
When considering questions related to outsourcing and insourcing, it is important to be conversant in the vast Theory of the Firm (ToF) literature, which dates at least from Coase (1937). Operations research and operations management initially took the boundaries of the firm as given and focused on tactical issues such as how to organize for efficiency (Taylor 1911) or how much inventory to carry (Whitin 1955). In contrast, economics and business strategy have long focused on higher-level questions of organization, under the names of vertical integration, vertical coordination, firm boundaries, make-versus-buy (or make-or-buy or make-buy), and outsourcing/insourcing. The ToF literature originally focused on defining a firm and explaining a firm’s existence, given the benefits of the market (Coase 1937). This inquiry spawned much theoretical and empirical work on where firm boundaries should be drawn; i.e., which activities should be performed inside the firm and which should be performed outside the firm. This section briefly discusses key approaches to this topic. We organize this as follows: First, we review theories based on outsourcing’s effect on incentives. The dominant theory here is transaction cost economics (TCE). Second, we review theories related to capabilities, with the dominant theory being the resource-based view (RBV). Finally, we briefly touch upon the real options perspective, which has received much less attention in the ToF literature.

1.1. Incentive-Based Theories and Perspectives

We first review a set of theories that focus on organizing to achieve cooperation among parties to a transaction.

1.1.1. Transaction Cost Economics (TCE)
The development of the massive stream of literature comprising transaction cost economics (TCE) has resulted in at least two Nobel Prizes in economics: Ronald H. Coase (awarded in 1991) and Oliver E. Williamson (awarded in 2009). Many publications in economics, strategy, and beyond draw from TCE, refining it and empirically testing its propositions (Macher and Richman 2008).

The basic idea of this theory is that coordinating a transaction between a buyer and a seller using the market mechanism results in both ex-ante and ex-post transaction costs. Activities generating these costs include searching for and selecting a business partner, negotiating on price and other terms, writing contracts (which will almost always be incomplete), monitoring and enforcing contractual compliance, and renegotiating contracts when unforeseen circumstances arise. When these costs become large, performing the activities within a single entity, thereby managing by fiat instead of contract, may be comparatively more efficient. The risk of appropriability or value capture by others due to, for example, the loss of intellectual property (Pisano 1990; Oxley 1997; Gulati and Singh 1998), is a related concern. While contracts can specify what is not all owed, violations still need to be enforced in court, which carries potentially high ex-post transaction costs. Williamson (1971, p.114) states that: “fiat is frequently a more efficient way to settle minor conflicts … than is haggling or litigation.” Williamson (1979, p.253) also notes that: “The advantage of vertical integration is that adaptations can be made … without the need to consult, complete, or revise interfirm agreements.” Put directly, given the “business judgment rule” in contract law, courts exercise forbearance in which corporate-level managers serve as a “court of appeal” for firms’ divisional-level conflicts, which thereby mitigates costly interfirm renegotiation or litigation (Williamson 1991).

TCE rests on two key assumptions about decision makers. One is that these decision makers exhibit bounded rationality, which Herbert Simon defined as “The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world” (1957, p.198, emphasis in original). Such decision makers
are consequently unable to stipulate in the formal contract all actions for all possible future contingencies. The second key assumption is that some managers may behave opportunistically. This means acting in their own self-interest, “with guile” (Williamson 1975, p.255) by playing outside the “rules of the game.” Given these assumptions, contingencies can arise in an exchange where one party may harm the other through post-contractual opportunistic behavior such as the economic holdup problem of renegotiating contract terms to capture greater economic value once the other party has become locked in to the exchange relationship (Klein et al. 1978; Williamson 1979).

TCE goes further to define characteristics of the exchange that increase the likelihood and severity of when such opportunistic behavior might arise, resulting in increased costs. The unit of analysis is the transaction. Different transactions, depending on key characteristics discussed below, should be governed differently, on a continuum from market (i.e., arms-length/transactional) to hierarchy (i.e., in-house). The “discriminating alignment hypothesis” (Williamson 1996) is that performance will be better when governance choice better aligns with transaction characteristics.

