Transaction Cost Economics As a Theory of Supply Chain Efficiency

Mikko Ketokivi
Operations & Technology, IE Business School - IE University, Maria de Molina 12-5, 28006, Madrid, Spain, mikko.ketokivi@ie.edu

Joseph T. Mahoney*
Strategy, Entrepreneurship & International Business, Caterpillar Chair of Business, College of Business, University of Illinois at Urbana-Champaign, 140C Wohlers Hall, 1206 South Sixth Street, Champaign, Illinois 61820, USA, josephm@illinois.edu

Transaction cost economics (TCE) is one of the most widely referenced organization theories in operations and supply chain management research. Even though TCE is a broadly applicable theory of governance, one of its specific topics of interest—the make-or-buy decision—readily aligns with some of the central research questions on how firms manage supply chains. However, both general management and operations management researchers sometimes misunderstand and misapply TCE’s aims, assumptions, and logic. A common mistake is to read TCE as a theory of competence or of power. While TCE relates to both, TCE is essentially a theory of efficient governance of transactions in particular and exchange relationships in general. Our purpose in this study is to review the intellectual and theoretical foundations of TCE, its primary aims, and its applicability as a theory of supply chain efficiency. To this end, we discover much common ground between TCE and research in operations and supply chain management. We close by discussing implications for future research, focusing on how operations and supply chain management researchers could contribute to broader academic conversations on management and governance.

Key words: transaction cost economics; supply chain efficiency; governance; make-or-buy decision; vertical integration

History: Received: January 2019; Accepted: November 2019 by Kalyan Singhal, 3 revisions.

1. Introduction

The common interest in transaction cost economics (TCE) and supply chain management is the focus on “the interaction between economic entities” (Zipkin 2012, p. 465). While TCE constitutes a general theory of the governance of exchange relationships and economic organization, it directs particular attention to the make-or-buy decision, sometimes described as “the canonical transaction” (Williamson 1996b, p. 41). This focus makes TCE readily applicable and relevant to research on operations and supply chain management. Indeed, not only is TCE one of the most cited and applied organization theories in operations and supply chain management research (Grover and Malhotra 2003), its potential in informing future research is also recognized (Anand and Gray 2017).

In this study, we have two main objectives. One is prompted by the fact that both the general management literature and the literature on supply chains manifest at times serious misunderstandings of TCE. We seek to clarify these misunderstandings by examining TCE’s key concepts, assumptions, and theoretical logic. Our second objective is to examine not only how supply chain management scholars can apply TCE in future research, but also how to do so in a way that contributes to broader, cross-disciplinary conversations. To this end, we are inspired by Adam Smith’s ([1776] 1976) thesis that the wealth of nations results not merely from the benefits of specialization, but specifically from the combination of specialization and trade. Extending Smith’s ([1776] 1976) logic, we submit that an accumulation of “theoretical wealth” requires a similar combination: researchers not only work in their own specialized fields, but also exchange ideas with specialists from other fields. Even though operations and supply chain management researchers have been “net importers” of theoretical ideas from other management fields (Amundson 1998), they have much potential to contribute to broader conversations as well. As Williamson noted with regard to the collaboration of TCE and operations and supply chain management research, “there should be give and take” (2008, p. 5). Our goal is to examine what this give and take might entail.

The remainder of this study is structured as follows. We first distinguish three theoretical lenses to supply chains: competence, power, and efficiency, which provide the context for TCE vis-à-vis other
theories. We then examine the logic of TCE as a theory of supply chain efficiency by using automobile assembly as an empirical illustration. Automobile assembly is a particularly useful example not only because it has been the focus of a great deal of TCE literature (e.g., Monteverde and Teece 1982, Walker and Weber 1984, 1987), but also because it is a context with which operations and supply chain scholars are readily familiar. We close by examining directions for future research.

2. Theorizing the Supply Chain: Three Theoretical Lenses

Consider the decision of designing the first-tier supply chain of components and subsystems for final assembly. Which parts will the final assembler produce in-house, and which will it outsource to external suppliers? Which aspects of the transaction explain and predict the likelihood of outsourcing a given component? Is the decision explained by attributes of the product itself? Production technology? Capabilities, identities, and histories of particular supplier candidates? Once an exchange relationship with a supplier is established, how does it develop over time? When will existing suppliers have an advantage at a contract renegotiation stage? When are supplier switching costs likely to be higher?

These questions are not only established and well researched in supply chain management (Tsay et al. 2018), but they are also foundational questions in TCE (Williamson 1971). While TCE is broadly applicable “to any issue that can be posed directly or indirectly as a contracting problem” (Williamson 1994, p. 86), the make-or-buy decision in particular—when to internalize a transaction via vertical integration—is the most established research question within TCE (Williamson 1996b, p. 41). Fortuitously, the make-or-buy decision is the most relevant governance question in supply chain management research as well.

