness among the firm's businesses.

Managerial decisions that influence the vertical, geographic or product market scope of the firm are the essence of corporate strategy. Managers can alter a firm's corporate strategy by expanding the firm's strategic scope through internal investments, foreign direct investments, or the acquisition of companies that add new product or geographic markets, or expand the firm's value chain. For example, Kraft Foods expanded its strategic scope in terms of both product and geographic markets when it acquired Cadbury in 2010. Similarly, managers can reduce the firm's strategic scope, by divesting the firm of businesses and thus reducing its product and/or geographic scope or by reducing its vertical scope through outsourcing. For example, General Electric sold off NBC Universal to Comcast, and thus exited the network broadcasting industry. Decisions such as acquisition and divestiture have an impact on the business portfolio of the firm and thus its corporate strategy.

Theoretical perspectives to explain the determination of a firm's corporate strategy include ▶ transaction cost economics (TCE) and the ▶ resource-based view of the firm (RBV). TCE provides an understanding of the determinants of a firm's strategic scope by conceptualizing the firm and the market as alternative mechanisms

that can be used to coordinate transactions (Coase 1937; Williamson 1985). The set of activities that a firm must undertake in order to produce an end product/service involve a multitude of transactions which can take place either internally (e.g., hierarchical) or externally (e.g., market). TCE maintains that the decision as to whether to rely on markets or hierarchy is based on 'the underlying differences in the attributes of the transactions' (Williamson 1985: 68). Transaction costs can be defined as those of 'planning, adapting, and monitoring task completion under alternative governance structures' (Williamson 1985: 2). For a given transaction the firm incurs either the bureaucratic costs of internal governance or the governance costs of markets. TCE posits that when the bureaucratic costs of internal governance are lower than the governance costs of markets, the tasks concerned with a given business activity will be absorbed into a firm (Coase 1937). According to TCE, managers will organize activities internally and thus expand the firm's strategic scope when the costs are less than those incurred by relying on external markets. Alternatively, managers reduce the firm's strategic scope when the internal bureaucratic costs are higher than the governance costs of the market. Thus, TCE posits that the firm's corporate strategy is determined by decisions that are driven by the desire to minimize the firm's transaction costs.

The determinants of a firm's corporate strategy can also be explained in terms of the resource-based view of the firm. From the RBV perspective, the firm consists of a bundle of resources, and the allocation of these resources determine the strategic scope of the firm (Penrose 1959). RBV theorizes that the firm may have excess resources that cannot be readily sold or traded, or may be more valuable within the firm than on the market, because of the characteristics of the resources (Peteraf 1993). For example, the strength and reputation of a firm's brand is an intangible resource, idiosyncratic to the firm. Consequently, a firm can realize economic benefits by leveraging its brand into other product markets and, in doing so, expand its product market scope. In this way a firm's excess resource capacity provides the determination for expansion in the strategic scope of the firm (Penrose 1959). Thus, RBV theorizes that the firm's corporate strategy is determined by decisions to profitably exploit the firm's bundle of resources.

The impact of corporate strategy on the financial performance of the firm has been examined extensively. It has been theorized that the potential economic benefits from an expansion of a firm's strategic scope arise from the realization of economies of scale and scope of serving more markets and from leveraging the firm's intangible resources into new markets (Caves 1971; Teece 1982). In addition, experiential learning from operating in new markets can enhance the firm's knowledge base and capabilities (Vermeulen and Barkema 2001). With greater strategic scope, the firm can also benefit from increased market power over suppliers, distributors and customers (Kogut 1985). In terms of product market scope, firms that have higher relatedness within their business portfolio have been found to outperform firms with lower levels of relatedness (Robins and Wiersema 1995), while more global firms have been found to outperform their domestic competitors (Hitt et al. 2006). However, both highly product-diversified and internationally diversified firms have been found to have lower financial performance than less diversified firms. Greater strategic scope results in increasing complexity, which in turn results in higher levels of organizational differentiation, necessitating greater collaboration and integration of the interdependencies (Lawrence and Lorsch 1967). As a result, rising managerial and administrative costs begin to outweigh the economic benefits associated with scale and scope economies. Recent evidence indicates that because of increased competitive pressures and globalization, managers have pursued increasingly focused corporate strategies (Comment and Jarrell 1995; Bowen and Wiersema 2005), which in turn has resulted in significant improvements in firm performance (Berger and Ofek 1995).

Decisions regarding the corporate scope of the firm determine corporate strategy. Thus, corporate strategy is defined by the extent of product diversity, internationalization and the degree of vertical integration pursued by the firm.

See Also

- ▶ Diversification
- ▶ Resource-Based View
- ▶ Transaction Cost Economics
- ▶ Vertical Integration

References

Andrews, K.R. 1980. *The concept of corporate strategy*. Homewood: Irwin.

Berger, P.G., and E. Ofek. 1995. Diversification's effect on firm value. *Journal of Financial Economics* 37: 39–65.

Bowen, H.P., and M.F. Wiersema. 2005. Foreign-based competition and corporate diversification. *Strategic Management Journal* 26: 1153–1171.

Caves, R.E. 1971. International corporations: The industrial economics of foreign investment. *Economica* 38: 1–27.

Chandler, A. 1962. *Strategy and structure*. Cambridge, MA: The MIT Press.

Coase, R.H. 1937. The nature of the firm. *Economica* 4: 386–405.

Comment, R., and G.A. Jarrell. 1995. Corporate focus and stock returns. *Journal of Financial Economics* 37: 67–87.

Hitt, M.A., L. Tihanyi, T. Miller, and B. Connelly. 2006. International diversification: Antecedents, outcomes, and moderators. *Journal of Management* 32: 831–867.

Kogut, B. 1985. Designing global strategies: Comparative and competitive value added chains. *Sloan Management Review* 27: 15–28.

Lawrence, P.R., and J.W. Lorsch. 1967. Differentiation and integration in complex organizations. *Administrative Science Quarterly* 12: 1–47.

Penrose, E. 1959. *The theory of the growth of the firm*. New York: Wiley.

Peteraf, M.A. 1993. The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal* 14: 179–192.

Porter, M.E. 1985. *Competitive advantage: Creating and sustaining superior performance*. New York: Free Press.

Robins, J., and M.F. Wiersema. 1995. A resource-based approach to the multi-business firm: Empirical analysis of portfolio interrelationships and corporate financial performance. *Strategic Management Journal* 16: 277–299.

Teece, D.J. 1982. Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior & Organization* 3: 39–63.

Vermeulen, G.A.M., and H.G. Barkema. 2001. Learning through acquisitions. *Academy of Management Journal* 44: 457–476.

Wiersema, M.F., and H.P. Bowen. 2011. The relationship between international diversification and firm performance: Why it remains a puzzle. *Global Strategy Journal* 1: 152–170.

Williamson, O.E. 1985. *The economic institutions of capitalism*. New York: Free Press.

Corporate Venturing

Gary Dushnitsky
London Business School, London, UK

Abstract

Firms' survival and growth critically depend on their ability to innovate. In addition to internal R&D, firms are increasingly pursuing (open) ▶ innovation through engagement with external partners. Corporate venturing emerges as an integral part of this strategy as firms harness innovative entrepreneurial ventures. Due to the unique nature of the phenomenon, the corporate venturing literature is informed by, and contributes to, the fields of entrepreneurship, finance and strategy. We provide a definition of corporate venturing and brief examination of its benefits, uses and implications to scholarly work.

Definition Corporate venturing, also known as corporate venture capital, is defined as a minority equity investment by an established corporation in a privately held entrepreneurial venture.

It is an external venturing strategy that many firms undertake as part of their inter-organizational relationships (e.g., strategic alliances, licensing or joint ventures). On the one hand, corporate venturing and strategic alliance are similar: they serve as mechanisms for two independent firms to exchange resources. On the other hand, they differ with respect to the nature of the relationship and its organization (Dushnitsky and Lavie 2010). Alliances imply mutual dependence of otherwise independent firms that engage in interactive coordination of various value chain activities such as joint R&D and marketing initiatives. In contrast,

CVC investment entails disparity between an investor and the consumer of monetary funds, and involves a unidirectional flow of financial resources from the investor to the funded venture that independently performs its value chain activities. Alliances have specific objectives that are negotiated and then pursued by both parties whereas CVC agreements pertain to the operations of the funded venture. In contrast, the scope of alliance operations is narrowly defined even when involving an equity stake position (Robinson and Stuart 2007). Moreover, many firms manage alliances and CVC through separate units aimed either at alliance management (Dyer et al. 2001) or venture capital investment (Keil 2002; Dushnitsky 2006). This organizational divide reflects managers' views of alliances and CVC as distinct activities, and is also manifested in distinctive staffing practices and personnel backgrounds (Block and Ornati 1987; Dushnitsky and Shapira 2010).

The corporate venturing literature is related to, yet distinct from, the corporate entrepreneurship and spin-out literatures. Consider the corporate entrepreneurship, also known as ▶ intrapreneurship, literature. This refers to a wide array of internally oriented activities, including investment in internal divisions, business development funds and so on (e.g., Guth and Ginsberg 1990; Zahra 1995; Thornhill and Amit 2001). The distinction between the concepts is twofold: (a) a CVC unit targets external entrepreneurs, while corporate employees are funded by intrapreneurship initiatives, and (b) a CVC investor and an entrepreneurial venture participate in the market for entrepreneurial financing, along with independent VCs and angel investors, whereas intrapreneurship entails the parent corporation backing business and employee initiatives that do not consider competing sources of capital. Next, consider the spin-outs, also known as spin-off, literature (e.g., Klepper 2001; Agarwal et al. 2004; Gompers et al. 2004). The direction of employee mobility differentiates the two: spin-out describes an outflow situation whereby an employee departs a corporation and starts his or her own business, whereas CVC is interested with inflows, namely, harnessing entrepreneurial knowledge or products.

Extant Work

Corporate venturing involves a parent corporation that launches a CVC programme, which in turn invests in entrepreneurial ventures. Extant work focuses on each of these entities, as well as the nature of the interactions between them, and the ultimate performance implications.

Why Does an Established Firm Choose to Invest in Entrepreneurial Ventures?

The objectives, or mandate, of CVC programmes vary. Independent – not corporate-affiliated– venture capital funds seek financial returns. An early survey finds that financial returns are the main objective for most programmes, yet a large minority emphasizes the pursuit of strategic objectives (Siegel et al. 1988). The latter covers such objectives as windows onto novel technologies, exposure to new products or geographical markets, or opportunity to expand a firm's network (Siegel et al. 1988; Sykes 1990; McNally 1997). More recent surveys (e.g., Ernst and Young 2002, 2009) find that most firms explicitly balance strategic and financial objectives, with the fraction of solely financially orientated CVC programmes on the decline.

Large sample analysis of actual CVC investment patterns substantiate the strategic role of CVC activity (Dushnitsky and Lenox 2003, 2005b; Basu et al. 2010; Sahaym et al. 2010): investing firms experience greater technological and competitive pressures to innovate and to target ventures that are likely to possess cutting-edge technologies, and have capabilities to absorb these technologies or integrate them into the corporate ecosystem.

Combined, these studies suggest that strategic objectives can range from seeking substitutes to sponsoring complements. On the one hand, investment activity may be used to identify novel products, services or technologies to replace existing corporate offerings (i.e., targeting potential substitutes). On the other hand, CVC activities

may complement corporate businesses by funding ventures that increase the value of existing lines of businesses (Brandenburger and Nalebuff 1996): nurturing an ecosystem by seeding ventures that offer complementary products or services (Riyanto and Schwienbacher 2006; Gawer and Henderson 2007).

How Is the CVC Programme Governed?

The relationship between a parent corporation and its CVC programme incorporates the organizational structure of the CVC programme, its autonomy level and compensation scheme. One observes four common CVC structures (Rind 1981; Siegel et al. 1988; Winters and Murfin 1988; Sykes 1990; McNally 1997; Birkinshaw et al. 2002; Ernst and Young 2009): (a) the programme is managed by a corporate business unit, (b) it is managed by a separate, specifically formed, subsidiary, (c) the programme co-managed investments along with a venture capital fund or (d) capital is allocated to venture capital funds that invest on behalf of the corporation.

The diversity in programmes' governance extends to their level of autonomy, both in terms of capital allocation and decision autonomy (Rind 1981; Siegel et al. 1988; Winters and Murfin 1988; Sykes 1990; McNally 1997; Birkinshaw et al. 2002; Keil 2002; Keil et al. 2008). Some programmes are allocated a large amount of capital upfront while others receive the necessary funds on an ad hoc basis. The discretion to make investment (i.e., fund a particular venture) and exit (i.e., sell a venture or take it public) decisions is fully delegated to the CVC programme in some corporations yet remains subject to scrutiny and corporate approval in others. Compensation practices are equally heterogeneous (Block and Ornati 1987; Sykes 1992; Dushnitsky and Shapira 2010). A minority of programmes award high-powered incentives to their personnel. Flat-rate corporate salary was the prevailing compensation scheme among CVC personnel in the past and remains a common practice today.

The interdependencies between the three governance facets have received little attention in the literature. Yet preliminary evidence suggests that firms align these facets to support benefits capture (Keil 2002; Keil et al. 2008), and that deviation from ideal structure types may result in inferior performance (Hill and Birkinshaw 2008).

Who Receives CVC Funding?

The relationship between CVC programmes and entrepreneurial ventures goes beyond financial investment. An important part of the relationships unfolds prior to the funding round itself. Potential investment targets are usually identified through referrals from venture capitalists (Hochberg et al. 2007; Dushnitsky and Shaver 2009; Hill et al. 2009; Keil et al. 2010), and at times originate from employees and business partners (Winters and Murfin 1988; Sykes 1990; Ernst and Young 2009). A key point is that the entrepreneur plays a proactive role in the process (Dushnitsky 2006) such that the formation of a CVC–entrepreneur investment is a mutual decision (Dushnitsky and Shaver 2009). Specifically, corporations are more likely to fund ventures with significant resource needs or those with a high level of complementarities. Moreover, investment between those with potentially substituting products is sensitive to intellectual property regime and CVC organizational structure.

Once an investment is consummated, the programme actively monitors its portfolio companies. CVC programmes communicate with ventures frequently and many programmes hold a board seat, or at least observer rights (Sykes 1990; Bottazzi et al. 2004; Cumming 2006, 2008; Masulis and Nahata 2011). One also observes bidirectional flow of information, capital and commercial assets between the corporation and the ventures (e.g., Siegel et al. 1988; McNally 1997; Birkinshaw et al. 2002; Maula et al. 2009; Yang et al. 2009; Bengtsson and Wang 2010). The entrepreneurs benefit from corporate advice, and at times the ventures educate the parent corporation about new technologies or business opportunities. Interestingly, there is evidence that the level of complementarities of a CVC–entrepreneur pair shapes the intensity of monitoring and other interactions between the two.