The most robust explanatory variable is asset specificity. Williamson (1985, p.56) submits that: "asset specificity is the big locomotive to which transaction cost economics owes much of its predictive content." Considering the level of joint investment in human or physical resources, the level of asset specificity is the degree to which such investments have lower economic value when used outside the context of the specific exchange relationship; such investments may also be called idiosyncratic to the transaction (Williamson 1979). High asset specificity in the buyer-supplier relationship has correspondingly high transactional hazards due to potential opportunistic behavior. Specifically, when assets are costly to redeploy then the appropriable economic quasi-rent (i.e., the difference between the first best and second best use value of the asset) may be substantial. Because of this, one of the exchange parties may try to take advantage and renegotiate the contract to appropriate part, if not all, of this economic quasi-rent (Klein et al. 1978; Williamson 1979). Asset
specificity may be low at the onset of a bilateral transaction, but this relationship may transform over time as the two exchange parties learn how to work with each other. As Williamson (1975, p.29) put it: “Although a large-numbers exchange condition obtains at the outset, it is transformed during contract execution into a small-numbers exchange”; often referred to as the “fundamental transformation” (Williamson 1985, p.61-3). Even without asset specificity, the time for a buyer to switch suppliers or a supplier to find new buyers can pose a challenge, which is called “temporal specificity” by Masten et al. (1991, p.9). For example, high perishability for goods (such as fruits and vegetables) can lead to economic holdup problems. If vertical coordination does not take place quickly then the economic value of these perishable goods diminishes and exchange partners can appropriate economic rents. Such transactional hazards often necessitate vertical integration as an economic safeguard to avoid such problems of “temporal specificity” (Chandler 1977; Masten et al. 1991; Bucheli et al. 2010).

Any increase in uncertainty that leads to challenges in incomplete contracting is also relevant, as this creates the need for “unprogrammed adaptation” (Williamson 1971, p.113). Per TCE, asset specificity and uncertainty together lead to situations where prior agreements need to be renegotiated, creating an opportunity for one party to behave opportunistically, such as by engaging in economic holdup of exchange partners (Macher and Richman 2008), which may motivate vertical integration as an economic safeguard (Williamson 1985). The POM literature has shown that, in the presence of demand uncertainty and (implied) low asset specificity, outsourcing to a common supplier or vendor can be beneficial for the risk-pooling benefits (Chaturvedi and Martínez-de-Albéniz 2016); this scenario was also noted by Williamson (1985). Further, the real options lens, discussed in Section 2.3 below, indicates that under technological uncertainty firms will be less inclined to invest in internal non-redeployable assets for fear of technological obsolescence; this is another case where uncertainty leads endogenously (i.e., strategically) to low asset specificity, and thus may favor outsourcing
Managers must be aware of all these types of uncertainties, and to capture this point, Weber and Mayer (2014) join bounded rationality and uncertainty constructs by noting the importance of “interpretative uncertainty” by decision makers.

Frequency of the transaction is the third, but by far least discussed, transaction characteristic of the theory. Williamson (1985, p.79) divides frequency into two buckets: occasional and recurrent. We submit that frequency has largely been neglected in much literature because researchers are analyzing or considering recurrent transactions (e.g., an on-going buyer-supplier relationship); this is the case in the types of transactions examined in most operations and supply chain settings. Indeed, Gibbons’ (2010) review of transaction costs, as they apply to vertical integration, states that: “frequency … is beyond the scope of this subsection’s focus on vertical integration and contracting” (p.273). For non-asset-specific transactions, frequency does not affect the appropriate governance choice (market). In the case of high asset specificity, Williamson (1985, p.60) submits that recurrent transactions are more likely to be performed in-house because “the costs of specialized governance will be easier to recover for large transactions of the recurring kind.” In addition, “vertical integration does not offer advantage over a contract for a one-time exchange” (Lajili et al. 2007, p.347-8). But, recurring transactions can also make reputation effects more pronounced, reducing the likelihood of opportunistic behavior (Klein and Leffler 1981); “It is perhaps because of these competing effects that researchers have been largely unable to confirm (or refute) the effects of transactional frequency on governance modes” (Macher and Richman 2008, p.7).