The starting point for a TCE analysis of economic exchange is the realization that economic transactions between a buyer and a supplier can occur both within firms (insourcing from internal suppliers) and across firms (outsourcing from external suppliers). The central objective in TCE is to determine which of the available options for governing the transaction is most appealing from an efficiency perspective. The decision is often more nuanced than a binary “make” or “buy” (e.g., Gereffi et al. 2005), but TCE readily incorporates various hybrid governance forms, such as franchising and joint ventures (Bercovitz 2002, Borys and Jemison 1989). In this study, we use the label governance choice to refer collectively to all the options for organizing the supply chain.2

In theoretical and empirical supply chain management (see Tsay et al. 2018, for a recent, comprehensive review) research addressing governance choice in the supply chain falls into three main theoretical categories (cf. Santos and Eisenhardt 2005):

1. Competence: Are production activities assigned to the technologically and organizationally most capable (e.g., in terms of quality and cost) actors? In operations and supply chain management, empirical work on production competence (Cleveland et al. 1989), competitive manufacturing (Hayes and Wheelwright 1984), and strategic operations (Hayes et al. 1996) are representative examples of this approach. More generally, research in this approach is often couched in competence- and resource-based theories of the firm (Barney 1991, Teece et al. 1997).

2. Power: Who controls various stages of the supply chain? To what extent is an actor in the supply chain able to influence the actions of other actors? In operations and supply chain management, empirical work on resource dependence (Handfield 1993) and supply chain power (Benton and Maloni 2005, Handley et al. 2019) are representative examples. Theories of inter-organizational power and resource dependence (Leblebici et al. 1991, Pfeffer and Salancik 1978) constitute the intellectual roots of this approach.

3. Efficiency: What kind of a supply chain relationship enables the transfer of components and intermediate products from one production stage to another in an economically efficient way? How can transacting firms ensure that they do not waste resources in the exchange relationship? The intellectual roots of the efficiency view are in TCE (Williamson 1971). Grover and Malhotra (2003, pp. 463–464) provide a summary of TCE-based empirical research articles on the management of industrial supply chains.

Operations and supply chain management researchers typically choose one of these three theoretical lenses, but most of the research studies reviewed by Tsay et al. (2018) focus on how heterogeneous capabilities link to outcomes such as profitability. Research in operations and supply chain management therefore tends to lean toward the competence-based view.

However, a closer examination of the research articles in Tsay et al. (2018) reveals a number of important extensions that could benefit future research that applies the efficiency view in general and TCE in particular. Our first observation is that the TCE-based articles reviewed have adopted only a small subset of
TCE’s key concepts and logic, which indicates that research in operations and supply chain management has not incorporated many key insights that TCE offers. Such narrow application of TCE runs the risk of making the theoretical logic incomplete, and potentially misleading. This risk presents not only a problem, but also a research opportunity.

Another recurring observation is that some of the research studies that claim to apply TCE’s reasoning do so in a limited way. This concern is reminiscent of some challenges associated with the application of the resource-based view (RBV), which merits attention here. Specifically, Barney noted that many of the citations to the foundational RBV article (i.e., Barney 1991) were “used primarily to help establish the context of some empirical research—for example, that the focus is on the performance implications of some internal attribute of a firm—and are not really direct tests of the theory developed in the 1991 article” (2001, p. 46). In this study, we seek to ensure that the application of TCE in operations and supply chain management does not fall prey to similar misalignment where references to TCE are merely perfunctory or ritualistic, or where TCE is invoked mainly to establish the context for empirical research on supply chain governance. We raise this concern because even though Tsay et al. (2018) identified TCE as the most commonly applied theoretical perspective when reviewing supply chain research on outsourcing, TCE’s theoretical reasoning is often difficult to apply explicitly and rigorously in actual research. A common symptom we identified was confounding the efficiency view (of TCE) with the competence-based and the power views. 3

To understand better these misunderstandings and to prevent them from occurring in future research, looking at assumptions, central variables, unit of analysis, and general theoretical logic of TCE is instructive. We start our inquiry by first introducing the competence-based and the power views, and then discussing the TCE-based efficiency view, giving particular attention to how its logic differs from the other two. Table 1 provides a summary of the differences.

### 2.1. Competence: Who Are the Most Capable Actors in the Supply Chain?

A resource- and competence-based view of supply chains suggests that the firm will decide what to do in-house and what to outsource based on considerations of its relative competence: Is the firm technologically more capable in manufacturing a specific part than its potential external suppliers? Is the firm able to exploit scale or scope economies in ways that others in the supply chain cannot (e.g., Chandler 1990)? Note that in the competence-based view of the firm, the identities of individual firms matter, and research attention is focused on how the heterogeneity of firms’ competences influences firm-level heterogeneity in economic performance. 4 A case in point, many small businesses today outsource their warehousing and logistics operations to Amazon: the small business holds its products in an Amazon fulfillment center, where Amazon packs and delivers products to end customers. This organizational arrangement between the small business and Amazon is a supply chain governance decision, where the division of labor reflects primarily the fact that Amazon is more competent than the small business in executing requisite logistics activities and associated transactions. The arrangement works only if Amazon makes the service economically appealing to the small business owner. It seems to pass the market test: in 2017, Amazon generated $33 billion in revenue from this third-party seller service.

The competence-based view focuses on topics such as value creation, comparative advantage, competitive advantage, and persistent above-normal economic profits. In operations and supply chain management research, measures of economic value creation and capture are commonly adopted.
dependent variables, and researchers often examine heterogeneous capabilities as key antecedents. Indeed, the published literature on competitive manufacturing focuses on a capabilities approach as its fundamental theoretical lens: competitive manufacturing (Hayes and Wheelwright 1984), production competence (Cleveland et al. 1989), and high-performance manufacturing (Schroeder and Flynn 2002) are commonly used labels. Further, many teaching cases focus on comparative and competitive advantages arising from differences in operational capabilities (e.g., Southwest Airlines, Toyota, and Zara). Often, the theoretical focus is on linking operational capabilities specifically to outsourcing decisions—the Zara “fast fashion case” is such an exemplar (Caro and Gallien 2010, Ghemawat and Iniesta 2003).