It is noteworthy that the CVC–entrepreneur relationships may affect other firm activities

(Arora and Gambardella 1990; Keil et al. 2008; Van de Vrande et al. 2009). Specifically, there is evidence of interdependencies between corporate venture capital and internal R&D (Chesbrough and Tucci 2004; Dushnitsky and Lenox 2005b), strategic alliances (e.g., McNally 1997; Colombo et al. 2006; Phelps and Wadhwa 2009; Dushnitsky and Lavie 2010) or M&A activity (Benson and Ziedonis 2009, 2010; Tong and Li 2011).

What Are the Performance Implications of Corporate Venture Capital?

Because an entrepreneurial venture, or a CVC programme, may benefit at the expense of the parent firm (or vice versa), we discuss each entity separately. The financial returns to the parent corporation are highly skewed. Many firms experience negative returns from their CVC activity, while a few record high positive financial performance (Allen and Hevert 2007). The extent to which a CVC-investing firm outperforms its industry peers is associated with the objective of the CVC programme: it mainly occurs for strategically orientated CVC (Dushnitsky and Lenox 2006). Indeed, a CVC-investing firm exhibits exploratory patenting behaviour (Schildt et al. 2005) as well as greater patenting output in comparison to industry peers (Dushnitsky and Lenox 2005a; Wadhwa and Kotha 2006). The magnitude of these benefits is a function of the parent firm's absorptive capacity as well as industry-level factors.

The performance of the CVC programmes echoes that of the parent firm. Programmes that perform well strategically also report favourable financial returns (Siegel et al. 1988; McNally 1997). The degree to which firms experience favourable performance varies across governance structures: it is sensitive to programme autonomy level (Gompers and Lerner 1998; Hill and Birkinshaw 2008; Hill et al. 2009) and compensation schemes (Sykes 1990; Dushnitsky and Shapira 2010).

Finally, CVC-backed ventures exhibit, on average, favourable performance in absolute terms (McNally 1997; Maula et al. 2009) as well as in comparison with VC-backed ventures (Gompers and Lerner 1998; Maula and Murray 2001; Ivanov and Xie 2010). Evidence regarding the mechanisms through which these benefits accrue remain anecdotal and are attributed to: (a) the CVCs' competence in picking winners and (b) ventures' ability to leverage corporate resources. It is also possible that a venture's benefits come at the expense of its corporate investor, for instance, if the latter offered inflated valuations.

Final Thoughts

Corporate venturing is attracting the attention of scholars and practitioners alike. A number of comprehensive reviews (Dushnitsky 2006; Maula 2007) outline the substantial knowledge accumulated on the topic and highlight opportunities for future work. Some of the main future trends include the expansion of CVC practices and investment beyond Silicon Valley and the need to manage CVC activity as programmes' longevity increases. Addressing these issues necessitates multidisciplinary effort across the fields of entrepreneurship, finance and strategy.

See Also

- ▶ Innovation
- ▶ Intrapreneurship
- ▶ Open Innovation

References

Agarwal, R., R. Echambadi, A. Franco, and M.B. Sarkar. 2004. Knowledge transfer through inheritance: Spin-out generation, growth and survival. *Academy of Management Journal* 47: 501–522.

Allen, S.A., and K.T. Hevert. 2007. Venture capital investing by information technology companies: Did it pay? *Journal of Business Venturing* 22: 262–282.

Arora, A., and A. Gambardella. 1990. Complementarity and external linkages: The strategies of the large firms in biotechnology. *Journal of Industrial Economics* 38: 361–379.

Basu, S., C. Phelps, and S. Kotha. 2010. Towards understanding who makes corporate venture capital investments and why. *Journal of Business Venturing* 26: 153–171.

Bengtsson, O., and F. Wang. 2010. What matters in venture capital: Evidence from entrepreneurs' stated preferences. *Financial Management* 39: 1367–1401.

Benson, D., and R. Ziedonis. 2009. Corporate venture capital as a window on new technologies: Implications for the performance of corporate investors when acquiring startups. *Organization Science* 20: 329–351.

Benson, D., and R. Ziedonis. 2010. Corporate venture capital and the returns to acquiring portfolio companies. *Journal of Financial Economics* 98: 478–499.

Birkinshaw, J., G. Murray, and R. van Basten-Batenburg. 2002. *Corporate venturing: The state of the art and the prospects for the future*. London: London Business School.

Block, Z., and O.A. Ornati. 1987. Compensating corporate venture managers. *Journal of Business Venturing* 2: 41–52.

Bottazzi, L., M. Da Rin, and T. Hellmann. 2004. The changing face of the European venture capital industry: Facts and analysis. *Journal of Private Equity* 7: 26–53.

Bottazzi, L., M. Da Rin, and T. Hellmann. 2008. Who are the active investors? Evidence from venture capital. *Journal of Financial Economics* 89: 488–512.

Brandenburger, A., and B. Nalebuff. 1996. *Co-opetition*. Cambridge, MA: Harvard Business Press.

Chesbrough, H. W. and Tucci, C. 2004. Corporate venture capital in the context of corporate innovation. Paper presented at the DRUID conference in Copenhagen, Denmark.

Colombo, M., L. Grilli, and E. Piva. 2006. In search for complementary assets: The determinants of alliance formation of high-tech startups. *Research Policy* 35: 1166–1199.

Cumming, D. 2006. Corporate venture capital contracts. *Journal of Alternative Investments* 9: 40–53.

Dushnitsky, G. 2006. Corporation venture capital: Past evidence and future directions. In *Oxford handbook of entrepreneurship*, ed. M. Casson, B. Yeung, A. Basu, and N. Wadeson. Oxford: Oxford University Press.

Dushnitsky, G., and D. Lavie. 2010. How alliance formation shapes corporate venture capital investment in the software industry: A resource-based perspective. *Strategic Entrepreneurship Journal* 4: 22–48.

Dushnitsky, G. and M. J. Lenox. 2003. When do firms undertake R&D by investing in new ventures? Best Paper Proceedings of the Annual Meeting of the Academy of Management.

Dushnitsky, G., and M.J. Lenox. 2005a. When do incumbents learn from entrepreneurial ventures? *Research Policy* 34: 615–639.

Dushnitsky, G., and M.J. Lenox. 2005b. When do firms undertake R&D by investing in new ventures? *Strategic Management Journal* 26: 947–965.

Dushnitsky, G., and M.J. Lenox. 2006. When does corporate venture capital investment create firm value? *Journal of Business Venturing* 21: 753–772.

Dushnitsky, G., and Z.B. Shapira. 2010. Entrepreneurial finance meets corporate reality: Comparing investment practices and performing of corporate and independent venture capitalists. *Strategic Management Journal* 30: 1025–1132.

Dushnitsky, G., and J.M. Shaver. 2009. Limitations to interorganizational knowledge acquisitions: The paradox of corporate venture capital. *Strategic Management Journal* 30: 1045–1064.

Dyer, J.H., P. Kale, and H. Singh. 2001. How to make strategic alliances work. *Sloan Management Review* 42: 37–43.

Ernst and Young. 2002. Corporate venture capital report.

Ernst and Young. 2009. Global corporate venture capital report.

Gawer, A., and R. Henderson. 2007. Platform owner entry and innovation in complementary markets: Evidence from Intel. *Journal of Economics & Management Strategy* 16: 1–34.

Gompers, P. and Lerner, J. 1998. *The determinants of corporate venture capital success: Organizational structure, incentives and complementarities*. NBER Working paper No. 6725.

Gompers, P., J. Lerner, and D. Scharfstein. 2004. Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986–1999. *Journal of Finance* 60: 577–614.

Guth, W., and A. Ginsberg. 1990. Guest editor's introduction: Corporate entrepreneurship. *Strategic Management Journal* 11: 5–15.

Hill, S., and J. Birkinshaw. 2008. Strategy–organization configurations in corporate venture units: Impact on performance and survival. *Journal of Business Venturing* 23: 423–444.

Hill, S., M. Maula, J. Birkinshaw, and G. Murray. 2009. Transferability of the venture capital model to the corporate context: Implications of corporate venture units. *Strategic Entrepreneurship Journal* 3: 3–270.

Hochberg, Y., A. Ljungqvist, and Y. Lu. 2007. Whom you know matters: Venture capital networks and investment performance. *Journal of Finance* 62: 251–301.

Ivanov, V., and F. Xie. 2010. Do corporate venture capitalists add value to start-up firms? Evidence from IPOs and acquisitions of VC-backed companies. *Financial Management* 39: 129–152.

Keil, T. 2002. *External corporate venturing: Strategic renewal in rapidly changing industries*. Westport: Quorum Books.

Keil, T., E. Autio, and G. George. 2008. Corporate venture capital, disembodied experimentation and capability development. *Journal of Management Studies* 45: 1475–1505.

Keil, T., M. Maula, and C. Wilson. 2010. Unique resources of corporate venture capitalists as a key to entry into rigid venture capital syndication networks. *Entrepreneurship: Theory and Practice* 34: 83–103.

Klepper, S. 2001. Employee startups in high-tech industries. *Industrial and Corporate Change* 10: 639–674.

Masulis, R., and R. Nahata. 2011. Venture capital conflicts of interest: Evidence from acquisitions of venture-backed firms. *Journal of Financial and Quantitative Analysis* 46: 395–430.

Maula, M. 2007. Corporate venture capital as a strategic tool for corporations. In *Handbook of research on venture capital*, ed. H. Landström. Cheltenham: Edward Elgar Publishing.
Maula, M., and G. Murray. 2001. Corporate venture capital and the creation of US public companies. In *Creating value: Winners in the new business environment*, ed. A. Hitt and N. Lucier. Oxford: Blackwell.
Maula, M., E. Autio, and G. Murray. 2009. Corporate venture capital and the balance of risks and rewards for portfolio companies. *Journal of Business Venturing* 24: 274–286.
McNally, K. 1997. *Corporate venture capital: Bridging the equity gap in the small business sector*. London: Routledge.
Phelps, C., and Wadhwa, A. 2009. *An option to partner: A dyadic analysis of CVC relationships*. Best Paper Proceedings of the Annual Meeting of the Academy of Management.
Rind, K.W. 1981. The role of venture capital in corporate development. *Strategic Management Journal* 2: 169–180.
Riyanto, Y.E., and A. Schwienbacher. 2006. The strategic use of corporate venture financing for securing demand. *Journal of Banking and Finance* 30: 2809–2833.
Robinson, D.T., and T.E. Stuart. 2007. Financial contracting in biotech strategic alliances. *Journal of Law and Economics* 50: 559–595.
Sahaym, A., H.K. Steensma, and J.Q. Barde. 2010. The influence of R&D investment on the use of corporate venture capital: An industry-level analysis. *Journal of Business Venturing* 25: 376–388.
Schildt, H.A., M.V.J. Maula, and T. Keil. 2005. Explorative and exploitative learning from external corporate ventures. *Entrepreneurship Theory & Practice* 29: 493–515.
Siegel, R., E. Siegel, and I. MacMillan. 1988. Corporate venture capitalists: Autonomy, obstacles and performance. *Journal of Business Venturing* 3: 233–247.
Sykes, H. 1990. Corporate venture capital: Strategies for success. *Journal of Business Venturing* 5: 37–47.
Sykes, H. 1992. Incentive compensation for corporate venture personnel. *Journal of Business Venturing* 7: 253–265.
Thornhill, S., and R. Amit. 2001. A dynamic perspective of internal fit in corporate venturing. *Journal of Business Venturing* 16: 25–50.
Tong, T.W., and Y. Li. 2011. Real options and investment mode: Evidence from corporate venture capital and acquisition. *Organization Science* 22: 659–674.
Van de Vrande, V., W. Vanhaverbeke, and G. Duijsters. 2009. External technology sourcing: The effect of uncertainty on governance mode choice. *Journal of Business Venturing* 24: 62–80.
Wadhwa, A., and S.B. Kotha. 2006. Knowledge creation through external venturing: Evidence from the telecommunications equipment manufacturing industry. *Academy of Management Journal* 49: 819–835.
Winters, T.E., and D.L. Murfin. 1988. Venture capital investing for corporate development objectives. *Journal of Business Venturing* 3: 207–222.
Yang, Y., V.K. Narayanan, and S. Zahra. 2009. Developing the selection and valuation capabilities through learning: The case of corporate venture capital. *Journal of Business Venturing* 24: 261–273.
Zahra, S. 1995. Corporate entrepreneurship and financial performance: The case of management leveraged buy-outs. *Journal of Business Venturing* 10: 225–247.

Co-specialization

David J. Teece
Berkeley Research Group, LLC, Emeryville, CA, USA
Haas School of Business, University of California, Berkeley, Berkeley, CA, USA

Definition Co-specialization exists when the value generated by two or more assets used in combination is substantially greater than the value of each asset in its next best use.

Assets are co-specialized when the value of each asset is a positive function of its use in conjunction with the other assets. Furthermore, the value of each asset in alternative uses must be substantially less for deep co-specialization to exist.

An asset is specialized when it cannot be put to alternative use without a substantial loss in value. In the classic mine-mouth coal-fired electric power facility (Joskow 1985), once the electricity-generating facility is built at the mine mouth, there is a contractual hazard associated with obtaining coal from the mine. Long-term contracts entered into may or may not suffice to provide adequate protection against the mine owners arbitrarily raising the price of coal to the dependent generating facility. The mine owners, on the other hand, may be able to sell coal in global markets, which limits the loss they would incur from not selling to the generating facility. If, however, the coal mine cannot easily dispose of its coal elsewhere (e.g., because of a lack of transport infrastructure), then the two facilities are co-specialized.

With co-specialization, new value can be created (and potentially appropriated) when owners of assets are made aware of their value in

combination with other idiosyncratic assets. The computer, software and electronics industries are particularly rich in co-specialization requirements and opportunities. Cross-border settings also frequently present instances of co-specialization, which is one of the factors behind foreign direct investment (Pitelis and Teece 2010).

Specialized assets may be undervalued in isolation because they are difficult to sell at full value to anyone but an owner of the necessary complement(s). If there are few such buyers, then prices can be low. The lack of active markets for these assets makes them, in turn, hard for competitors to obtain, even when the competitor has a suitable complementary asset. Thus, co-specialized assets are a potential basis for sustainable competitive advantage. Because they can yield special value, they are sometimes said to be strategic assets.

Protecting co-specialization benefits frequently requires integrated (vertical or horizontal) operations. Asset specificity is in fact a key concept in ▶ transaction cost economics (Williamson 1985).

An organization's ability to identify, develop and utilize specialized and co-specialized assets built or bought is at the heart of the ▶ dynamic capabilities approach (Augier and Teece 2006). The ability of management to effectuate the coordination of co-specialized assets through own development, design or astute purchase is an essential (dynamic) capability necessary to seize new opportunities and manage threats.

The strategic management implications of co-specialization and specialized assets can be found in the pre-capabilities literature, most notably in Teece (1986). The Porter ▶ five forces framework approach ignored these ▶ complementarities, choosing instead to highlight substitutes.

Decisions on whether to build, buy or ally other specialized assets will depend upon many factors. In addition to traditional transaction costs (risk of opportunistic behaviour), Teece (1986, 2006) identifies (1) relative positioning of other asset owners, (2) cash availability and (3) whether the asset (and any associated operational capabilities) can be built in time.