1.1.2. Incomplete Contracts/Property Rights Theory

The more formalized incomplete contracting/property rights theory initiated by Grossman and Hart (1986) and Hart and Moore (1990) – the so-called GHM models – (see also Hart 1995 and Tirole 1999) emphasizes that ownership matters. From this perspective, ownership is based on the (ex-post) residual rights of control in the case of missing contractual provisions. Differences between the
market and vertical integration are entirely ascribed to the differences in asset ownership that distinguish these alternative governance modes. The key commonalities of this formalized theory and TCE (Williamson 1985) are that governance modes are evaluated comparatively, and that in each theory higher levels of asset specificity favor vertical integration. The key difference is that the GHM models focus exclusively on *ex ante* incentives to invest and neglect *ex post* negotiating costs and governance inefficiencies (Whinston 2001; Williamson 2002). For example, the GHM models: ignore incentive distortions and potential bureaucratic failures that occur in firms; explicitly deny that internal audits in the vertically integrated firm differ in any way from external audits in market organization; deny the adaptability advantages of fiat; disregard the potential bureaucratic failures of internal organization; and posit that third-party enforcement by courts is perfectly efficacious. In general, analytical modeling – such as the GHM models – provides an “audit trail” in terms of the necessary and sufficient conditions leading to conclusions. However, both researchers and practitioners who base their recommendations and decisions on such modeling need to be sensitive to the losses in translation from verbal arguments to formal models, and the possibility that simplifying assumptions necessary for analytical tractability may lead to prescriptions that miss important information. As will be discussed later, this is a tradeoff well known to POM scholars. A key difference between POM and the ToF literature is that POM’s history is heavily analytical whereas the ToF’s history is largely conceptual and empirical.

1.1.3. Measurement Perspective

Arguably a subset of TCE, the measurement perspective focuses not on the level of asset specificity but rather on the ability of the buyer to evaluate all-important aspects of the output of the supplier. Alchian and Demsetz (1972) suggest how and why observation of the joint output from the supplying party does not enable the buying party to infer individual productivity. Their example is “team production,” such as loading a piece of heavy furniture onto a truck, wherein each individual’s
contribution would be difficult to discern by merely observing the output, and thus requires close monitoring of each individual's behavior (effort). The question then becomes: who monitors the monitor? Alchian and Demsetz' (1972) solution is that the monitor becomes the residual claimant in the firm. Note that instead of ownership being defined as ex-post residual control rights as described in the GHM models above, ownership is now defined in terms of ex-ante residual income rights or residual claimancy. To the extent that the firm has superior monitoring capabilities vis-à-vis external monitoring, this perspective supports internal organization in the context of team production. Barzel (1982) provides a similar measurement argument, which focuses on measuring all important aspects of quality, stating succinctly (p.42): “Distinct firms will form and trade with each other at junctures where output can be readily measured, but where output is difficult to measure the different steps will be performed within the firm.” Different governance modes may differentially attenuate this value-capture problem. Thus for both Alchian and Demsetz’ (1972) measurement problem in team production and Barzel’s (1982) quality measurement problem, governance mode choice and measurement costs are interdependent.