The unit of analysis in the competence-based view is typically a legal entity, such as the firm, a strategic business unit, or an operating unit. The objective is to explain and predict heterogeneity among these entities, and its implications for differential economic performance (Bromiley and Rau 2016, Hitt et al. 2016, Ketokivi 2016).5

2.2. Power: Who Controls Supply Chains?
In the power perspective, attention focuses on actors who can exert power on others in the supply chain, effectively controlling assets without owning them. An example is a large final assembler who can “squeeze” its small suppliers by unilaterally setting prices. The power perspective also posits that actors in the supply chain seek ways to avoid dependence on any individual actor or firm in the chain. Final assemblers rarely produce all components (let alone raw materials) in-house and are thus dependent on suppliers, but the objective is to avoid dependence on a specific supplier. The exchange hazards of single-source supply are well-documented (Burke et al. 2007, Klein et al. 1978). In general, work on power (Benton and Maloni 2005), bargaining power (Wang et al. 2013), and resource dependence (Handfield 1993) in supply chains are staples in operations and supply chain management research.6

The unit of analysis in the power perspective often coincides with the competence-based view: the focus is on an organizational entity, such as the firm. However, in many research studies on power, the focus is on several entities simultaneously. This approach is understandable given that a power relationship involves one entity that has influence over the behavior of another. For example, Gray et al. (2009) examine performance implications of an original equipment manufacturer (OEM) outsourcing manufacturing to a comparatively more powerful contract manufacturer (CM). Importantly, performance implications for both the OEM and the CM were examined.

In the power perspective, outcomes other than profitability are also often examined. In an example of this approach, Wang et al. (2013) examine bargaining power and price setting in the OEM–CM relationship.

3. The Efficiency View: Transaction Cost Economics
We now turn to the third, efficiency view, where TCE serves as the primary theoretical foundation. Of the three perspectives, TCE is the only one that directs attention to governance costs. Because TCE uses many idiosyncratic concepts with specific, distinctive meanings, original TCE texts are often difficult to follow for those not immersed in the paradigm. The objective of this section is to introduce TCE’s assumptions, central concepts, and logic, through the example of automobile final assembly. Words that appear italicized in this section are central terms—premises, concepts, and other terminology—that are used in TCE in a specific and often distinctive meaning; specificity and hierarchy are such examples. We counsel to resist interpretation of italicized words in a conventional or colloquial way, because this may lead to misunderstandings. Williamson’s (1994, pp. 101–103 and 1996b, pp. 377–379) glossaries of TCE terminology complement the discussion that follows. In this section, we use footnotes rather extensively to benefit the reader by offering elaborations and extensions of the issues, as well as more detailed references to the TCE literature.

The starting point of a TCE analysis of an economic transaction is the assumption that the transacting parties share the mutual interest of organizing the transaction in an economically efficient manner to increase value creation. The foundational premise is that effective management of transactions requires efficient governance, which consists of cooperation and coordination: “[G]overnance [e.g., the business firm] is the means by which order is accomplished in a relation in which potential conflict threatens to undo or upset opportunities, to realize mutual gains” (Williamson 1996b, p. 12). Further, the study of governance “is concerned with the identification, explication, and mitigation of all forms of contractual hazards” (Williamson 1996b, p. 5). In TCE, contractual hazards arise from risks and uncertainties associated with long-term trading relationships, which are difficult to specify fully in a long-term contract.7

3.1. The Cost of Transacting: Beyond Production Costs
The name transaction cost economics derives from the fundamental insight that economic exchange has an ex ante preparation cost and an ex post execution cost (Coase 1937). On the ex ante side, economic exchange
involves drafting, negotiating, and safeguarding potentially complex contracts between two or more legal entities. Contractual parties also incur ex post costs “that arise when contract execution is misaligned as a result of gaps, errors, omissions, and unanticipated disturbances” (Williamson 1996b, p. 379). Consequently, daily execution of complex contracts requires constant attention, monitoring, adjustment, enforcement, renegotiation, and sometimes third-party mediation and arbitration, and in the extreme case, litigation. Further, because ex ante and ex post costs are interdependent, they must be addressed simultaneously, rather than sequentially (Williamson 1985, p. 21). The general prescription is to be forward-looking and to anticipate problems (e.g., by placing an arbitration clause in the contract), instead of reacting to them after they have already occurred.

In TCE, the label governance cost (Williamson 1985, p. 61) is used to refer to the ex ante and ex post costs associated with exchange. In corporate accounting practices, governance costs seldom appear as a separate cost category, instead, they are treated as an overhead cost. Such thinking may be a key reason for why executives direct more attention to salient direct production and procurement costs. Consider the firm’s legal department as an example: corporate legal expenses are essentially governance costs that arise when the firm engages in transactions with other companies. At the same time, these costs are seldom attributed—in the spirit of activity-based costing—to specific transactions.