See Also

- ▶ Complementarities
- ▶ Dynamic Capabilities
- ▶ Five Forces Framework
- ▶ Scope Economies
- ▶ Transaction Cost Economics

References

Augier, M., and D.J. Teece. 2006. Understanding complex organization: The role of know-how, internal structure, and human behavior in the evolution of capabilities. *Industrial and Corporate Change* 15: 395–416.

Joskow, P. 1985. Vertical integration and long term contracts: The case of coal-burning electric generating plants. *Journal of Law, Economics & Organization* 1: 33–80.

Pitelis, C.N., and D.J. Teece. 2010. Cross-border market co-creation, dynamic capabilities and the entrepreneurial theory of the multinational enterprise. *Industrial and Corporate Change* 19: 1247–1270.

Teece, D.J. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy* 15: 285–305.

Teece, D.J. 2006. Reflections on profiting from innovation. *Research Policy* 35: 1131–1146.

Williamson, O.E. 1985. *The economic institutions of capitalism*. New York: Free Press.

Cost

Marvin Lieberman
University of California Los Angeles, Anderson School of Management, Los Angeles, CA, USA

Abstract

This entry begins by highlighting the different concepts of cost which are discussed in the literature. Following this short section, the entry discusses the accounting concept of cost, which distinguishes between fixed and variable costs. The entry then considers economic concepts, including opportunity costs, marginal cost and sunk costs. The final paragraph considers some of the economic aspects that are more difficult to measure, such as externalities.

Definition 'Cost' is the amount of money that must be expended to achieve specific objectives. Although seemingly a straightforward concept, various notions of 'cost' have been developed to serve different purposes.

'Cost' is the amount of money that must be expended to achieve specific objectives (in a business context, typically the production or acquisition of goods or services). Although seemingly a straightforward concept, various notions of 'cost' have been developed to serve different purposes. Economic concepts of 'cost' deviate from standard accounting measures, and economic characterizations can depend on the time frame of analysis. Furthermore, economic concepts such as 'opportunity costs' and 'transaction cost', intended as guides for managerial decision-making, are distinct from actual monetary expenditures. These conceptual distinctions can have important managerial and strategic implications. Some common notions of 'cost' are described briefly below and in greater detail elsewhere in this volume.

Accounting Measures of Cost

Accounting costs capture the recorded value of monetary expenditures made by the firm. There are three basic categories of accounting costs: materials costs, (direct) labour costs and overhead costs. The latter includes subcategories such as indirect labour (including managerial labour), depreciation, rent, utilities and so on. Details of cost accounting can be found in textbooks such as Horngren and colleagues (2011).

A major purpose of cost accounting is to develop measures of fixed and variable cost per unit of output, to provide information for decision-making. Unit cost calculations are often sensitive to how overhead costs are determined and allocated, given that overheads account for the bulk of total costs in many companies today. Moreover, the extent to which overheads are fixed or variable depends on the time frame of analysis. Thus, accounting estimates of unit cost are always arbitrary to some degree.

Accounting measures give little or no information about notions of cost that are not directly linked to the firm's transactions, such as ▶ sunk costs, opportunity costs and transaction costs. Such dimensions of cost often have the most important managerial and strategic implications.

C

Economic Concepts of Cost

Economists have developed numerous concepts of cost that differ from the accounting measures. The primary economic concept is that of 'opportunity cost': the value of a resource in its next best use. Opportunity cost differs from out-of-pocket cost in that it values a resource on the basis of its best alternative application (obtainable by reallocating the resource within the firm, or by selling it outside). This value may be higher or lower than the amount of money originally expended to acquire the resource. Although not directly observable, opportunity cost is the notion that economists consider most relevant for managerial decisions pertaining to resource allocation.

The classical theory of the firm described in economic textbooks is rooted in concepts of fixed, variable and ▶ marginal cost. Fixed costs are those that do not vary with the volume of output over the time frame of analysis. (A typical example is the cost of renting equipment and facilities, which is set at the start of each time period and thus does not depend on the volume of output produced.) Variable costs are those that increase with the volume of output (e.g., raw materials, temporary workers). The firm's total costs are the sum of its fixed and variable costs. Marginal cost is the change in total cost associated with one additional unit of output or, equivalently, the variable cost of producing one more unit. The classical economic theory of the firm implies that decisions relating to output and pricing should be based on marginal cost.

Modern perspectives emphasize the idea that some types of fixed costs are also 'sunk'. Sunk costs are expenditures that cannot be recouped if the firm exits from the business. Advertising and R&D are important categories of sunk costs, and expenditures on physical facilities often have

a sunk cost component. (Note that rent is not sunk, unless a continued stream of rental payment is contractually required. Similarly, the purchase cost of a building or equipment is not sunk if the assets can be sold or used for other purposes without loss. However, if assets take a specialized form that reduces the value that can be obtained in best alternative use or outside sale, the reduction in value from the original purchase price represents a sunk cost.)

Sunk costs are important for several reasons. Investments made by a firm can shift the firm's cost structure, potentially reducing the firm's marginal cost. When these investments take the form of sunk costs and are made ahead of competitors, they can serve as a means for strategic commitment and market pre-emption (Ghemawat 1991). Similar logic reveals that industries with high sunk costs tend to have high producer concentration (Sutton 1991).

Sunk costs also have important managerial implications. Economic logic implies that sunk costs, once incurred, should be rationally ignored in making subsequent decisions (which should be based solely on forward-looking marginal or opportunity costs). Nevertheless, sunk costs are often misperceived by managers in ways that distort decision-making. Many managers and individuals treat sunk costs as if they are continuing costs that need to be recouped in the future, a phenomenon known as the 'sunk cost fallacy'. One explanation for such behaviour is 'loss aversion', a common cognitive bias identified by psychologists and behavioural economists (Kahneman 2011).

Transactions cost theory is based on concepts of sunk costs that are specific to a particular business transaction or set of partners. 'Transactions costs' arise when firms make relation- or partner-specific investments in order to pursue economic transactions.

In many situations, the costs incurred by a firm do not include all of the economic costs paid by society. For example, production of a product or service may generate air or water pollution; this imposes costs on society that may not be fully charged to the firm. In such cases, the 'social costs' of the good or service can exceed the private costs because of external diseconomies or 'externalities'.

See Also

- ▶ Externality
- ▶ Marginal Cost
- ▶ Market Structure
- ▶ Measurement of Social Value, the
- ▶ Sunk Costs
- ▶ Transaction Cost Economics

References

Ghemawat, P. 1991. *Commitment: The dynamic of strategy*. New York: Free Press.

Horngren, C.T., S.M. Datar, and M.V. Rajan. 2011. *Cost accounting: A managerial emphasis*, 14th ed. Upper Saddle River: Prentice Hall.

Kahneman, D. 2011. *Thinking fast and slow*. New York: Farrar, Straus & Giroux.

Sutton, J. 1991. *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration*. Cambridge, MA: MIT Press.

Cost Leadership

Marvin Lieberman
University of California Los Angeles, Anderson School of Management, Los Angeles, CA, USA

Abstract

Also known as cost advantage, cost leadership is one of the generic strategies proposed by economist Michael Porter. Cost leadership requires aggressive pursuit of cost controls and requires that cost minimization be central to managerial attention and corporate culture. The cost leader is likely to earn attractive returns, as it can remain profitable even if strong industry rivalry drives competitors' margins to zero. Depending on the industry context, the cost leader may set prices below those of competitors and offer a more basic product line.

Definition Cost leadership is the strategy of attaining the position of lowest cost among competitors in an industry.

The firm with lowest ▶ cost of operation in a given industry or market holds the position of cost leadership. Also known as cost advantage, cost leadership is one of the generic strategies proposed by economist ▶ Michael Porter (1980, 1985). Although many firms in an industry can simultaneously hold positions of differentiation advantage (the other main generic strategy proposed by Porter), only a single firm can be the cost leader. The cost leader is likely to earn attractive returns, as it can remain profitable even if strong industry rivalry drives competitors' margins to zero.

A cost leadership strategy requires aggressive pursuit of cost controls and requires that cost minimization be central to managerial attention and corporate culture. Porter (1996) argues that a successful cost leader must align all its policies and internal activities to complement and reinforce each other in the direction of minimizing costs. Firms such as WalMart, IKEA and Southwest Airlines have historically achieved such alignment, as illustrated by Porter (1996).

Porter describes several techniques that help to attain cost leadership. One is to exploit 'cost drivers' such as economies of scale, scope and accumulated experience. By pursuing high volume, the firm may be able to achieve high asset turnover and move rapidly down the learning curve. One example of this approach is Intel, which has maintained cost leadership in the microprocessor industry through scale economies and aggressive cost reductions sustained via large rates of output, compared with its closest rival, AMD.

Another technique proposed by Porter (1985) is to analyse costs by assessing the firm's value chain. By scrutinizing the entire set of cost elements internal and external to the firm, and key linkages among those elements, management may be able to identify opportunities for cost reduction. Moreover, innovative cost leaders such as IKEA have succeeded by dramatically reconfiguring the value chain, departing from what had been standard practice in the industry.

The profit potential of a cost leadership strategy may depend on industry characteristics. In industries with relatively undifferentiated products, such as chemicals, metals and other raw materials, the cost leader's offerings are attractive to the entire customer market, and the benefits of cost leadership are apparent. In industries with differentiated products, the cost leader often sells a relatively basic product and/or a narrow product line, at a price (and potentially, quality level) below that of more differentiated competitors. The success of a cost leadership strategy in such industries depends on having a set of customers who find the low price attractive, despite some sacrifice of quality or variety.

While the viability of a cost leadership strategy may appear to depend on high volume, that is not necessarily the case. Cost leaders often thrive in market niches or localized markets, even though they fail to become dominant players in the broader industry.

A cost leadership strategy has various risks. Efforts to become the low cost producer (which requires displacement of the existing industry leader, unless the industry is new) may fail to prove successful. Attempts by multiple firms to attain the position of cost leader can lead to strong industry rivalry and intense price competition. Once achieved, a successful cost leadership strategy requires constant vigilance to be sustained over time. Typically, a strong culture of frugality must be maintained, and temptations to expand the product line or move upscale must be resisted. The cost leader may develop a reputation for low quality, and the segment of price-sensitive customers may diminish.

See Also

- ▶ Business Strategy
- ▶ Competitive Strategy
- ▶ Cost
- ▶ Porter, Michael E. (Born 1947)

References

Porter, M.E. 1980. *Competitive strategy*. New York: Macmillan.

Porter, M.E. 1985. *Competitive advantage*. New York: Macmillan.

Porter, M.E. 1996. What is strategy? *Harvard Business Review* 61–78.

Cost-Benefit Analysis

William R. Stromeyer and Jay B. Barney
Ohio State University, Fisher College of Business, Columbus, OH, USA

Abstract

Cost–benefit analysis is an analytical technique that developed out of welfare economics and the concept of Pareto efficiency. It is utilized primarily in the public domain as a means of determining which alternative to the status quo will provide the greatest increase in social welfare. By assessing both the expected benefits of a policy or project, as well as the anticipated costs, it permits the assignment of expected net present values to each alternative. The largest challenges to conducting a robust cost–benefit analysis include predicting and monetizing impacts, utilizing the appropriate discount rate and accounting for uncertainty.

Definition Cost–benefit analysis is the systematic cataloguing and identification of impacts, including both expected social benefits and anticipated social costs, for the possible outcomes of alternative policies or projects. Monetization of these impacts facilitates the calculation of each alternative's net social benefit. Cost–benefit analysis determines the soundness of each proposal and permits the ranking of alternatives by net social present value.

History

Cost–benefit analysis (CBA) has a long history in economic thought and policy formation, dating back at least to 1844 and Jules Dupuit's classic work on the utility of public works (Dupuit 1995). The utilization of CBA came into prominence in the United States with the passage of the River and Harbor Act of 1902, which required a board of engineers to report on the desirability of river and harbour projects by taking into account the benefits accrued from additional commerce and the related costs (Prest and Turvey 1965). Subsequently, an increased focus on promoting the social good led to the passage of the 1939 Flood Control Act, which included a clause specifically requiring that accrual of benefits be in excess of the estimated costs. The practice of, and requirement for, CBA then spread to other agencies and departments.

In 1950, an inter-agency committee was tasked with producing a set of generally accepted principles for conducting cost–benefit analyses. This endeavour provided a formal link between government policy and welfare economics. This codification, along with similar work in the UK, served as the foundation of cost–benefit techniques utilized for highway and motorway investments that were occurring in both countries. Over the last 40 years the techniques of CBA have been further developed, to the extent that substantial guidance now exists for almost all projects that utilize public funds or resources.

Brief Overview of Conceptual Foundations

Modern welfare economics and CBA are premised on the concept of *Pareto efficiency*. A project or policy improves Pareto efficiency if it is possible to find a set of transfers that makes at least one person better off without making anyone else worse off. CBA is a technique for determining if the expected benefits of a proposed project or policy outweigh the costs, including the foregoing of other alternatives. Further CBA permits the ranking of alternative projects by either a measure of net present value (NPV) or cost–benefit ratio. Under certain restrictive assumptions, selection of the project or policy with the highest cost–benefit ratio should lead to an increase in the social surplus above and beyond the current status quo. Under optimal conditions the net transfers from the formation of this social surplus will lead to a move towards a Pareto efficient state.

The utilization of CBA requires an understanding of two important concepts: willingness to pay and opportunity costs. Willingness to pay (WTP) is a method for valuing the outputs of a policy, and

opportunity costs is a method for valuing the resources required to implement the policy or project. WTP assesses the amount that each individual affected by the policy would need to pay out or receive in order to be indifferent between the proposed policy and the status quo. The opportunity cost is the value that inputs would have if deployed in the next best alternative use. For a policy to have a net positive benefit it is necessary that the aggregate benefit for all affected individuals, measured by WTP, is positive and that this amount is greater than the opportunity cost of the required inputs. The relationship between net benefits and Pareto efficiency is maintained when a CBA values all impacts in terms of willingness to pay and all costs in terms of opportunity costs. Under these conditions a positive net benefit indicates that it is possible to compensate those who bear the costs of the policy to a sufficient degree that no one is made worse off and at least one individual is made better off (Boardman et al. 2011).

While the theoretical underpinnings of CBA are sound, actually structuring decision rules around Pareto efficiency is extremely difficult owing to the informational requirements necessary to conduct such analyses and the impracticality of measuring individual preferences. Instead, CBA decisions are often based on the *Kaldor–Hicks* criterion that states that a policy is acceptable if and only if those who will gain can fully compensate those who will lose (Hicks 1939; Kaldor 1939). This criterion is the basis for the *potential Pareto efficiency rule*: adopt a policy if and only if the net benefits are positive. The aim of this rule is to maximize aggregate social wealth, lead to an averaging of winners and losers across policies, account for relevant parties other than well-organized groups with political connections, and foster a more equal distribution of wealth.