Agency theory (e.g., Holmstrom 1979; Holmstrom and Milgrom 1991) is not, and was not intended to be, a theory of firm boundaries per se. That said, some insights from this literature apply to our discussion. For example, Jensen and Meckling (1976) note that minimizing agency costs involves minimizing the sum of (1) the monitoring costs incurred by the principal, (2) the economic bonding costs incurred by the agent, and (3) the residual loss (the latter being an expansive category). We note the interdependence between monitoring (measurement) costs and bonding (transaction) costs. Consider a franchise contract in which the principal is the franchisor and the agent is the franchisee. To protect the value of the franchise system’s brand name the franchisor may be required to monitor the franchisee extensively. Suppose the franchisor requires the franchisee to post an economic bond or economic hostage in the form of the franchisee making a franchise-specific
investment that loses value upon franchise termination (Williamson 1985). This credible commitment by the franchisee to the franchisor gives the franchisee incentive to maintain high quality, which then lowers the monitoring costs that need be incurred by the franchisor. The introduction of these economic hostages produces efficiencies in the franchise system independent of who initiates the proposal. Here, as elsewhere, it is useful to consider contracting in its entirety (Williamson 1985), and a comparative assessment of imperfect governance form alternatives to align with the economic problem at hand is applicable.

1.2. Resource-, Knowledge-, and Capability-Based Approaches

In contrast to theories that focus on incentive alignment to achieve cooperation between parties in an exchange, several “views” focus on the benefits of coordination of internal activities vs. outsourced activities (Conner 1991; Conner and Prahalad 1996). As noted by Mayer and Salomon (2006, p.942): “because transactions cost economics fundamentally concerns characteristics of exchange, its logic typically holds firm capability constant.” Clearly, however, relative capabilities matter. Indeed, Williamson (1999, p.1103) explicitly discusses how TCE and capabilities can complement each other: “Rather, therefore, than ask the question ‘What is the best generic mode (market, hybrid, firm, or bureau) to organize X?’ which is the traditional transaction cost query, the question to be put instead is “How should firm A—which has pre-existing strengths and weaknesses (core competencies and disabilities)—organize X?”

1.2.1. Resource-Based View (RBV)

The resource-based view also has a long history, formally dating at least back to Edith Penrose’s (1959) _The Theory of the Growth of the Firm_ and even David Ricardo (1817) (some of the key concepts appear in Adam Smith (1776)). Penrose (1959) emphasized the internal resources of the firm as the drivers of, or impediments to, its growth. This approach contrasted with the literature at the time, which focused not on growth but on the optimal size of the firm, and more on factors external
to the firm, such as industry position. A recent article in *Production and Operations Management* (Kor et al. 2016) summarized the key ideas of Penrose (1959), which highlighted the importance of firm-specific experience and the coordination challenges created by growth. Penrose (1959) emphasized that with experience comes improvement, likely freeing up resources to allow growth. The “Penrose effect” or “Penrose theorem” posits that: “[m]anagerial capability is the binding constraint that limits the growth of the firm” (Kor et al. 2016, p.1732).

More recent seminal works are attributed to Wernerfelt (1984) and to Barney (1991). Barney (1991) is most often credited for defining the characteristics of resources that lead to sustainable competitive advantage: valuable, rare, inimitable, and non-substitutable (VRIN).

Whereas transaction cost and measurement theories discussed above delineate conditions under which a firm should outsource, works under the umbrella term of the RBV focus more on economic rents (for this review we consider economic rents to mean positive NPV and sustainable competitive advantage). However, scholars have applied the RBV to firm boundaries in various ways. One is to conclude that firms should keep in-house those resources that are VRIN and outsource the rest (Santos and Eisenhardt 2005). Corollaries to this are that resources that are co-specialized to those that are VRIN must also be kept in-house (Teece 1986; Mahoney and Pandian 1992), and that certain in-house activities may need to be kept in-house to develop the VRIN resources of the future that take advantage of inter-temporal and inter-project spillovers (Kang et al. 2009).