Production costs are relevant to the investigation of governance decisions, because some firms may have comparative production cost advantages over other firms due to economies of scale and economies of scope (Baumol et al. 1982, Chandler 1990). However, a fundamental insight of TCE is that while economies of scope may lead to co-location, they do not necessarily imply co-ownership by way of vertical integration (Teece 1980, Williamson 1985). TCE recognizes the importance of production costs, but directs particular attention to the cost of transacting: “[F]irms are not concerned with transaction costs to the exclusion of revenues and production costs... A great deal of action nevertheless turns on transaction cost economizing, the importance of which is ignored or finessed by most constructions” (Williamson 1996b, p. 8).

A useful way of thinking of production costs in the TCE context is the following: Let us posit that firms are aware of the relevant production cost differences between alternative, feasible courses of action. The objective of a TCE analysis is to determine whether the magnitude of governance costs is high enough to offset potential production cost differences. The use of general- vs. special-purpose production equipment is a good example. Specialized equipment offers productivity benefits over general-purpose equipment (Hayes et al. 2005, Williamson 1985). However, if the lower governance costs associated with comparatively lower risk general-purpose equipment offset the productivity disadvantage, investing in general-purpose technology may result in comparatively lower total costs. An obvious advantage of general-purpose technology is its applicability to a variety of productive tasks.

The assertion that general purpose equipment is associated with comparatively lower risk and lower governance costs than special-purpose equipment (Williamson 1985, 1988) brings us to a consideration of TCE’s central concept, asset specificity: “the degree to which an asset can be deployed to alternative uses and by alternative users without sacrifice of productive value” (Williamson 1996b, p. 59). Specifically, the higher the sacrifice in productive value (due to non-redeployability), the higher the asset specificity.

Williamson (1996b) discusses five distinct types of asset specificity. First is site specificity in which the buyer and the seller are geographically co-located in successive supply chain stages, reflecting ex ante decisions to minimize inventory and transportation costs. For example, a coal-mining plant may be co-located with an electricity plant (Joskow 1988). Once sited, the assets in place are immobile. Second is physical asset specificity, in which one or both of the exchange parties to the transaction make investments in equipment that involves design characteristics specific to the transaction, and which have lower economic values in alternative uses. For example, a supplier in the automobile industry may invest in customer-specific dies to stamp components (Klein et al. 1978). Third is human capital specificity, in which investments in relationship-specific human capital often arise through various learning-by-doing processes (Harris and Helfat 1997). For example, software programmers may be required to commit their time to developing a company-specific proprietary programming language. Fourth, temporal specificity links specificity to time (Masten et al. 1991). For example, perishable fruits and vegetables must be handled in the supply chain in a timely manner, and may require vertical integration as a safeguard (Bucheli et al. 2010). Finally, dedicated assets are found in contexts where a supplier makes discrete investments in production capacity for the prospect of selling a substantial volume of output to a particular customer (Williamson 1983). If the contract were terminated prematurely, the supplier would have substantial excess capacity, which may be problematic even if it does not exhibit physical asset specificity.

Asset specificity may sound like a technology management issue. However, a key message in TCE is that

---

Ketokivi and Mahoney: Supply Chain Efficiency
specficity becomes relevant in economic transactions whenever one of the transacting parties, or both, commit to specificity in a way that locks them into the specific exchange relationship. Such lock-in gives rise to relational contracting (Williamson 1985, p. 71).

3.2. Governance Choice and Efficiency: The Discriminating Alignment

The TCE focuses on economizing (Williamson 2009), which Knight (1941, p. 252) equated to “reducing waste,” which resonates with theories of operations and supply chain management, such as lean manufacturing (Liker 2004, Liker and Choi 2004) and swift and even flow (Schmenner 2001, Schmenner and Swink 1998). Like TCE, these operations and supply chain theories focus on economic efficiency, which links directly to productivity. However, much like theories of competence and power, theories in supply chain management tend to abstract out the productivity impact of efficient governance. Consequently, TCE can offer an important complement to current theorizing of operations and supply chains.

Insofar as efficient choice is concerned, TCE posits weak-form selection (Williamson 1985, p. 23): given two or more known alternatives with their respective cost structures in a supply chain governance decision, TCE predicts that the exchange parties will choose the comparatively more efficient governance option. This fundamental proposition is known as the discriminating alignment hypothesis: “[T]ransactions, which differ in their attributes, are aligned with governance structures, which differ in their costs and competencies, in a discriminating (mainly, transaction-cost-economizing) way” (1991a, p. 277). The discriminating alignment hypothesis is not merely theoretical conjecture: the empirical evidence it has garnered since the 1970s is, in a word, remarkable (Macher and Richman 2008, Rindfleisch and Heide 1997, Shelanski and Klein 1995).

Achieving alignment of transaction characteristics with governance structures is complicated, because governance choices are embedded within broader institutional environments (Libecap 1989, North 1990, Ostrom 1990), which not only enable but also constrain firm behavior. In the TCE context, North’s (1991, p. 97) definition of institutions as “humanly devised constraints that structure political, economic and social interaction” is useful. The institutions relevant to TCE consist of both formal rules, such as constitutions, laws, and property rights, and informal constraints, such as sanctions, taboos, customs, traditions, and codes of conduct. In terms of changes in formal rules, a change in labor laws, for example, can change the calculus of efficient governance of hierarchy vis-à-vis relational contracting. In TCE, the institutional environment is incorporated through shift parameters, “changes in which elicit shifts in the comparative costs of governance” (Williamson 1996b, p. 253). Shift parameters constitute an organizational and managerial challenge, because, as Ahmadjian and Oxley (2005) highlight, it is often difficult to “export” lessons learned in one environment into a new, possibly drastically different environment.