Process

The following is a list of steps that comprise a generic CBA.

1. Identify and specify the alternative policies or projects
2. Compile a list of affected parties
3. Select the means of measuring costs and benefits and identify related elements
4. Predict outcome of costs and benefits over the duration of the project
5. Monetize all costs and benefits
6. Determine appropriate discount rate
7. Calculate the NPV of alternatives
8. Conduct sensitivity analysis
9. Produce and provide recommendations

Valuation Theory and Monetizing Impacts

A crucial task in the conducting of a CBA is the valuation of costs and benefits, including predicting, assessing and monetizing the impacts of the proposed project or policy. Valuations of benefits and costs tend to be separated into those that occur in primary markets and those that occur in secondary markets. Primary markets are those goods or services that will be directly affected by the proposal. For example, the construction of a new subway will have a direct effect on the supply and demand for public transportation. Secondary markets account for those goods and services that will be affected by changes in the primary market. These goods tend to be substitutes for or complements to the identified primary market. Valuation techniques are further differentiated by the condition of the market under consideration: efficient, distorted or non-existent (Musgrave 1969).

While valuation theory provides the basis for monetization, the challenge of identifying, predicting and monetizing all impacts of a project or policy can still be daunting. Predicting impacts is focused on future outcomes, but the only source of reference tends to be what has happened in the past. Inferences from prior policy outcomes inform predictions for the likely outcomes of new proposals. Several methodologies for identifying impacts have been developed, including the prediction of incremental impacts relative to the status quo, the use of data from ongoing policy and projects, evaluation and comparison to a similar policy, meta-analysis of similar policies, prediction via general elastics and simulation.

The actual monetization of impacts tends to be accomplished either through surveys which measure stated preferences or via the observation of market behaviours that demonstrate revealed preferences. Valuing impacts from observed behaviour is generally accomplished by experiments or quasi-experiments, direct estimation of demand curves or via the indirect market method. Surveys designed to capture stated preferences are generally referred to as contingent valuation methods, since respondents are not actually engaging in a transaction. While economists prefer to observe market behaviour, contingent valuations are often needed for public goods that have no clear market proxy.

Discounting and Determining the Discount Rate

In order to permit the comparison of alternatives, conducting a CBA requires an adjustment of costs and benefits to account for the time value of money. The time value of money captures the tendency of individuals to prefer consumption in the current period versus future periods, and to account for the fact that resources deployed in the current period can generate wealth. The specification of a discount rate allows for the calculation of the expected future value or the expected present value of costs and benefits. Generally, CBA is conducted with present value amounts, as this permits the computation of the NPV of a policy. NPV analysis permits the selection of policies with the highest NPV under the assumption that policies are independent and mutually exclusive. In order to compare projects with different time frames it is often necessary to either implement the *roll-over* method or *equivalent annual net benefit* method. CBA can also be conducted in either *nominal* values (current dollars) or in inflation-adjusted *real* values. Real values account for changes in purchasing power and price inflation. However, utilizing real values requires the estimation of future inflation, which is generally accomplished either through estimating the change in the potential future value of the consumer price index (CPI), inflation forecasts, survey measures of inflation expectations, or taking the implied inflation yield from treasury inflation protected securities (TIPS).

As CBA is applied to public policies and projects the appropriate discount rate to use is called the social discount rate (Dasgupta and Pearce 1972), which aggregates individual preference for marginal rate of time preference and marginal rate of return on private investment. Four approaches to determining the social discount rate include estimating the marginal rate of return on private investment, estimating the social marginal rate of time preference, utilizing the government's long-term borrowing rate, and using the prior three sources to compute a weighted average. However, for most public projects the actual rate to be used is determined by a central government agency or department. In the United States the Office of Management and Budget usually prescribes this rate as specified in Circular No. A-94 and subsequent revisions (OMB 1992).

One of the main points of contention over CBA is the proper specification of the discount rate, which is particularly influential for projects or policies with long-term effects or long life spans. A low discount rate tends to weigh current and future values equally, thus future generations are given roughly equal weight to our own. Large discount rates penalize benefits that accrue further into the future, and are thus more likely to favour the current generation. Of particular debate recently has been the appropriate discount rate to utilize when assessing climate change interventions or other environmental issues (Price 2000).

Dealing with Uncertainty

Inherent in the CBA process is the need to deal with uncertainty, as proposed policies and projects often have attending costs and benefits that stretch over many years. There are two primary methods for addressing uncertainty. The first treats uncertainty as quantifiable risk. *Expected value analysis* assigns probabilities to the various specified contingencies, and takes a weighted expectation of their net present values. The validity of this technique depends on taking an adequate number of contingencies and correctly assigning probabilities. Further, it requires that contingencies are mutually exclusive and that periods are independent of one another.

If, on the other hand, the accrual of net benefits or the probability of these accruals depends on prior contingencies, then more advanced techniques need to be used. One such alternative technique is *decision analysis*, in which a sequential, extended tree is developed, linking each period to the next via a decision node and probabilistic contingencies. Using backward induction, each potential pathway can be reverse solved to determine an expected NPV. This permits identification of the best alternative solution for each period, depending on what has occurred in the prior periods. This determination of period-by-period solutions permits the integration of the value of information and the formation of quasi-option valuation. This is the value gained from delaying decisions in order to acquire new information that will facilitate more informed decision-making.

A crucial step in CBA is the utilization of *sensitivity analysis* to test the robustness of the analysis to changes in the assumptions (Dasgupta and Pearce 1972). Three methods for conducting this are *partial sensitivity analysis*, *worst-case and best-case analysis* and *Monte Carlo analysis*. Partial sensitivity analysis examines how the analysis changes when a limited number of crucial assumptions are changed, holding all other assumptions constant. Worst-case and best-case analysis provide two alternatives to the specified base-case of the analysis by taking assumptions to their logical extremes. Monte Carlo analysis assigns a probability distribution to each of the assumptions of import, and then, through computer-aided resampling, generates an overall distribution for the outcome. This method takes account of the greatest range of information available, and also generates information about the variance in the expected outcomes.

Public Versus Private Applications

The techniques and theory of CBA are generally related to the public domain. Public applications focus on the optimization of social welfare and are tied to the social discount rate. Although CBA per se is not often directly used in the private sector or at the firm level, there are similar concepts in practice. Private usage at the firm level tends to focus on firm-value outcomes, and is often conducted in the process of ▶ capital budgeting or in the assessment of new firm strategies or opportunities. In this case, the optimization target of the analysis will most likely be the firm's stock price, and thus the analysis should utilize the weighted average cost of capital for the discount rate. Firm undertakings that are likely to involve significant shifts in the firm's strategy or that require substantial firm resources require the same types of uncertainty analysis as used in general CBA. Further, if a project will entail new capital it may be necessary to adjust the WACC for anticipated changes in the firm's capital structure. Recently, there has been an increased interest in utilizing CBA for determining how a private firm may best address corporate social responsibility issues (McWilliams and Siegel 2001).

Limitations

A primary critique of CBA is that the accuracy of the analysis is highly dependent on how well costs and benefits have been estimated, which impacts have been considered, and which parties have been granted standing in the analysis. Inaccurate analyses can lead to inefficient decision-making under the Kaldor-Hicks criteria or the potential Pareto efficiency rule. In conducting a CBA there is a trade-off between convenience and usability versus evaluative rigour. The primary methodologies of valuation in CBA are market-centred in that they rely on market allocation and the observation of explicit or implied market supply and demand. Such techniques may not be appropriate in assessing the value of crucial public goods, such as the environment, or difficult to value impacts, such as lives saved or the intrinsic value of freedom (Sen 2000).

See Also

- ▶ Capital Asset Pricing Model (CAPM)
- ▶ Capital Budgeting
- ▶ Externality
- ▶ Public Policy: Strategy in the Public Interest
- ▶ Risk and Uncertainty
- ▶ Stakeholder

References

Boardman, A.E., D.H. Greenberg, A.R. Vining, and D.L. Weimer. 2011. *Cost–benefit analysis: Concepts and practices*, 4th ed. Upper Saddle River: Pearson Education.
Dasgupta, A.K., and D.W. Pearce. 1972. *Cost benefit analysis: Theory and practice*. New York: Harper & Row.
Dupuit, J. [1844] 1995. De la mesure d l'utilité des travaux publics. *Revue française d'économie* 10: 55–94.
Hicks, J.R. 1939. The foundations of welfare economics. *The Economic Journal* 49: 696–712.
Kaldor, N. 1939. Welfare propositions of economics and interpersonal comparisons of utility. *The Economic Journal* 49: 549–552.
McWilliams, A., and D. Siegel. 2001. Corporate social responsibility: A theory of the firm perspective. *Academy of Management Review* 26: 117–127.
Musgrave, R. 1969. Cost–benefit analysis and the theory of public finance. *Journal of Economic Literature* 7: 797–806.
OMB (The Office of Management and Budget). 1992. *Circular* No. A-94 (rev. edn). Washington, DC.
Prest, A.R., and R. Turvey. 1965. Cost–benefit analysis: A survey. *The Economic Journal* 75: 683–735.
Price, C. 2000. Valuation of unpriced products: Contingent valuation, cost–benefit analysis and participatory democracy. *Land Use Policy* 17: 187–196.
Sen, A. 2000. The discipline of cost–benefit analysis. *Journal of Legal Studies* 29: 931–952.

Creative Destruction

Sohvi Leih[1] and David J. Teece[2,3]
[1]University of California, Haas School of Business, Berkeley, CA, USA
[2]Berkeley Research Group, LLC, Emeryville, CA, USA
[3]Haas School of Business, University of California, Berkeley, Berkeley, CA, USA

Abstract

Creative destruction describes the process of how economic progress emerges from strong competition which destroys weaker competitors while Karl Marx explained how capitalism finds this process, it was Joseph Schumpeter who popularized the term. He described capitalism as driving "the perennial gale of creative destruction" Schumpeter has been influential in explaining how economies evolve. Economic progress is not a simple linear growth process, but is characterized by a non-linear process involving the creation of novelty and the destruction of old processes and structures. Entrepreneurs building new companies can shape market outcomes as can the managers of incumbent firms.

Classification

Innovation and strategy

Definition

Creative destruction describes the revolutionalizing process in which a new product, process, or method replaces the old; in other words, it describes the process of how capitalist economic development occurs based on the destruction of prior institutional, business, and market structures. What makes capitalism distinctive is the decentralized and distributed capacity for introducing new patterns of behavior; whether they are technological, organizational or social, they are the fuel that drives economic change (Metcalfe 1998, p. 3). Entrepreneurs building new companies are often the instrument of creative destruction; incumbent firms are usually the victim.

The term "creative destruction" is often attributed to Austrian American economist, Joseph Schumpeter, who viewed the term as "the essential fact about capitalism." He believed that the process of creative destruction is at the core of economic growth. In *Capitalism, Socialism, and Democracy*, Schumpeter (1942) wrote:

> The opening up of new markets, foreign or domestic, and the organizational development from the craft shop to such concerns as U.S. Steel illustrate the same process of industrial mutation – if I may use that biological term – that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism. (Schumpeter 1942: 83)

Schumpeter's description of capitalism as "the perennial gale of creative destruction" has been influential in our understanding of how economies evolve. "The essential point to grasp is that in

dealing with capitalism we are dealing with an evolutionary process," Schumpeter wrote (p. 82). Economic progress is not a simple linear growth process, but it is characterized by a non-linear process in which new firms invent new products and processes which displace older products and processes and the firms that make them.

Schumpeter's view of the economic world stands in stark contrast with the static equilibrium model prevailing in economics during Schumpeter's years. These neo classical approaches contain no provision for innovation, entrepreneurship, and technology. Schumpeter does not assume that markets tend toward equilibrium, but instead that entrepreneurs and technologies create disequilibrium, which leads to new profit opportunities. Schumpeter highlights the inability of static equilibrium analysis to capture the essential long-term features of capitalist reality (Rosenberg 2013, p. 7).

Schumpeter also viewed innovation as the engine of growth, and recognized that innovation is endogenously generated by profit-seeking firms. However, he viewed innovation in the narrow sense ("revolution from within the economic system") also providing the fuel for the creative destruction process. That is, innovations destroy existing industry structures and result in new industry structures. Entrepreneurship and competition facilitate creative destruction. Schumpeter summed it up as follows:

> The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers' goods, the new methods of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprise creates. (p. 83)

Schumpeter Versus Marx

While the term 'Creative destruction' has often been attributed to Joseph Schumpeter (1942), the idea perhaps originated from Marx as discussed in the work of Werner Sombart. Schumpeter also discussed Marx's thought at length in *Capital, Socialism and Democracy* (1942). Although Marx didn't explicitly use the term "creative destruction," in the *Communist Manifesto* by Marx and Engels (1848), creative destruction is implied in the processes of the accumulation and annihilation of wealth under capitalism. Marx further developed the idea in *Grundrisse* (1857) and *Das Kapital* (1863). Marx and Engels (1848) described the crisis tendencies of capitalism in terms of "the enforced destruction of a mass of productive forces." As Marx expressed it in the *Communist Manifesto*, the bourgeoisie "has played a most revolutionary role in history." Capitalism,

> It cannot exist without constantly revolutionizing the instrument of production, and thereby the relations of production, and with them the whole relations of society. The need of a constantly expanding market for its product chases the bourgeoisie over the whole surface of the globe. The bourgeoisie, by the rapid improvement of all instruments of production, by the immensely facilitated means of communication, draws all nations, even the most barbarian, into civilization. The bourgeoisie, during its rule of scarce 100 years, has created more massive and more colossal productive forces than have all preceding generations together. (Marx and Engels 1976, pp. 487–89)

Marx and Shumpeter both share "a vision of economic evolution as a distinct process generated by the economic system itself" (Clemence 1951: 160). Marx, no less than Schumpeter, perceived capital accumulation to occur irregularly, in bursts, with cyclical consequences (Marx 1906, pp. 672, 693–94).

Schumpeter and Marx shared a common vision, including the inherent instability of capitalism and the inevitability of "crises", and the eventual destruction of capitalist institutions and the arrival of a socialist form of economic organization as a result of the working out of the internal logic of capitalist evolution (Rosenberg 2011, p. 1215).

Although both Schumpeter and Marx highlight the discontinuous nature of the capitalist dynamic, the causality is different. Schumpeter views the individual entrepreneur's initiative as the core driver for economic development and technological change (Foster 1983, p. 328). Schumpeter describes a broader process by which capitalism, through its "creative" success, leads on to its own "destruction" as an economic system and prepares

the way – technologically, institutionally, and psychologically – for a socialist economic system to succeed and supersede it (Elliott 1980).

However, in Marx's view, the structure of accumulation itself establishes the important source of capitalist development (Foster 1983, p. 328). To put the matter briefly, Marx creates his model of stationary equilibrium (simple reproduction) by "assuming away" accumulation but not the capitalist, while Schumpeter's model of the circular flow abstracts from the existence of the entrepreneur himself (Foster 1983, p. 328).