The above discussion suggests that RBV considerations applied to outsourcing require precision in the definition of a resource, and how the outsourcing decision relates to it. Research has shown that both resource and TCE-related considerations matter in governance choice (Argyres 1996), and that existing resources in a given activity (operationalized as production experience) make firms more likely to internalize that activity (Leiblein and Miller 2003).
Further, Mayer and Salomon (2006) show that under situations where economic holdup is a concern, strength in what they call "governance capabilities" can favor outsourcing because these capabilities can reduce the holdup hazards. Mayer and Salomon (2006, p.956) conclude that: “governance capabilities (a potentially valuable, rare, inimitable, and non-substitutable firm-specific capability) may be central to governance decisions.” Such capabilities are similar to the “alliance capability” in the relational view of Dyer and Singh (1998) and Kale et al. (2002). Both imply that choices to outsource when holdup risks are high may not be “mistakes” (Mayer and Salomon 2006, p.956). In the book The Machine that Changed the World that documented the Toyota Production System, Womack et al. (1990, p.127) emphasize the importance of governance capabilities: “The make-or-buy decision that occasioned so much debate in mass production firms struck Ohno and others at Toyota as largely irrelevant as they began to consider obtaining components for cars and trucks. The real question was how the assembler and the supplier could work together smoothly to reduce costs and improve quality, whatever formal, legal relationship they might have.”

1.2.2. Knowledge-Based View (KBV)

The KBV goes beyond simply maintaining that knowledge is the critical VRIN resource that firms can develop. As a theory of the firm, the KBV (Kogut and Zander 1992; Grant 1996a, 1996b) focuses on the ease with which knowledge can be developed and transmitted. A well-established classification describes knowledge as either codified or tacit (Polanyi 1962; Nonaka 1994). The KBV primarily concerns itself with the transmission and use of tacit knowledge, and the relative ease with which this knowledge can be developed and shared within and between firms. Kogut and Zander (1992, p.384) note that: “Firms exist because they provide a social community of voluntaristic action structured by organizing principles that are not reducible to individuals.” To the extent that a technical dialogue (Monteverde 1995) develops more effectively within a firm than without, a firm boundary
can inhibit inter-firm flow of knowledge. Concerns with intellectual property protection may also inhibit the free flow of knowledge between firms relative to within firms (Teece 1986; Ziedonis 2004).

The KBV identifies characteristics of a transaction that drive whether an activity should be performed in-house or outsourced. When a transaction depends on the exchange of tacit knowledge, this view recommends that (all else equal) the activity be performed in-house. This recommendation is often consistent with that of TCE; that is, transactions requiring the exchange of tacit knowledge also typically would involve high levels of (human) asset specificity. However, the KBV is focused on coordination (vis-à-vis incentive) benefits of internal organization. That said, disentangling pure coordination from pure incentive motives can be elusive. Indeed, Foss (1996) and Mahoney (2001) independently noted that the importance of improvements in language and exchange are enhanced even further when opportunism is taken into account. As is often the case, even these subtle points have been considered by Oliver Williamson (e.g., Williamson 1975, p.25): “A further advantage of internal organization is that, as compared to recurrent market exchange, efficient codes [of communication] are more apt to evolve and be employed with confidence by the parties. Such coding also economizes on bounded rationality. Complex events are summarized in an informal way by using what might be an idiosyncratic language. Although, in principle, the parties to recurrent market contracts could devise the same language, thereby realizing the same economies, such exchanges are more subject to risks of opportunism – hence, are less apt to be developed as fully.” The standardization of language to which Williamson (1975) refers may take the form of accounting systems, blueprints, and other reporting systems (Nelson and Winter 1982; Mahoney 2001).

1.2.3. Dynamic Capabilities

A criticism of the above perspectives is that they are static. Another theory of the firm focuses on the path-dependent nature of competitive advantage. This “dynamic capabilities” approach focuses not on the current state of resources or knowledge, but rather on how adept firms are at adaptation
and their capability of coping with change. Dynamic capabilities are defined as “a firm’s ability to integrate, build, and reconfigure internal and external competences” (Teece et al. 1997, p.516), which can be sources of persistent performance differences among firms in rapidly changing environments. Williamson (1996, p.227) states: “One way to unpack the ‘capabilities’ view of the firm is to ask what – in addition to an inventory of physical assets, and a census of its workforce – is needed to describe the capabilities of the firm. Features of organization that are arguably important include the following: (1) the communication codes that the firm has developed (Arrow 1974); (2) the routines that it employs (Cyert and March 1963; Nelson and Winter 1982); and (3) the corporate culture that has taken shape (Kreps 1990).”