3.3. Exploring Key TCE Terminology: The Case of Automobile Assembly

In the following, we introduce the central TCE concepts—in addition to the ones provided in the previous section—using automobile final assembly as the context. In the automotive supply chain, inter-firm exchange is ubiquitous: a common factor across modern automakers is that they operate in the supply chain as final assemblers that outsource the majority of components. For example, General Motors (GM) decided that instead of producing car seats in-house, it is better to purchase them from Adient, self-described as “a global” leader in the automotive seating supply industry with leading market positions in the Americas, Europe and China [that] maintains long-standing relationships with the largest global automotive original equipment manufacturers (OEMs). Adient’s proprietary technologies extend into virtually every area of automotive seating solutions” (Adient 10-K form, fiscal year 2018).

3.3.1. Asset Specificity, Uncertainty, and Frequency. GM’s decision to outsource seating production is likely based, in part, on Adient’s comparative production advantage due to its specialized expertise in seat production as well as aggregation economies of scale and scope (Wong 2017). TCE readily acknowledges comparative production advantages, but invites us to examine whether the magnitude of governance costs warrants attention. In the case of the GM–Adient relationship, the transaction is likely associated with substantive governance costs. One dimension concerns production location: Adient’s seat production is geographically co-located with GM’s final assembly plant, and is therefore site-specific. Co-location is crucial, because seats are inputs to an assemble-to-order production system: Adient must deliver seats to GM’s final assembly line according to a pre-determined schedule, both “just-in-time” and “in-sequence.” Adient has therefore made a massive relationship-specific investment with GM.

If Adient and GM either specified a complete contract that ensured continuity of the exchange relationship or provided for appropriate compensation in the event that one of the parties terminated the agreement, then Adient’s dependence on GM (and vice versa) would not be a problem. However, transacting relationships are typically subject to
uncertainty, the second dimension at the core of TCE. Specifically, TCE suggests that exchange relationships are subject to three main types of uncertainty (e.g., Walker and Weber 1984). One is technological uncertainty: the seats may require engineering effort, which means that their exact price and quality may not be ex ante known. Another example of technological uncertainty links to technological life cycles: When will current technology become obsolete (Balakrishnan and Wernerfelt 1986)? The second type is demand uncertainty: the demand for automobiles is significantly affected by factors outside the control of the exchange parties—input prices, interest rates, and general consumer confidence are but a few examples. Thus, the number of seats that GM will require in any given month or year is variable and unpredictable, leading to potential maladaptation problems such as order cancellations (Williamson 1983, p. 526). The third type is behavioral uncertainty: it is generally impossible for one exchange party to predict how the other party will behave in an unforeseen circumstance that the contract does not cover. In fact, exchange parties may not even be able to predict their own behavior in unforeseen and unprecedented circumstances. Because of complexity and unpredictability, comprehensive “contingent claims contracts” where “all relevant future contingencies pertaining to the supply of a good or service are described and discounted with respect to both likelihood and futurity” are infeasible (Williamson 1985, pp. 69, 333).

The third dimension of transactions is frequency of transacting. Exact numbers are typically confidential, but considering Adient’s global seating market share of roughly one-third and GM’s annual worldwide sales volume of roughly ten million automobiles in the past decade, and assuming GM purchases approximately one-third of Adient’s seating, we estimate the magnitude of car seats Adient supplies GM to be in the thousands each day. Importantly, this dimension is not just about volume but indeed frequency: because of the need for just-in-time and in-sequence delivery, Adient must indeed deliver car seats daily to GM’s assembly plants across the globe.14

TCE’s logic suggests that the transactional factors (asset specificity, uncertainty, and frequency) are sources of differential governance costs (Williamson 1979), which warrant attention in the GM-Adient case because all three factors are at a high level. If executives at GM and Adient have acted according to TCE’s prescription and analyzed the buyer-supplier relationship in its entirety (Williamson 1985, p. 29), they have considered comparative production and governance cost implications of these three factors.

3.3.2. Bounded Rationality and Its Implications. A key assumption in TCE is that contracts are unavoidably incomplete due to bounded rationality, which Simon described as behavior that is “intendedly rational, but only limitedly so” (1997, p. 88). According to Simon, limitations to rationality arise from the fact that “the capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problem whose solution is required for objectively rational behavior in the real world” (1957, p. 198). Williamson, in turn, pointed to limits in both cognition and communication: “Bounded rationality involves neurophysiological limits on the one hand and language limits on the other” (Williamson 1975, p. 21). Importantly, Simon further maintained that bounded rationality is the very reason why the management theorist “cannot simply chuck psychology overboard and place the theory of organization [solely] on an economic foundation” (1997, p. 88). TCE is therefore—despite its name—strictly speaking not an economic theory, but rather, an inter-disciplinary theory that combines law, economics, and organization theory (Williamson 1993). To be sure, TCE extends far beyond conventional microeconomic analysis (Mahoney and Pandian 1992, Scott 2001), and has had a major impact on the field of strategic management in particular (Mahoney 2005, Nickerson 2010).