Examples A classic example of Schumpeterian creative destruction is the introduction of the personal computer. The emergence of the personal computer introduced by then small firms such as Apple, Dell and HP made many of the industry's leading minicomputer firms disappear. Incumbent computer manufacturers including DEC, Data General, Control Data, Prime Computer, and UNIVAC lost significant amounts of their value and employment while a new set of dominant players gained the value that previous leaders of their industry lost (Spencer and Kirchhoff 2006).

This pattern has happened in other industries. Christensen (1997) observes that even in the less technologically fast moving industries such as power shovels, the shift from cable to hydraulics left the industry with a new set of leaders while the old dominant players faded away. Similar effects are found in the typesetting (Tripas 1997) and digital imaging (Tripas and Gavetti 2000) industries.

The success of Netflix might be another example of creative destruction. Netflix has been so disruptive to existing industries that its impact is often referred to as the "Netflix effect." Employment in the video/disk rental industry has decreased by 93% in a decade – from 153,000 jobs in 2005 to fewer than 11,000 in 2015 (Perry 2015).

The process of Schumpeterian can cause temporary economy-wide distress, such as loss of employment or income. Schumpeter acknowledges (1942) that displaced individuals might react with strong resentment:

> Secular improvement that is taken for granted coupled with individual insecurity that is acutely resented is of course the best recipe for breeding social unrest (Schumpeter 1942, p. 145). . .This type of reaction leads to the labor movement, and it is supported by intellectuals whose hostility increases with every achievement of capitalist evolution. (p. 154)

However, economists and policy analysts have concluded that such distress is an inherent part of economic growth, and that societies that allow creative destruction grow more productive and wealthy over time. For example, in 1900 almost 40 of every hundred Americans were farmers but in 2000, it took just two out of every hundred workers to feed America. Despite the decline in the number of farmers, the US is still a major agricultural exporter thanks largely to advances in agricultural productivity (Cox and Alm 2008).

Related Works in Economics and Management

Some economists and many management scholars, following Schumpeter, have long recognized the "disruptive" impact of innovation. As two Berkeley faculty noted 25 years ago:

> The development, commercialization, and diffusion of product and process technologies have long been the most fundamental competitive forces in advanced industrial economies, generating economic growth, enhancing consumer welfare, and in the process, constantly challenging and frequently overturning the established order within and among industries. If one calibrates competition by the intensity of rivalry among industry participants, then innovation is unquestionably the major force driving competition. (Jorde and Teece 1993)

In the same vein, Abernathy and Clark (1985) developed the "transilience map" a framework analyzing innovations according to their effects on markets and competences. In their framework, they discussed disruptive innovations that disrupt markets and competencies. Bower and Christensen further develop this concept and note that disruptive technologies may not be '... radically new from a technological point of view' but have superior 'performance trajectories' along critical dimensions that customers value (Bower and Christensen 1995; Christensen 1997).

Herbert Simon (1982) had a similar view of creative destruction by arguing that potential threats to a firm's survival can lead to the change of routines. According to Simon's model of satisficing behavior, firms maintain their routines as long as they can uphold satisfactory performance. Otherwise, they destroy their old routines and search for superior routines. Schumpeter would have disagreed with such a view of "creation" since Schumpeter views creation as an independent event rather than an adaptive response to pressures. Entrepreneurs' innovative behaviors are individualistic and heroic actions, and it is the introduction of innovations into the old system that causes the destruction of old routines.

A version of the creative destruction thesis has entered the management literature and become immensely popular and influential in the form of Clay Christensen's "Disruption Innovation" thesis (Christensen 1997). Christensen posits that Schumpeterian creative destruction follows particular pathways. He argues that incumbents have an advantage in serving the current customer base, but they get blindsided and fail to recognize the needs of new customers. New entrants arrive and disrupt incumbents, often by entering the market with cheaper and possibly lower performance products; but they eventually upgrade and win share from the incumbent. Implicitly, Christensen assumes that incumbents can effectively transform. . . but only if they put their mind to it. Failures are caused by bad management focusing on the immediate road ahead and not seeing challenges and opportunities that exist around the corners. Clearly, disruption is a syndrome that is common in highly competitive innovation driven markets.

Dynamic capabilities (Teece and Pisano 1994; Teece et al. 1997; Teece and Leih 2016) and disruption are related ideas, with dynamic capabilities being the more general framework that readily accommodates creative destruction/disruptive innovation. However, dynamic capabilities does not endeavor to predict the particular pathways by which innovators using new technology overturns the status quo and challenges incumbents. There remains much to be learned about particular contexts and strategies available to new entrants and to incumbents alike.

Utterback notes that Christensen's concept is seductive due to the clarity of its examples and "its claimed power and generality" (Utterback and Acee 2003; Christensen et al. 2002). While recognizing the importance of the issues Christensen is addressing, Utterback goes on to summarize the Christensen thesis and then challenges it:

> In Christensen's theory of disruptive technology the establishment of a new market segment acts to channel the new product to the leading edge of the market or the early adopters. Once the innovation reaches the early to late majority of users it begins to compete with the established product in its traditional market.

And juxtaposed with his own research findings:

> Here we have presented an alternative scenario in which a higher performing and higher priced innovation is introduced into leading established market segments and later moves toward the mass market. Diffusion for example, of fuel injection started with the luxury and sports car segments and then migrated into other segments. The first use of electronic calculators was in the scientific community. Later simpler, less expensive and portable models expanded the total market by creating new segments which later included the mass market. Cooper and Schendel similarly discuss the down-market progression of the ball point pen which was originally more expensive than the fountain pen. Continued development resulted in the "throw away" pen which opened up new market segments. (Acee 2001, p. 43)

The evidence would seem to suggest that there are multiple pathways by which innovation disrupts the status quo, as Utterback has documented. Innovation can also strengthen the status quo of incumbents as Tushman and Anderson (1986) has indicated and of complementors as Teece (1986, 2006, 2017[1]) has suggested.

Interestingly, the management literature suggests that the problems of "disruption" (failure to respond to new technologies) is as much a decision making problem as it is a lack of capabilities. Cooper and Schendel (1976) noted half a century ago that:

[1]Profiting from Innovation in the Digital Economy is forthcoming in Research Policy, 2017.

> . . .a typical sequence of events involving the traditional firm's responses to a technological threat begins with the origination of a technological innovation outside the industry, often pioneered by a new firm. Initially crude and expensive, it expands through successive sub-markets, with overall growth following an S-shaped curve. Sales of the old technology may continue to expand for a few years, but then usually decline, the new technology passing the old in sales within 5 to 14 years of its introduction.

Cooper and Schendel conclude that failure to respond appropriately occurs because "decisions about allocating resources to old and new technologies within the organization are loaded with implications for the decision makers; not only are old product lines threatened, but also old skills and positions of influence."

These observations drive home that responding to innovation driven competition requires a variety of strategies, depending on whether the innovation is competency destroying, or competency enhancing; and when it is the former, it's not just a matter of upgrading capabilities. It's also a matter of selecting the right projects, as the dynamic capabilities framework indicates.

See Also

- ▶ Clausewitz, Carl von (1780–1831)
- ▶ Disruptive Technology
- ▶ Innovation
- ▶ Schumpeter, Joseph (1883–1950)
- ▶ Systemic Innovation

References

Abernathy, W.J., and K.B. Clark. 1985. Innovation: Mapping the winds of creative destruction. *Research Policy* 14: 3–22.

Acee, Happy J. 2001. *Disruptive technologies: An extended view*. SM Thesis, Management of Technology, Massachusetts Institute of Technology.

Aghion, P., and Howitt, P. A model of growth through creative destruction. NBER paper No. 3223.

Bower, J.L., and C.M. Christensen. 1995. Disruptive Technologies: Catching the Wave. *Harvard Business Review* 73 (1): 43–53.

Christensen, C.M. 1997. *The innovator's dilemma: When new technologies cause great firms to fail*. Boston: Harvard Business School Press.

Christensen, C.M., M.W. Johnson, and D.K. Rigby. 2002. Foundations for growth: How to identify and build disruptive new business. *MIT Sloan Management Review* 43 (3): 22–31.

Clemence, R.V. 1951. *Essays. Joseph A. Schumpeter*. Cambridge, MA: Addison-Wesley.

Cooper, A.C., and D. Schendel. 1976. Strategic responses to technological threats. *Business Horizons* 19: 61–69.

Cox, W.M., and Alm, R. 1992. The Churn. Federal Reserve Bank of Dallas. Annual report. 5–18.

Cox, W.M., and Alm, R. 2008. Creative destruction. Library of Economics and Liberty.

Davis, S., J. Haltiwanger, and S. Schuh. 1996. *Job creation and destruction*. Cambridge, MA: MIT Press.

Elliott, J.E. 1980. Marx and Schumpeter on capitalism's creative destruction: A comparative restatement. *Quarterly Journal of Economics* 94: 45–68.

Foster, J.B. 1983. Theories of capitalist transformation: Capital notes on the comparison of Marx and Schumpeter. *Quarterly Journal of Economics* 327–331.

Jorde, T.M., and D.J. Teece. 1993. Rule of reason analysis of horizontal arrangements: Agreements designed to advance innovation and commercialize technology. *Antitrust Law Journal* 61: 579–619.

King, A. Understanding dynamic pathways by combining two important theories, working paper.

King, A.A., and B. Baatartogtokh. 2015. How useful is the theory of disruptive innovation? *MIT Sloan Management Review* 57 (1): 77.

Marx, Karl. 1906. Capital, Vol. I. New York: Modern Library, (n.d.) Originally published by Charles H. Kerr, Chicago.

Metcalfe, S. 1998. *Evolutionary economics and creative destruction*. London: Routledge & Kegan Paul.

Perry, M.J. 2015. *The 'Netflix effect': An excellent example of 'creative destruction'*. American Enterprise Institute.

Reinert, E.S. 2002. Schumpeter in the context of two canons of economic thought. *Industry and Innovation*. 9 (1/2): 23–39.

Rosenberg, N. 2011. Was Schumpeter a Marxist? *Industrial and Corporate Change* 20 (4): 1215–1222.

Rosenberg, N. 2013. *Schumpeter and the endogeneity of technology: Some American perspectives*. Routledge.

Salter, W.E.G. 1960. *Productivity and technical change*. Cambridge: Cambridge University Press.

Schumpeter, J.A. 1942. *Capitalism, socialism and democracy*. New York: Harper and Brothers.

Simon, H.A. 1982. *Models of bounded rationality*. Cambridge, MA: MIT Press.

Spencer, A.S., and B.A. Kirchhoff. 2006. Schumpeter and new technology based firms: Towards a framework for how NTBFs cause creative destruction. *International Entrepreneurship and Management Journal* 2 (2): 145–156.

Teece, D.J. 1986. Profiting from technological innovation. *Research Policy* 15: 285–305.

Teece, D.J. 2006. Reflections on profiting from innovation. *Research Policy* 35 (8): 1131–1146.

Teece, D., and S. Leih. 2016. Uncertainty, innovation, and dynamic capabilities: An introduction. *California Management Review* 58 (4): 5–12. https://doi.org/10.1525/cmr.2016.58.4.5.

Teece, David, and Gary Pisano. 1994. The dynamic capabilities of firms: An introduction. *Industrial and Corporate Change* 3 (3): 537–556.

Teece, D.J., G. Pisano, and A. Shuen. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* 18 (7): 509–505.

Tripsas, M. 1997. Unraveling the process of creative destruction: Complementary assets and incumbent survival in the typesetter industry. *Strategic Management Journal* 18: 119–142.

Tripsas, M., and G. Gavetti. 2000. Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal* 21 (10/11): 1147–1161.

Tushman, Michael L., and Philip Anderson. 1986. Technological discontinuities and organizational environments. *Administrative Science Quarterly* 31: 439–465.

Utterback, J.M., and H.J. Acee. 2003. Disruptive technology. Retrieved from www.sussex.ac.uk: http://www.sussex.ac.uk/Units/spru/events/KP_Conf_03/documents/Utterback_Acee.pdf.

Creative Industries

Paolo Boccardelli
LUISS, Business School, Rome, Italy

Abstract

Through the exploitation of individual creativity, skill and talent, creative industries generate a potential for wealth and job creation by the means of the development and exploitation of intellectual property rights. These activities can be divided into two main components: the creative production and the complementary resources and activities to bring the outcome of the creative efforts to the markets. Based on a value chain perspective these industries are defined in four different phases: origination, production, distribution and consumption of the creative outcome. While the origination and production phases are characterized by a limited potential for gaining economies of scale and thus by small operators, the distribution and consumption ones can be dominated by large firms.

Definition Creative industries are those industries which have their origin in individual creativity, skill and talent and which have a potential for wealth and job creation through the generation and exploitation of intellectual property in different products and services markets.

The definition of 'creative industries' is far from being clear and universally accepted. In fact, there is considerable confusion over the precise definitions of the terms creative industries, cultural industries and the cultural sector (Stapleton 2007).

Despite this confusion the term 'creative industries' has received considerable attention during recent years and it refers to suppliers of a range of products that 'we broadly associate with cultural, artistic, or simply entertainment value' (Caves 2000: vii). Although this attention is certainly due to the emotional and symbolic value of creative and cultural products in society, since in many countries they benefit from financial support by the public authorities, the economic role of these industries has recently grown. Indeed, before the emergence of the global economic downturn in 2008, the entertainment and media industries were forecast to grow from $1.3 trillion in 2005 to reach 1.8 trillion by 2010. In the 1990s, the creative economy in the Organisation for Economic Co-operation and Development (OECD) countries grew at an annual rate twice that of service industries and four times that of manufacturing (Howkins 2001).

Although in the US the definition of creative industries is generally conservative and based on those businesses involved in the creation or distribution of the arts (Americans for the Arts 2008), as of January 2008 the analysis of these industries reports that more than 612,000 businesses exist and that they employ about three million individuals in the country, with a growth rate larger than in the rest of the US economy. Between 2007 and 2008 the growth in businesses was about 12% (10% in the rest of the economy), while the growth in employees was 11.6% (four times larger than the rest of the economy at 2.4%) (Americans for the Arts 2008 – based on the Dun and Bradstreet database). According to the 2007 OECD report, in the US the economic contribution of

cultural industries to gross value added (GVA) in 2002 accounted for about $340 billion (3.3%) (OECD 2007). More recently (2008), based on data from Datamonitor and Hoover's, the economic contribution of creative industries to the GVA was estimated at about $580 billion.

The growth of the cultural and creative sector in the European Union from 1999 to 2003 was 12.3% higher than the growth of the overall economy (European Commission 2006), and its turnover in the EU contributed 2.6% of the European Union's GDP in 2003. The cultural sector employed at least 5.8 million people in Europe in 2004, which is more than the total working population of Greece and Ireland put together (European Commission 2006). Over the period 2000–2005, international trade in creative goods and services experienced an average annual growth rate of 8.7%, with a value of world exports of $424.4 billion in 2005, representing 3.4% of total world trade, according to UNCTAD (UN 2008).