Zollo and Winter (2002) join the organizational learning literature with the dynamic capabilities literature. Going beyond the view of capabilities as (tacit) routines, deliberate learning mechanisms such as explicit knowledge articulation and codification activities are emphasized as complementary means through which firms build their capabilities.

1.2.4. Problem-Solving Perspective

Another perspective that prescribes when activities should be performed in-house or outsourced is the “problem-solving perspective” (Nickerson et al. 2012). Here the key consideration is the nature of the problem to be solved. Nickerson and Zenger (2004) explain and predict how knowledge sets can be organized to efficiently search for and create new knowledge, which mitigates knowledge-formation and knowledge-transfer problems and generates the following predictions: non-decomposable problems (Simon 1962) are assigned to consensus-based teams; nearly decomposable problems are assigned to authority-based teams; and decomposable problems are assigned to the market (e.g., outsourcing). Macher (2006) empirically corroborates this perspective. The POM literature (e.g., Fine 2000) has considered the idea that supply chain architecture (especially the degree of vertical integration vs. outsourcing) correlates with product architecture (modular vs. integral, cf.
that dictates the decomposability of the product design and process management tasks, which are usually heavily knowledge-based.

1.3. Real Options Perspective

The real options lens is much less frequently employed in the ToF literature than TCE and RBV.¹ A recent piece on real options theory (ROT) (Trigeorgis and Reuer 2017, p.57) wondered about the “interplay between ROT and other, more established perspectives” and specifically how it can “better connect to and be integrated with other theories in strategy.” ROT explicitly notes that outsourcing decisions are not only motivated by minimizing costs but also can create transactional value (Zajac and Olsen 1993). That is, there is “governance inseparability” (Argyres and Liebeskind 1999), wherein the governance choice for one transaction may enable or constrain the governance choice for other transactions. A real options lens can encourage outsourcing or insourcing.

“Growth options” are real options that offer the right to further develop an asset, and to make follow-on investments. As Leiblein (2003, p.949) articulated, “[g]rowth options are particularly valuable in high-technology industries where there are often weak appropriability regimes and inter-generational knowledge spillovers are significant. In these contexts, it will often be desirable to internalize activities associated with an early generation of a product.” Moreover, Kang et al. (2009) note that governance choice can be influenced by growth options in the form of inter-temporal business with the current exchange partner and of inter-project spillovers in gaining business with other companies.

In addition to growth options, flexibility options (Leiblein 2003; Sanchez and Mahoney 1996) imply that “under uncertainty, it may be optimal to utilize market like mechanisms that provide greater flexibility” (Leiblein 2003, p.949). As an example, Kouvelis et al. (2001) modelled and empirically

¹ For example, a Google Scholar search on December 1, 2017, of “theory of the firm” and 1) “transaction cost” had 30,100 results; 2) “resource based” had 39,800 results; and 3) “real options” had 4,810 results.
showed that macroeconomic volatility of a foreign market motivates multinational firms to employ flexible entry modes (e.g., a joint venture over a wholly-owned subsidiary). By ROT logic, unlike behavioral uncertainty discussed earlier in the context of TCE, technological uncertainty about the obsolescence of specific assets makes outsourcing a more likely governance choice (Balakrishnan and Wernerfelt 1986; Lajili et al. 2007). Further, hybrids such as joint ventures may be viewed as real options, which provide an opportunity but not an obligation to acquire (and vertically integrate) as uncertainty is resolved over time (Kogut 1991).

ROT is a growing, but less mature, part of the ToF literature reviewed in this section. The precise conditions leading to a prescribed organizational form are not as clearly established as in the more dominant theories reviewed in the prior sections.
REFERENCES