TCE maintains that even though the transacting parties think they can generally rely upon one another, due to bounded rationality and information asymmetry, ensuring reliable execution of an exchange is impossible, under conditions of uncertainty in particular. Here, one driver of behavioral uncertainty has received special attention: TCE posits that transacting parties may engage in behavior that extends beyond simple self-interest (acting selfishly, but in good faith, “within the rules of the game”) to opportunism16 (acting selfishly in bad faith, “outside the rules of the game”). Because of bounded rationality, it is impossible to identify ex ante who is inclined to act opportunistically, and under what conditions. A central prescription in TCE is to “organize transactions so as to economize on bounded rationality while simultaneously safeguarding the transactions against the hazards of opportunism” (Williamson 1996b, p. 48).

The main point in TCE concerning uncertainty is that in an unpredictable world where exchange parties make irreversible, relationship-specific commitments, one is well advised to be forward-looking. Even though “not all future contingencies for which adaptations are required can be anticipated at the outset” (Williamson 1979, p. 237), the transacting parties should seek the proper safeguarding of the exchange relationship to make it robust to uncertainty.
Importantly, the emphasis here is on protecting and safeguarding the exchange relationship, not the individual firms in it: the unit of analysis in TCE is the transaction, not the firms engaging in the exchange.

3.3.3. Why Discrete Market Transactions Are Better for Simple Transactions. “Simple” transactions—where asset specificity, uncertainty, and frequency are low—are more effectively handled as discrete market transactions that rely on the price system (Macneil 1974, p. 378, Williamson 1996b, p. 370). Indeed, the price system can work well even when uncertainty and transaction frequency are high. For example, GM hardly needs to pay as much attention to how it sources light bulbs or tires compared to seats or steering columns. When transactions entail higher levels of asset specificity, trying to manage the exchange relationship through relational (short- or long-term) contracts can become cumbersome and involve high governance costs. At some point, these governance costs may become so high that it becomes economically more efficient to transfer the transaction from the market to a hierarchy, that is, inside the firm.

Many potential problems that drive up transaction costs in the GM–Adient relationship would be reduced if Adient were GM’s internal supplier. For example, there would be no need to determine what kinds of credible commitments are needed to safeguard the relationship-specific investment on Adient’s part, because the investment would effectively be transferred to GM’s corporate balance sheet. More generally, conflicts between the internal seat supplier and the final assembler could be addressed as internal managerial problems instead of involving third parties in arbitration or litigation. In fact, internal conflicts must be addressed as managerial problems: a buyer cannot take an internal supplier to court, because the company would be effectively suing itself. In effect, “a legal basis for the firm” (Masten 1988) exists, and given the law of forbearance, “hierarchy is its own court of ultimate appeal” (Williamson 1991a, p. 274).18

3.3.4. Bilateral Dependence and Credible Commitments. By building a seat production plant next to GM’s final assembly plant, the relationship-specific investment appears on Adient’s balance sheet. In contrast with theories of power, TCE emphasizes that, over time, even unilateral relation-specific investments become mutually beneficial (Kang et al. 2009). Thus, even though the investments may appear on the supplier’s balance sheet, “the buyer cannot turn to alternative source of supply and obtain the item on favorable terms since the cost of supply from unspecialized capital is presumably great” (Williamson 1979, p. 240). If GM lost Adient as a supplier, it either must find another supplier willing to make the same relationship-specific investment or start producing seats internally. Because of bilateral dependence that arises over time from mutual lock-in, it would be myopic for GM to jeopardize its exchange relationship with Adient by acting opportunistically.19 This insight constitutes perhaps the most salient difference between the power and the efficiency views.

Is it possible for a powerful and opportunistic buyer to take advantage of a relationship-specific investment made by the supplier? Such holdup problems (Goldberg 1976, Klein et al. 1978) sometimes occur and strong buyers may take advantage of weaker suppliers. However, given TCE’s premise that exchange parties are not myopic and focus on maintaining the transaction and seeking mutual gains, TCE both predicts and prescribes the adoption of proper safeguards to address holdup problems ex ante. Williamson noted that “[p]arties to a contract who look ahead, recognize potential hazards, work out the contractual ramifications, and fold these into the ex-ante contractual agreement obviously enjoy advantages over those who are myopic or take their chances and knock on wood” (Williamson 2000, p. 601). This is why realized (as opposed to potential) holdup problems are generally outside the scope of TCE, which focuses on mutually beneficial exchange relationships, where forward-looking exchange parties adopt the proper safeguards, and consequently, are willing to commit to specificity. To this end, addressing the economic holdup problem ex ante is prescribed. The preferred outcome, based on analyzing the transaction in its entirety, is that the transacting parties arrive ex ante at mutual credible commitments (Williamson 1983) that secure voluntary—not forced or captive—cooperation. The position taken by TCE is that voluntary cooperation is secured when both parties see exchange relationship continuity as beneficial.20 If a holdup problem has already occurred, something that TCE sought to prevent has happened.