An even broader interpretation of creative economies has been put forward by Richard Florida (2002, 2003, 2007), who underlined the role of creativity in promoting the economic, social and cultural dynamism of a country. According to Florida (2007) almost one-third of the US workforce can be identified with the creative class, and the creative sector accounted for nearly half all wage and salary income, about $1.7 trillion, including in these data many other creative professionals in business, finance and law. In line with this analysis, Landry underlined the emergence of 'creative cities', as the *loci* where cultural activities can be viewed as an integral component of the city's economic and social life (Landry 2000; UN 2008). A good example of a creative city is London, where the creative sector is the second biggest in the city, and where it grew faster than any other major industry except financial and business services and accounted for between 20% and 25% of job growth in the city over this period (The Work Foundation 2007).

This accounting of a wider economic contribution of creative sectors may be partly due to some emerging trends. Although the term 'cultural industry' was originally coined to identify the development, production and release of cultural content in mass media (e.g., books, magazines, newspapers, sound recordings, films), today the emerging convergence between technologies, media and industries has led scholars and practitioners to apply the term to suppliers of mass media content as well as producers of the traditional arts that do not lend themselves to mass reproduction, such as live performances and the creative arts. This definition, although broad enough to encompass many operators as well as a large part of the workforce, may be further enriched with some other economic activities which are focused on intellectual creation and its exploitation in the marketplace, such as design, fashion, crafts, architecture, sports, software and even tourism.

This last definition has been inspired by the British Department for Culture, Media and Sport (DCMS), which, in the late 1990s, played an important role in promoting the debate on cultural industries and the way they can be harnessed as drivers of economic development. The widely quoted definition by the DCMS points out that creative industries are 'those industries which have their origin in individual creativity, skill and talent and which have a potential for wealth and job creation through the generation and exploitation of intellectual property' (DCMS 2001: 4). It is noteworthy that the designation 'creative industries' that has developed since then has broadened the scope of cultural industries beyond the arts and has marked a shift in approach to potential commercial activities that until recently were regarded purely or predominantly in non-economic terms (UN 2008).

This broad and general definition aims to encompass all the cultural and creative activities that can be streamlined in an industrial and commercial flow of value creation and exploitation. Accordingly, the prominent role of intellectual property (IP) rights in their different shapes and species (e.g., patents, copyrights) affects the industrial organization and dynamics of the creative sectors. However, this definition of the creative sector, in its general meaning, can be considered controversial and tortuous; indeed, numerous attempts by economists, statisticians

and cultural geographers have taken into consideration different mechanisms and phenomena in order to arrive at a suitable categorization of the industry. Much of the literature thus tends to convert the definition provided by the DCMS into a list of specific industries, corresponding to sectors clearly identified in the national and international statistics. In such a way, the classification provided by the DCMS (2006) used the Standard Industrial Classification (SIC), and it identified the following 11 sectors: Advertising; Architecture; Arts and Antique Markets; Crafts; Design; Designer Fashion; Film, Video and Photography; Music and the Visual and Performing Arts; Publishing; Software, Computer Games and Electronic Publishing; Television and Radio. However, sectors included in the SIC codes do not perfectly overlap with creative industries. By comparing each cluster of activities we find that some of them must be analysed at a four-digit level, while some others are not codified at all.

Several other approaches have been developed to define the creative industries (see also the Creative Economy Report 2008 of the United Nations for a synthetic review of these approaches). Beside the symbolic texts model, which focuses attention instead on popular culture, the concentric circles model and the World Intellectual Property Organization (WIPO) copyright model are worth mention (UN 2008). The former asserts that creative ideas originate in the core creative arts and that these ideas diffuse outwards through a series of layers or 'concentric circles', with the proportion of cultural to commercial content decreasing as one moves farther outwards from the centre (UN 2008). In this way, four main layers can be defined: *core creative arts* (e.g., literature and music); *other core cultural industries* (e.g., films, museums and libraries); *wider cultural industries* (e.g., publishing, sound recording, television, radio, video and computer games); *related industries* (e.g., architecture, advertising, design and fashion).

The WIPO model then analyses all the activities around the development, production and commercialization of IP rights (WIPO 2003), resulting in three main segments: *core copyright industries* (e.g., advertising, film and video); *interdependent copyright industries* (e.g., consumer electronics, paper and musical instruments); *partial copyright industries* (e.g., architecture, design and fashion).

However, a different and more productive way of defining the creative industries is based on the distinction between creative and artistic activities, on the one hand, and *humdrum inputs* on the other hand (Caves 2003). The former can be considered the core part of the creative and cultural production, while the latter identify all the complementary resources and activities to bring the outcome of the creative efforts to the different markets. This distinction enables the use of a value-chain perspective, which allows the identification of those operators and activities that directly contribute to the production of the creative content and those which are needed to support this activity.

Based on a value-chain perspective, the creative industries are defined in four different phases:

- Origination, which identifies the creation of cultural and creative ideas and artefacts to embed in a specific product (e.g., designs, songs, stories, pictures, scripts, games, styles, characters);
- Production, which encompasses all the activities to create commercially viable products (e.g., motion picture production, CD recording, TV production);
- Distribution, which allows the diffusion of creative and cultural products in different platforms and channels (e.g., broadcasting, publishing and sale of books, CDs, games, live event production);
- Consumption, which represents the final stage of the process, and endows end consumers with the opportunities to experience cultural and creative products in different ways (e.g., personal equipment like television sets, radios or computers, in concert halls and theatres).

As in many other industries, in the creative businesses some features rule the economic organization of the value chain. Creative businesses are dominated by a wide range of small and in

many cases independent originating operators, relatively few production firms, and to some extent a structured distribution phase. This basic feature is largely explained by the role played by the economies of scale in the origination phase, which, unlike the industrial sectors, appear to be limited. The differentiation and the almost infinite variety of products available do not allow an easy exploitation of scale effects (Caves 2000). Besides this, the industry is affected by a high rate of uncertainty in the economic returns due to volatility on the demand side caused by the emotional character of the products, and moral hazard and information asymmetries on the supply side (Conant 1960; de Vany and Walls 1996, 1999; Caves 2000). The latter can easily be observed in the vertical relationships characterizing the value chain, where the economic relations must be carefully regulated in order to absorb and balance uncertainty and risks between originators, producers and distributors (Caves 2003). Thus, consumer valuation is extremely difficult to predict: taste is acquired through consumption and is subject to a number of factors, including culture, fashion, self-image and lifestyle (the 'nobody knows' property). Distributors usually assume the role of promoting the products, making huge investments that normally are well beyond the budgets of small firms at the top of the supply chain. The proliferation of creative products and the need to spread the risk across many of them so that profitable products subsidize failures are additional features that are associated with economies of scale in distribution.

Moreover, each product is a prototype that embodies a relevant component of creativity and non-recurrent content and employs a wide range of diversified competences (e.g., for film production, see Conant 1960; Caves 2000; Lampel and Shamsie 2000, 2003).

These characteristics strongly affect the way in which the industry is organized. Indeed, the prototypical and non-recurrent nature of creative production is often associated with a project-based organization of the industry and the firms. In such a way, the common structure is project-based, and operations are built on teams of principals that are formed to perform a single project and then disband. In addition, in the vast majority of the creative industries the most valuable and strategic component of resources is dispersed in an industry-wide community of independent professionals (e.g., creative and artistic human capital in the film industry), which must be appointed for specific and not stable projects. The frequently collective nature of creative production, and the need to develop creative teams with diverse skills, who often also possess diverse interests and expectations about the final product, is another key characteristic of these industries. On the other hand, the artistic nature of the production in several cases may offer some opportunities to define contractual relationships that are economically more convenient for the producers than for the artists, who may base their utility function on the artistic creation itself and not on the economic exchange (for *art for art's sake* effect, see Caves 2003).

Nevertheless, the management of the human capital within the creative businesses is a key and a distinctive issue. As pointed out by Teece (2010), human capital resources can be divided into three main 'classes': the 'literati', the 'numerati' and the entrepreneurs. The first two are the highly educated classes, and they represent expert talents. Both groups have specialized skills, but while the literati tend more to the synthesis and the communication of ideas, the numerati are more inclined to analysis. In creative productions both types of talents are diffusively employed: the literati are likely to be used more specifically but not only within cultural productions, the numerati in the software, the architecture and the design businesses.

Managing this type of talented human resources requires a light touch that enhances cooperation and creativity (Teece 2010). Indeed, it requires decentralized and distributed managerial efforts. Moreover, management must foster a creative environment by setting the proper incentives, by removing barriers and by building 'virtuoso' teams that rely more on creativity than on tight project management (Teece 2010).

The economic organization of the creative industries requires at least as much knowledge-intensive as capital-intensive resources, and it

relies heavily on networks to mobilize crucial resources (Lampel and Shamsie 2003). Although it is commonly adopted in sectors like film productions and live entertainment (see, for example, Faulkner and Anderson 1987; Robins 1993; Jones 1996; Miller and Shamsie 1996; DeFillippi and Arthur 1998; Mezias and Mezias 2000), this approach is increasingly influencing other sectors such as fashion and software design, in which the temporary nature of organizational arrangements and of comparative advantages is ruling the game. Accordingly, these features of the creative businesses amplify the role of the capabilities to select strategic resources in the relevant factor markets, to orchestrate and to deploy them in ever-changing projects, which are the vehicles to reach punctuated but temporary advantages on the product markets.

Despite the need to coordinate diverse creative activities within a relatively short time frame, the durability of several creative products – and particularly those based on cultural content and the relative capacity of their producers to extract economic rents – appears to remain long after the period of production. This is, however, mainly dependent on the capability to effectively manage IP rights over the years, over different markets and different channels.

See Also

- ▶ Appropriability
- ▶ Complementary Asset
- ▶ Small World Networks: Past, Present and Future

References

Americans for the Arts. 2008. *Creative industries 2008: The state report*. Washington, DC.

Caves, R.E. 2000. *Creative industries: Contracts between art and commerce*. Cambridge, MA: Harvard University Press.

Caves, R.E. 2003. Contracts between arts and commerce. *Journal of Economic Perspectives* 17: 73–83.

Conant, M. 1960. *Antitrust in the motion picture industry: Economic and legal analysis*. Berkeley: University of California Press.

DCMS (Department of Culture, Media and Sport). 2001. *Creative industries mapping document,* 2nd ed. Department of Culture, Media and Sport. Available at http://www.culture.gov.uk/reference_library/publications/4632.aspx. Accessed 1 Mar 2012.

DCMS (Department of Culture, Media and Sport). 2006. *Creative industries statistical estimates statistical bulletin*. Department of Culture, Media and Sport. Available at http://www.culture.gov.uk/what_we_do/research_and_statistics/4848.aspx. Accessed 1 Mar 2012.

De Vany, A., and W.D. Walls. 1996. Bose–Einstein dynamics and adaptive contracting in the motion picture industry. *Economic Journal* 439: 1493–1514.

De Vany, A., and W.D. Walls. 1999. Uncertainty in the movies: Can star power reduce the terror of the box office? *Journal of Cultural Economics* 4: 285–318.

DeFillippi, R.J., and M.B. Arthur. 1998. Paradox in project-based enterprise: The case of film making. *California Management Review* 40: 125–139.

European Commission. 2006. *Economy of culture in Europe*. Study prepared for the European Commission by KEA. Available at http://ec.europa.eu/culture/key-documents/doc873_en.htm. Accessed 1 Mar 2012.

Faulkner, R.R., and A.B.. Anderson. 1987. Short-term projects and emergent careers: Evidence from Hollywood. *American Journal of Sociology* 4: 879–909.

Florida, R. 2002. *The rise of the creative class and how it's transforming work, leisure, community and everyday life*. New York: Basic Books.

Florida, R. 2003. *Cities and the creative class*. New York: Routledge.

Florida, R. 2007. *The flight of the creative class: The new global competition for talent*. New York: Harper Business.

Howkins, J. 2001. *The creative economy: How people make money from ideas*. London: Penguin.

Jones, C. 1996. Careers in project networks: The case of the film industry. In *The boundaryless career: A new employment principle for a new organizational Era*, ed. M. Arthur and D. Rousseau. New York: Oxford University Press.

Lampel, J., and J. Shamsie. 2000. Critical push: Strategies for creating momentum in the motion picture industry. *Journal of Management* 2: 233–257.

Lampel, J., and J. Shamsie. 2003. Capabilities in motion: New organizational forms and the reshaping of the Hollywood movie industry. *Journal of Management Studies* 40: 2189–2210.

Landry, C. 2000. *The creative city: A toolkit for urban innovators*. London: Comedia/Earthscan.

Mezias, J.M., and S.J. Mezias. 2000. Resource partitioning, the founding of specialist firms, and innovation: The American feature film industry, 1912–1929. *Organization Science* 3: 306–323.

Miller, D., and J. Shamsie. 1996. The resource-based view of the firm in two environments: The Hollywood film studios from 1936 to 1965. *Academy of Management Journal* 3: 519–543.

OECD (Organisation for Economic Co-operation and Development). 2007. *International measurement of the economic and social importance of culture* (document STD/NAFS (2007)1), prepared by J. C. Gordon and H. Beilby-Orrin. Paris: OECD Statistics Directorate. Available at http://www.oecd.org/dataoecd/56/54/38348526.pdf. Accessed 1 Mar 2012.

Robins, J.A. 1993. Organization as strategy: Restructuring production in the film industry. *Strategic Management Journal* 14: 103–118.

Stapleton, J. 2007. *Study on the economic, social and cultural impact of intellectual property in the creative industries*. Study prepared for the World Intellectual Property Organization. Available at http://www.wipo.int/ip-development/en/creative_industry/studies.html. Accessed 1 Mar 2012.

Teece, D.J. 2010. Human capital, capabilities, and the firm: Literati, numerati, and entrepreneurs in the twenty-first century enterprise. In *The Oxford handbook on human capital*, ed. A. Burton-Jones and J.C. Spender. Oxford: Oxford University Press.

UN (United Nations). 2008. *Creative economy report 2008. The challenge of assessing the creative economy: Towards informed policy-Making*. UNCTAD/DITC/2008/2. Available at: unctad.org/en/docs/ditc20082cer_en.pdf. Accessed 1 Mar 2012.

WIPO (World Intellectual Property Organization). 2003. *Guide on surveying the economic contribution of the copyright industries*. Geneva. Available at www.wipo.int/copyright/en/publications/pdf/copyright_pub_893.pdf. Accessed 1 Mar 2012.

The Work Foundation. 2007. *Staying ahead: The economic performance of the UK's creative industries*. Available at www.theworkfoundation.com/assets/docs/publications/176_stayingahead.pdf. Accessed 1 Mar 2012.

Cross-Licensing

Peter Grindley
Berkeley Research Group, London, UK

Abstract

Cross-licensing is an exchange of rights by firms to use each other's property. Patent cross-licensing is used widely in complex technology industries such as ICT and pharmaceuticals to provide 'freedom to design' without risk of infringement, to avoid litigation and to round out product lines. Cross-licenses typically involve balancing royalty payments based on the net contributions of each party's ▶ patents to the other's products. If contributions balance, the licenses may occasionally be royalty-free.

Definition Cross-licensing is an agreement between two or more parties in which each party grants rights to the others to use its intellectual property (IP). There are many types of cross-licenses. Most prevalent are bilateral patent cross-licenses used in complex technology industries such as information and communications technology (ICT) and pharmaceuticals.

Firms use cross-licensing to exchange rights to use each other's intellectual property (IP). Cross-licensing may include all types of IP – ▶ patents, copyrights, trademarks and know-how. Most prevalent are patent cross-licenses used in complex technology industries to provide 'freedom to design' without risk of infringement, to avoid litigation or to round out product lines. Cross-licenses typically include balancing royalty payments based on the net contributions of each party's patents to the other's products. If contributions balance they may occasionally be royalty-free. Most often a cross-license is a bilateral agreement tailored to the specific needs of the parties. A patent pool, in which several firms license their patents for a technology as a group, is also a type of cross-license.

Types of Cross-Licensing

A main use of cross-licensing is in field-of-use patent portfolio cross-licenses (Grindley and Teece 1997). In complex technology industries such as ICT and pharmaceuticals many firms are active in the same technology area. Technological development is rapid and builds on existing technology with short life cycles. Firms generate large numbers of patents and may infringe each other's patents, often unintentionally. It may be difficult to avoid other patents, especially if they are implicated in a standard. This leads to overlapping patent 'thickets' from different firms (Shapiro 2001; Ziedonis 2004). Firms, therefore, often cross-license entire portfolios of patents to ensure 'freedom to design' or 'patent peace' to develop

technology without worrying about mutual infringement. This avoids the costs of firms designing around one another's patents or of searching for possible infringements.

Portfolio cross-licensing simplifies the task of ▶ licensing large numbers of patents. Although cross-licenses may apply to specific patents, more often they include all of the firm's patents for application in a field of use without identifying individuals, as well as new patents granted during the licence period. At the end of the patent period, say 5 years, there may be capture rights for the licensee to continue to use patents issued up to that date, or all rights may cease. Specific patents or fields of use may be excluded if the owner believes they are of exceptional strategic value.

Firms also use cross-licensing to round out their product offerings. Pharmaceutical companies might cross-license each other's proprietary technology, providing each other with more balanced product lines. Firms might agree to cross-license each other's rights to use brand names.

In a typical patent cross-license no know-how is exchanged. The firms already have the capabilities to develop technology and a cross-license is effectively an agreement not to sue each other for infringement. Other types of cross-license may also include knowhow and training in how to use a firm's technology. Know-how licences are more complex than pure patent licences and typically have a higher royalty rate. They may be part of a more elaborate agreement or joint venture.

Royalties

Although cross-licensing provides freedom to design this does not mean that cross-licenses are royalty-free. Royalties are paid based on the relative contributions of one firm's technology to the other's products, with a balancing payment made to the party with the most valuable portfolio.

Royalties are typically assessed in licensing negotiations using a version of the 'proud list' procedure (Grindley and Teece 1997). This estimates the likely contribution of each party's patents to the product earnings of the other. A sample group of each firm's most valuable patents may be rated on a scale of 0–1 for quality, validity and economic contribution to the other firm's products. These factors are multiplied by a reference royalty rate to develop an effective rate for each patent. The rates are calculated for all main patents and applied to the expected royalty base of affected product sales to determine a royalty amount per annum. The implied amounts for the two firms are balanced to give a net payment to the firm with the most valuable portfolio. From this base the parties negotiate a final balancing royalty, according to the firms' bargaining powers. Final royalties may also allow for market pressures on royalties such as potential 'royalty stacking' if a licensee expects a series of claims from other licensors (Geradin et al. 2008; Lemley and Shapiro 2007). The royalty is typically expressed as a running royalty rate applied to the licensee's total sales. Alternatively, the parties may prefer a lump sum or combination.

The final royalty payment represents the balance of the two portfolios rather than their absolute values so there may be some variation. Cross-licensing royalties paid by a firm with many patents to trade will be lower than for a firm with few valuable patents. This makes it difficult to compare royalties in different cross-licenses which depend on specific product and patent combinations. Occasionally, the contributions, or potential litigation strength, of each portfolio may be roughly equal and the parties may waive royalty payments.

Firms may strengthen their bargaining positions and reduce their royalty payments by building up their patent portfolios, either by research and development (R&D) activities or by purchasing patents from other firms, who either may have unused patents or who have now left the industry. Although buying patents can be expensive this can make good business sense as it would reduce a royalty bill.

Licensing and Competition

Firms should be aware of potential competition concerns related to cross-licensing. A cross-license should not act to raise competitors' costs unfairly or be a front for collusion. However, different royalty rates for differently situated

firms should not in themselves indicate unfair competition. An entrant with no technology to trade is likely to pay higher royalties than an established firm with an active R&D programme. Differences in royalties may reflect the true costs of technology needed to participate in the industry, paid either in cash or in kind.

See Also

- ▶ Licensing
- ▶ Patents
- ▶ Standards

References

Geradin, D., A. Layne-Farrar, and J. Padilla. 2008. The complements problem within standard setting: Assessing the evidence on royalty stacking. *Boston University Journal of Science & Technology Law* 14: 144–176.

Grindley, P., and D. Teece. 1997. Managing intellectual capital: Licensing and cross-licensing in semiconductors and electronics. *California Management Review* 39: 8–41. Reprinted in *Essays in Technology Management and Policy*: *Selected Papers of David J. Teece*. Singapore: World Scientific, 2003.

Lemley, M., and C. Shapiro. 2007. Patent holdup and royalty stacking. *Texas Law Review* 85: 1991–2049.

Shapiro, C. 2001. Navigating the patent thicket: Cross licenses, patent pools, and standard setting. In *Innovation policy and the economy*, vol. 1, ed. A. Jaffe, J. Lerner, and S. Stern. Cambridge, MA: The MIT Press.

Ziedonis, R. 2004. Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms. *Management Science* 50: 804–820.

Cross-Subsidization

David Bardolet
Bocconi University, Milano, Italy

Definition Cross-subsidization describes the tendency of multi-business firms to shift resources (especially cash) generated in their highly performing divisions towards their poorly performing divisions, resulting in a distribution of resources that does not always conform to rationally optimal rules.

Studies of capital investment in multi-business firms have uncovered a pattern of significant overinvestment in poorly performing business units, as compared with capital markets' investment in similar stand-alone businesses, as well as significant underinvestment in highly performing business units. This phenomenon, which seems to rest on the ability of corporations to shift cash across their divisions, has been dubbed cross-subsidization or 'corporate socialism'.

One of the first studies to observe cross-subsidization in the field is Lamont (1997), which found that oil companies cut investments across the board, including investments in their non-oil divisions, after a decrease in oil prices. Interestingly, some of those budget cuts were performed on non-oil divisions that actually benefited (i.e., generated higher cash flows) from the oil price crisis. Other literature has generalized these findings to other industries, despite some methodological caveats (see Stein 2003). In general, these studies find little correlation between a business unit's capital expenditures and diverse measures of that business unit's current performance and future potential. This suggests the existence of 'corporate socialism', by which corporations spread their allocations among all their divisions much further than any 'rational' investment rule would allow.

The fact that multi-business firms deviate from optimal investment rules (e.g., 'invest more capital in the project with the highest future rate of return') does not necessarily imply that such allocations are inefficient. Some authors argue that cross-subsidization is a tool that allows corporate management to support worthy business units that would not receive enough investment outside the corporate umbrella due to uncertainty about their prospects (Stein 2003). Other authors posit that highly performing divisions within a company are

usually more efficient in using their capital and hence require relatively less of it (Maksimovic and Phillips 2002). However, most empirical research on cross-subsidization seems to give greater weight to the argument that corporations might be inefficiently subsidizing poorly performing divisions using cash flows generated in highly performing ones, although the size and scope of this inefficiency remains an open question.

A number of theoretical explanations of cross-subsidization have been proposed. First, cross-subsidization could be a result of divisional managers acting as rent-seeking agents, allocating effort between running their divisions, which tends to enhance firm profit, and lobbying the CEO, which often attracts extra resources to the division at the expense of firm profit (Scharfstein and Stein 2000). Moreover, the CEO herself acts as a rent-seeking agent, using capital allocation as a substitute for other forms of compensation (e.g., salary, perquisites) to divisional managers. Thus, by diverting capital from well-performing divisions (in which managers receive a better return for their management efforts than for their lobbying efforts) to poorly performing divisions (in which managers have a stronger incentive to lobby than manage), the CEO can conserve discretionary funds for more attractive personal uses.

Secondly, cross-subsidization might be a result of an incentive scheme for divisional managers. In this account, the CEO/principal minimizes incentives for divisional rent-seeking by spreading capital across all divisions of the firm, assuming that if there is less competition for resources among them, divisional managers will be more likely to favour strategies that enhance the profits not only of their own division but those of other divisions (Rajan et al. 2000). In a similar fashion, some authors argue that the CEO, lacking private information on the expected value of divisional investments and lacking resources to carefully audit every request for funds, sets an initial compromise allocation that is 'generous' for less promising divisions and 'stingy' for more promising divisions. Managers who are underfunded can then request additional capital by subjecting their projects to a more careful audit. These schemes allow the corporation to strike a balance between optimal allocations and monitoring costs.

Finally, cross-subsidization might be the result of an innate psychological tendency of individuals to spread out allocations over an option set (Bardolet et al. 2011). This tendency can be driven by the implicit use of a specific allocation heuristic, such as the '1/n rule', or by the consideration of emotional and moral factors like equality and fairness, which drive individuals to allocate benefits and burdens relatively evenly among members of a group.

See Also

- ▶ Agency Problems
- ▶ Bounded Rationality
- ▶ Capital Budgeting
- ▶ Conglomerates
- ▶ Corporate Strategy
- ▶ Resource Allocation Theory

References

Bardolet, D., C. Fox, and D. Lovallo. 2011. Naïve diversification and partition dependence in organizational capital allocation: An experimental investigation. *Strategic Management Journal* 32: 1465–1483.

Lamont, O. 1997. Cash flow and investment: Evidence from internal capital markets. *Journal of Finance* 52: 83–109.

Maksimovic, V., and G. Phillips. 2002. Do conglomerate firms allocate resources inefficiently across industries? Theory and evidence. *Journal of Finance* 72: 721–767.

Rajan, R., H. Servaes, and L. Zingales. 2000. The cost of diversity: The diversification discount and inefficient investment. *Journal of Finance* 55: 35–80.

Scharfstein, D.S., and J. Stein. 2000. The dark side of internal capital markets: Divisional rent-seeking and inefficient investment. *Journal of Finance* 55: 2537–2564.

Stein, J. 2003. Agency, information and corporate investment. In *Handbook of the economics of finance*, ed. G.M. Constantinides, M. Harris, and R. Stulz. Amsterdam: North-Holland.

Customer Loyalty Programmes

John Roberts
University of New South Wales, Sydney, Australia

Abstract

To design and evaluate customer loyalty programmes, the firm must ensure that they create value in the mind of the consumer relative to their net cost. Some of that value must accrue to the firm and some to the consumer. Value to the firm can come from increased usage, greater share of wallet, increased average price, longer customer lifetimes, improved cross-selling and access to new customers. To the customer, the programme must offer net benefits in her/his mind relative to the competition.

Definition Customer loyalty programmes are reward systems that firms put in place to encourage consumers to increase their consumption and tenure with the firm, as well as to make the firm attractive to new customers.

Customer loyalty programmes are reward systems that firms put in place to encourage consumers to increase their consumption and tenure with the firm, as well as to make the firm attractive to new customers. The programmes do this by increasing perceived switching costs for the consumer and also by improving the attractiveness of the company's offerings (Wirtz et al. 2007). Considerable controversy surrounds the cost benefits of loyalty programmes. For example, Liu (2007) finds limited evidence of their effectiveness and states that the efficacy of customer loyalty programmes is not well established. In contrast, Meyer-Waarden (2007) finds evidence that loyalty programmes increase consumption.

The Objective of Customer Loyalty Programmes

To work, a loyalty programme has to make sense from the perspective of both the firm that initiates it and the consumers at whom it is targeted. Figure 1 clarifies those dual needs.

The objective of the firm is to create value for the consumer, which can be measured by the consumer's reservation price, R_{ij}, that is, the maximum that s/he is prepared to pay. It attempts to do this at a marginal cost, V_{ij}, which is less than the reservation price. The price, P_j, will fall somewhere between R_{ij} and V_{ij}. $(R_{ij} - P_j)$ is the consumer surplus that accrues to the customer, while $(P_j - V_{ij})$ is the contribution per unit for the firm. If consumer i has a need for I_i units in the category, and the market share of company j in meeting this is M_{ij}, then the sales to i by j will be $S_{ij} = I_i \star M_{ij}$.

The value of the transaction to the company in terms of contribution is described by $C_{ij} = (P_j - V_{ij}) \star I_i \star M_{ij}$. The value to the consumer of dealing with company j, if s/he is behaving myopically, is $U_{ij} = (R_{ij} - P_{ij}) \star I_i \star M_{ij}$. If the consumer is comparing the value of company j's products to those of other companies, then the value to the consumer will be the value company j provides relative to the value from the best alternative company, given by the expression

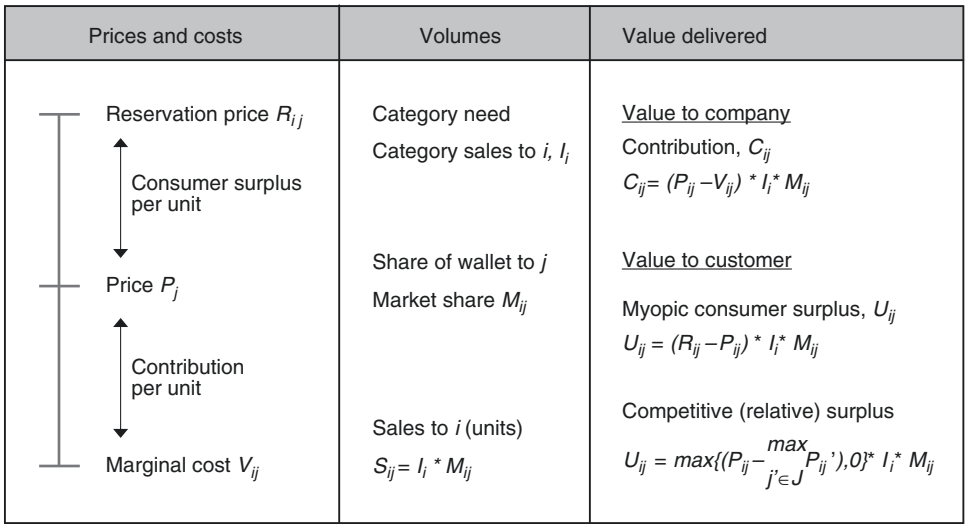
$$U_{ij} = \max\left\{\left(P_{ij} - \max_{j' \in J} P_{ij'}\right), 0\right\}^{*} I_i * M_{ij.}$$

These loyalty programmes will make sense only if the cost to the company in creating the programme is more than compensated for by the value that it creates for the consumer. However, while necessary, that criterion is not sufficient. The company must also be able to capture enough of that value to meet its costs.