Operationally, a credible commitment can be thought of as “a contract in which a promisee is reliably compensated should the promisor prematurely terminate or otherwise alter the agreement” (Williamson 1996b, p. 377).21 Outside the TCE paradigm, such mutual credible commitments have been described as “clever institutional arrangements [that] are a functional substitute for [trust]” (Granovetter 1985, p. 489). However, Williamson (1994, p. 97) cautioned against confusing the general notion of trust with its potential functional substitute. Ultimately, credible commitments are both theoretically and empirically salient without additional labels.22
3.3.5. Fundamental Transformation: The Dynamic of TCE. Relationship specificity often has an important dynamic aspect. In the GM–Adient relationship, even if the car seat market were perfectly competitive at the outset in that there were so many actors in the market that the actions of any single firm would be inconsequential (i.e., the identities of individual firms did not matter), relationship specificity might still develop over time as experience and learning accumulated in the buyer–supplier relationship. In TCE, this dynamic is called the fundamental transformation (Williamson 1985, p. 61): a competitive market (where the effect of any one individual firm is negligible) transforms into a market characterized by small-numbers bargaining (where individual firms may matter greatly, and specific firms can become bilaterally dependent). Therefore, even situations where the component supplied is standardized and multiple suppliers are available can develop relationship specificity over time, affecting contract renegotiation and switching costs. TCE predicts existing suppliers to have an advantage over new bidders in contract renegotiation precisely because the fundamental transformation has occurred. The emergence of small-numbers bargaining in this case is thus a direct consequence of specificity (Williamson 1985, p. 31).

3.3.6. TCE is Relentlessly Comparative. TCE does not suggest that internalizing a transaction solves all problems. The main proposition is that under conditions of high specificity, uncertainty, and frequency, internalizing the transaction may be economically more efficient than managing the relationship through a buyer–supplier contract. In this sense, TCE is relentlessly comparative (Williamson 1996b, p. 25) in that it “entails an examination of the comparative costs of planning, adapting, and monitoring task completion under alternative governance structures” (Williamson 1996b, p. 58). Further, TCE readily acknowledges that all feasible governance structures are flawed (Williamson 1996b, p. 7). The main claim, therefore, is that exchange uncertainty and the associated risk can be handled more efficiently within the firm than between two autonomous firms separated by a legal boundary. This is because even a small degree of authority or fiat (Barnard 1938, Simon 1997) can give transactions within the firm a comparative advantage over discrete or relational market transactions where authority is absent. Specifically, the managers of an internal supplier are willing to accept at least those directives of GM’s corporate management that are within the supplier’s zone of acceptance (Simon 1997), making the firm “a more elastic and adaptive mode of organization” (Williamson 1991a, p. 274) than the market. Furthermore, in the case of disputes, applying business judgment rules, the law of forbearance, arbitration, and other cooperative arrangements within the firm has an advantage over the court system where contractual disputes can become adversarial and cause irreversible damage to the exchange relationship: “relationships are effectively fracted if a dispute reaches litigation” (Williamson 1985, p. 71). TCE maintains that the firm acquires a number of important quasi-judicial functions (Williamson 1975, p. 30), whereby disputes can be addressed internally, in a more adaptive, sequential, and cooperative manner.

At the same time, the market mechanism has advantages over the firm in certain respects. If products and services are standardized and there are multiple buyers and sellers, the price system ensures effective exchange. The purchase of most consumer goods involves simple discrete market transactions where goods and services transfer from the seller to the buyer at a predetermined market price. TCE is particularly applicable in situations where identities matter substantially, most commonly due to some form of specificity that either exists already at the outset, or develops over time (Williamson 1994, p. 83).

3.3.7. Incentive Intensity. The essential advantage of discrete market transactions over within-firm exchange is that markets assign responsibility to individual economic actors more efficiently: an owner-managed supplier has unambiguous profit-and-loss responsibility and appropriates the net receipts (and suffers the losses) of its own efforts. The same cannot be said about the general manager of a corporate division that supplies another division. In internal transactions, problems start at setting the appropriate transfer prices at which products and service move from one internal unit to another. Further, even within decentralized firms, net receipts are often at least partially pooled, which makes the contributions of individual divisions somewhat ambiguous. In TCE terms, discrete market transactions have higher incentive intensity than within-firm transactions. As a general rule, high-powered incentives are more efficient in coordinating activities than managerial actions taken within the firm. Consequently, TCE often prescribes high-powered incentives: “Markets promote high-powered incentives and restrain bureaucratic distortions more effectively than internal organization… [High-powered incentives] favor tighter production cost control” (Williamson 1985, pp. 90–91).

There are, however, limits to the applicability of high-powered incentives (Williamson 1985, pp. 156–161). An example in the context of production management is the principle of treating production units and warehouses as cost centers, not profit centers. The objective of staying within a stated budget (cost center) leads to a lower-powered incentive than...
holding the unit profit-and-loss responsible (profit center). At the same time, attempts to create internal profit centers within the firm tend to be problematic, and a cost-center arrangement may be preferred.26

3.3.8. Focus on Remediable Problems. Should GM consider an alternative way of supplying seats to its final assembly? Is seat design and production within the firm possible? Would acquiring Adient (that supplies car seats to all major automakers) be feasible? In TCE, these questions fall under the rubric of remediable: “A condition is held to be remediable if a superior feasible alternative can be described and implemented with net gains” (Williamson 1996b, p. 379). With regard to outsourcing decisions, TCE acknowledges that many governance mode alternatives—such as backward integration into raw materials production—are simply infeasible. In the case of the GM–Adient relationship, GM’s purchase of Adient’s assets does not seem like a feasible alternative. But in other cases, backward integration may be both feasible and beneficial: one of the most analyzed TCE cases involves GM’s purchase of its then-external supplier Fisher Body in 1926. Prior to the merger, one of the vexing problems with the buyer–supplier relationship was Fisher’s reluctance to co-locate its body plants with GM’s assembly plants (Klein et al. 1978, p. 309).

Having discussed the objectives, assumptions, and logic of TCE, we now examine future research opportunities, focusing on how operations and supply chain management researchers can apply TCE and participate in broader conversations on governance.