As Fig. 1 shows, for a programme to make sense from the perspective of the company, the incremental profit or contribution that it generates must exceed its cost of implementation. Another requirement for the programme to be effective is that it provides net value to the target market for which it is designed.

Evaluating Consumer Loyalty Programmes

To evaluate a specific consumer loyalty programme from the company's perspective, we



Customer Loyalty Programmes, Fig. 1 Loyalty programme from perspective of company *j* and customer

must examine the cost of the programme relative to the incremental benefits it brings in terms of increased category usage, higher share of wallet, ability to further cross-sell and increased expected lifetime value. Unfortunately, these calculations often become considerably more difficult because customer loyalty programmes frequently trigger a competitive response from other companies in the marketplace. While Meyer-Waarden (2007) finds positive profits accruing from customer loyalty programmes (in packaged goods markets), he also determines that where multiple loyalty-card memberships of geographically close retailers exist, they reduce the lifetime duration of the initiative. Indeed, Dowling and Uncles (1997) are highly critical of customer loyalty programmes, calling them 'surprisingly ineffective'. They believe this is the case because the programmes are often targeted at the wrong consumers and simply start an escalating war by reducing average prices.

In sum, we must examine customer loyalty programmes from the company's perspective by balancing any benefits from additional cross sales, increased share, longevity and category usage against the cost of the programme, but this must be done under a number of scenario analyses depending on the reaction of other companies in the marketplace.

From a customer's perspective, the calculation is simpler: we can examine the value of the programme in terms of the degree to which it provides rewards that are salient to him/her. Obviously, ▶ market research is the best way to determine which programmes give the greatest 'bang for the buck'.

When to Use Loyalty Programmes

Wansink and Seed (2001) suggest that loyalty programmes work best in categories that have high margins, that consumers heavily invest in over their lifetime and that are difficult to differentiate. Wansink and Seed also point out that loyalty programmes can be extremely effective in industries in which companies can gain an information advantage of better knowledge of their customers. Loyalty programmes that can provide fine-grained information allow the firm to customize individual offerings to its customer base. For example, in their book *Scoring Points*, Humby et al. (2003) describe how the English supermarket chain Tesco has been able to customize its offerings based on its loyalty programme, thereby increasing sales, improving margins through higher utility and reducing vulnerability.

Principles of Good Loyalty Programmes

Dowling and Uncles (1997) suggest that many, and indeed most, loyalty programmes are not successful. They attribute this to the set of assumptions that managers make about consumers' reactions to programmes, which are often in error. To design a successful loyalty programme, managers need a structured view of what they are trying to achieve and how consumers and the competition will respond. Figure 1 provides us with exactly such a framework. The company's objective of increasing value can come from the last column by improving performance in the second one (growing category needs, increasing the share of wallet that the company attracts or gaining sales over time for a longer period). Increased value can also come from higher prices. While most loyalty programmes involve price reductions rather than price increases, they can provide the information needed to identify consumers for whom the firm can add value with higher-priced, higher-value offerings.

Loyalty programmes that succeed must work for both parties. As Wansink and Seed (2001) suggest, they must not appear to be self-serving – that is, meeting only the objectives of the company without offering the consumer salient benefits. Interestingly, Wansink and Seed find that programmes of only moderate value are most effective, since the firm must get some of the economic gains generated from them.

Programmes that have been found to work extremely well are those that tend to reward those consumers with higher response rates rather than just those with high usage levels. For example, Hilton Hotels found with their HHonors programme that gold members were more profitable than diamond members because the gold members' redemption rates were lower. Usage is not necessarily a good indicator of areas of leverage for increased profitability. Liu (2007) reaches a similar conclusion: customers with initial low patronage levels actually underwent higher improvement in customer lifetime value than those with moderate or high levels.

Loyalty programmes that are consistent with the brand equity of the issuing firm are likely to be more successful than those that are not related to the value proposition. That is, a programme that concentrates on savings might work well for Wal-Mart supermarkets, while a programme that works well on exclusive access to scarce products might work better for Nordstrom. A programme that focuses on donations to environmental causes would be consistent with the position of the Body Shop, while for Axe deodorant, a programme that offers macho video games might be more valuable.

Finally, loyalty programmes that have high positive externalities are likely to be more effective than those for which such externalities do not exist. For example, a category like carbonated beverages is likely to expand with promotional activities; by contrast, purchase acceleration may well occur for laundry detergent, but it is unlikely that programmes will increase usage levels in this category. Other such externalities include word of mouth that loyal users may generate and new applications that could increase category usage. Shugan (2005) suggests that programmes that offer immediate benefits (such as training, which will introduce perceived lock-in costs) are likely to be more effective than those that offer future benefits, such as deferred rewards. This assertion is, however, not necessarily supported by the research of Kivetz et al. (2006), who find that future rewards both accelerate consumption as the consumer moves towards them and improve retention after the consumer has received them.

Trends in Loyalty Programmes

Loyalty programmes are an embedded way of doing business in many industries. However, their nature is changing due to changes in the retailing environment. A number of these changes that managers need to take into account include:

Increased Promiscuity

Retention rates in most industries are down, meaning that loyalty programmes have a tougher job of locking in customers over the long term. This reduces the customer lifetime value of programmes and thus their net effectiveness.

Social Networks

As social networks develop, and tribes and cliques of members (including affinity groups) emerge in the community, rewards can be offered not just to individual consumers, but to groups of consumers. It is quite common to see rewards being offered to, for example, favourite charities, parent and teacher associations, and other groups to which consumers belong. Social networks are particularly in evidence in terms of loyalty programmes with companies such as the Mongolian mobile company, MobiCom, which charges its consumers much more to call consumers out of MobiCom's network than to call those in it, thus imposing an implied cost of defection.

Improved Information and Data-Mining

As firms get better at understanding consumer reactions to different offerings and mass customization of their service levels, the benefits of loyalty programmes increase because they can be cost-effectively tailored to individual groups or small segments of customers.

Business-to-Business Value Chains

Most of the discussion so far has looked at loyalty marketing to final consumers, but business-to-business loyalty programmes are also extremely important. We have witnessed a great deal of third-party transactions where a firm may piggyback on the loyalty programme of a strong issuer. For example, most airlines sell their frequent flyer points to third parties such as florists, other retailers and hotel chains. These loyalty points can then be redeemed with the airline. At one point, the value of outstanding frequent flyer points at United Airlines was three times the firm's market capitalization.

Summary

Customer loyalty programmes are ways of providing tangible and psychological rewards to customers for doing business with the firm. Their effectiveness depends entirely on the value set of the customer and the reactions of ▶ competitors. When designing such programmes, it is therefore essential to understand the objectives of the firm and the value that the programmes create in consumers' minds relative to the value that can be offered by other firms.

See Also

- ▶ Competitors
- ▶ Expected Utility
- ▶ Market Research
- ▶ Price Discrimination
- ▶ Prisoner's Dilemma

References

Dowling, G.R., and M. Uncles. 1997. Do customer loyalty programmes really work? *Sloan Management Review* 38: 71–82.

Humby, C., T. Hunt, and T. Phillips. 2003. *Scoring points*. London: Kogan Page.

Kivetz, R., O. Urminsky, and Y. Zheng. 2006. The goal-gradient hypotheses resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research* 43: 39–58.

Liu, Y. 2007. The long-term impact of loyalty programmes on consumer purchase behaviour and loyalty. *Journal of Marketing* 71: 19–35.

Meyer-Waarden, L. 2007. The effects of loyalty programmes on customer lifetime duration and share of wallet. *Journal of Retailing* 2: 223–236.

Shugan, S.M. 2005. Brand loyalty programmes: Are they shams? *Marketing Science* 24: 185–193.

Wansink, B., and S. Seed. 2001. Making a brand loyalty programmes succeed. *Brand Management* 8: 221–222.

Wirtz, J., A.S. Mattila, and M.O. Lwin. 2007. How effective are loyalty reward programmes in driving share of wallet? *Journal of Service Research* 9: 327–333.

Cyert, Richard M. (1921–1998)

Mie Augier
GSBPP, Naval Postgraduate School, Monterey, CA, USA

Abstract

Richard Cyert is perhaps most well known (among organization scholars and strategic management scholars) for his co-authorship with James March of the book *A Behavioral Theory of the Firm*. But he also made a number

of significant contributions to economics and statistics, and was for many years President of Carnegie Mellon University.

Richard Michael Cyert was born on 22 July 1921 in Winona, Michigan. He was educated at the University of Minnesota and graduated in economics in 1943. After serving as officer in the US Navy for 3 years during the Second World War, Cyert entered graduate school at Columbia University, New York, in 1946. He completed his degree in 1951. He received many rewards and honorary memberships for his works and was, until his death, President Emeritus and R.M. and M.S. Cyert Professor of Economics and Management at Carnegie Mellon University. He is well known for his contribution to behavioural economics (in particular the ▶ behavioural theory of the firm), business administration, Bayesian economics, and for his leadership of Carnegie Mellon University.

Cyert worked extensively within the fields of behavioural economics, economics in general, decision theory and management. Some of his most important books are *A Behavioral Theory of the Firm* written with ▶ James March (1963) and *Baysian Analysis and Uncertainty in Economic Theory* with DeGroot (1987). Professor Cyert went to Carnegie Mellon University (then Carnegie Institute of Technology) in 1948 where he later became dean of the Graduate School of Industrial Administration (GSIA) (1962–1972) and president (1972–1990) of the university. Even when taking on his leadership duties as a dean and president, he remained very active in publishing and research. First an instructor of economics, he went on to become assistant professor of economics and industrial administration, associate professor and head of the department of industrial management, professor and dean of GSIA and president of Carnegie Mellon University. During his life, he initiated, contributed to and maintained a keen interest in behavioural economics. In this article a general introduction to some of Cyert's works is presented.

Richard Cyert belonged to that small but select group of economists/political scientists/organization theorists who helped initiate and develop the behavioural economic and behavioural organization theory programme in America in the 1950s and 1960s. He is well known for being one of the three 'founding fathers' (along with James G. March and Herbert Simon) of the behavioural economics initiated at Carnegie Mellon University in this period. Cyert's interest in behavioural economics started with his doctoral thesis on price-setting in oligopolistic markets. Cyert found that neoclassical theory gave him very little support as a prescription for description of managerial and firm behaviour. As a result, Cyert and his colleagues at Carnegie Mellon University laid the research foundations for a series of contributions to organizational and behavioural approaches to economics and management (Augier and March 2002).

It was during his years as a doctoral student that Cyert realized that for economics to go anywhere it had to begin collaborating with other disciplines, such as organization theory, sociology, management and psychology. This interdisciplinary view was stimulated, encouraged and developed at Carnegie Mellon University during his interaction and collaboration with Herbert Simon and James G. March. This is especially clear in his work developed in order to go inside the black box of the neoclassical theory of the firm and understand the internal decision-making processes of the firm, especially in his work with James G. March. Their first co-authored paper, 'Organizational structure and pricing behavior in an oligopolistic market' was published in the *American Economic Review* in 1955, and, about 7 years later, they completed *A Behavioral Theory of the Firm*. In this book Cyert and March outlined a theory that was built around a political conception of organizational goals, a bounded rationality conception of expectations, an adaptive conception of rules and aspirations, and a set of ideas about how the interactions among these factors affect decision-making in firms (Augier and March 2002). They emphasized the idea of problemistic search: the idea that search within a firm is stimulated mostly by problems and directed to solving those problems, and the idea of organizational learning. In *A Behavioral Theory of the Firm*, organizations learn from their own experiences and the experiences of others.

When *A Behavioral Theory of the Firm* was published, Cyert became the dean of the GSIA, but he continued (as did March and Simon) working within behavioural economics.

Cyert also collaborated with Robert Trueblood and Morris DeGroot. With Trueblood he worked on statistical sampling methods and statistical decision theory (Trueblood and Cyert 1957). And in the late 1960s, Cyert began working with Morris DeGroot, who was trained in Bayesian statistics. They published their first paper in 1970 and their book was published in 1987 (Cyert and DeGroot 1987). Most people would probably not be inclined to equate 'Bayesian' economics with 'behavioural' or 'managerial' economics or strategy. But in an important way Cyert's approach to Bayesian economics was *both* a natural outgrowth of his work with March on behavioural economics *and* a contribution to behavioural economics itself (Day and Sunder 1996; Augier and March 2002). The argument that Cyert's work on Bayesian economics can be seen as a contribution to behavioural economics is twofold. First, in doing this kind of work, Cyert was interested in building a theory of *real economic behaviour* by taking *uncertainty* into account (Cyert 1970). This move is consonant with modern behavioural emphasis on the uncertainty and behavioural aspects of economics. Second, he built on his work by March and the idea of organizational learning. An important example of how learning can contribute to the development of Bayesian economics is found in Cyert's work on adaptive utility. Noticing the observable difference between the assumed fixed utility of decision-making to the observed choices, Cyert wanted to apply the concept of learning to the concept of utility in such a way that changes in utility functions over time (as a result of learning) could be accounted for. This intertemporal aspect of learning is clearly a behavioural idea.

The behavioural theory of the firm has become an important intellectual foundation for several traditions in strategic management, such as the ▶ dynamic capabilities perspective and perspectives including learning and adaptation; much of this traces back to Cyert's (and March's) work.

C

See Also

▶ Behavioural Strategy
▶ March, James G. (Born 1928)

References

Augier, M., and J.G. March. 2002. Richard M. Cyert: The work and the legacy. In *The economics of choice, change and organization: Essays in memory of Richard M. Cyert*, ed. M. Augier and J.G. March. Cheltenham: Edward Elgar.

Day, R., and S. Sunder. 1996. Ideas and work of Richard M. Cyert. *Journal of Economic Behavior & Organization* 31: 139–148.

Trueblood, R.M., and R.M. Cyert. 1957. *Sampling techniques in accounting*. Englewood Cliffs: Prentice Hall.

Selected Works

1955 (with J. G. March). Organizational structure and pricing behavior in an oligopolistic market. *American Economic Review* 45, 129–139.

1958 (with W. R. Dill and J. G. March). The role of expectations in business decision making. *Administrative Science Quarterly* 3: 307–340.

1963 (with J. G. March). *A behavioral theory of the firm*. Englewood Cliffs: Prentice Hall.

1970. Implications for economic theory of a behavioral approach to the firm. In *Behavioral approaches to modern management*, ed. W. Goldberg. Göteborg: Företagsekonomiska institutionen vid Handelshögskolan.

1970 (with M. H. DeGroot). Baysian analysis and duopoly theory. *Journal of Political Economy* 78: 1168–1184.

1987 (with M. H. DeGroot). *Bayesian analysis and uncertainty in economic theory*. Totowa: Rowman & Littlefield.

1990. ed. *The management of non-profit organizations*. Lanham: Lexington Books.