4. The Future of TCE: Broader Cross-Disciplinary Conversations

While application of TCE in operations and supply chain management is well established (Grover and Malhotra 2003, McIvor 2009, Tsay et al. 2018, Wagner and Bode 2014), the conversations have largely been limited to within the operations and supply chain management community, which tends to be a net importer of theoretical ideas from other disciplines (Amundson 1998). Similar borrowing is widespread in the broader management research community, which builds extensively on theories in economics, psychology, and sociology (Agarwal and Hoetker 2007, Whetten et al. 2009). Our goal in this section is to examine the unique contributions that operations and supply chain management researchers could provide to broader conversations. Instead of merely borrowing and applying ideas from other domains, the ground is fertile for reciprocal contributions. Before we enter into the specifics, we discuss two important preliminaries by taking a brief look at the current state of the TCE conversation and the importance of “scaling up” theory.

4.1. Where is the TCE Conversation in 2020?

When operations and supply chain management researchers cite TCE, they cite primarily Williamson’s 1975 and 1985 books. However, much has happened in the past 35 years that warrants attention. Early emphasis was on minimizing transaction costs and analyzing antitrust implications (Williamson 1975). In contrast, Williamson (1985) devoted greater attention to total costs, developing a tradeoff framework that incorporated both production and governance costs. In his 1985 book, Williamson also provided an in-depth discussion of vertical integration (Williamson 1985, pp. 85–130). In our view, this joint treatment of both production and governance costs in particular brought TCE a giant step closer to research on operations and supply chain management.

Later developments (e.g., Williamson 1996b) incorporated the revenue side as well, therefore broadening TCE’s focus beyond cost minimization (cf. Glenk and Reichelstein 2019). Of course, to the operations and supply chain management audience, it should be self-evident that asset specificity, for example, can have implications for revenues as well: Hayes and Wheelwright (1984), among others, considered the strategic role of proprietary production equipment, which likely has specificity implications. More generally, the discussion on whether an industrial firm should use general- or special-purpose technology has always been relevant in operations management. These discussions extend to the revenue side as well (Hayes and Upton 1998).

Asset specificity has had a central role in the early treatments of TCE in particular. However, contemporary TCE conversations have shifted focus from asset specificity to issues such as adaptation frequency considerations and comparative adjustment costs (Argyres et al. 2019, Bigelow et al. 2019). In this still emergent stream of research, theorists set aside specificity and focus on the relationship between vertical integration and the difficulty of adaptation (e.g., Gibbons 2005, Tadelis 2002, Wernerfelt 2004). Some of the empirical contributions in this literature include Forbes and Lederman (2009), and Zhou and Wan (2017), which work toward establishing the empirical relevance of theories of the firm—based on both TCE and other theories of the firm—that focus on adaptation and adjustment. Because adjustments tend to involve changes in production output in responses to changes in demand (Argyres et al. 2019, p. 359), operations and supply chain management scholars could provide important contributions to these conversations.

In sum, early TCE conversations focused on cost minimization, but the contemporary topics discussed...
under the TCE umbrella are considerably broader and include adjustment costs, opportunity costs, and revenues. This broader focus readily aligns with many research questions in operations and supply chain management.

4.2. The Importance of Scaling Up a Theory
The other preliminary pertains to the unit of analysis, which in TCE is the transaction. TCE’s focus is thus not on the transacting parties (let alone just one of them) but in identifying the conditions under which the (dyadic) exchange relationship is organized as efficiently as possible. In contrast with theories of competence and of power, TCE typically does not focus on who benefits from this efficiency: efficiency is a characteristic of the transaction, and as such, a mutual concern of the transacting parties. TCE focuses on how transaction value is created by being mutually efficient; value appropriation is outside TCE’s scope.

However, even though the basic unit of analysis is the transaction, TCE is ultimately concerned with how a system of economic transactions—the entire firm, even the entire supply chain or industry—is organized (Williamson 1996b, p. 8). Here, distinguishing between level of theory (in TCE: the system) and unit of analysis (in TCE: the transaction) is useful (cf. Hitt et al. 2007). Ultimately, TCE addresses the foundational questions of what determines the scale and scope of the firm (Teece 1980, Williamson 1985). Further, applying TCE to explain and predict the scale and the scope of each actor in the supply chain leads to an understanding of the economic organization of the entire chain. Williamson considered such “scaling up” an indispensable characteristic of theory: “The object of a simple model [of an individual transaction] is to capture the essence, thereby to explain hitherto puzzling practices and make predictions that are subjected to empirical testing. Often, however, simple models can also be “tested” with respect to scaling up. Does repeated application of the basic mechanism out of which the simple model works yield a result that recognizably describes the phenomenon in question?” (2010a, p. 685).

In TCE, the ultimate phenomenon of interest is not the individual transaction but economic organization more broadly—this is why theories must “scale up.” To this end, Williamson offered the following prescription: “[C]laims of real world relevance, including public policy relevance, of any proposed theory of the firm that cannot be shown to scale up from toy model status to approximate the phenomenon of interest (e.g., the modern corporation) should be regarded with caution” (2010a, p. 685). Scaling up is a crucial aspect of both the utility and the potential progression of a theory. Theories progress in academic conversations where scholars put forth a collective effort to scale up a common theoretical predisposition (Ketokivi et al. 2017a; Kuhn 1970).

Two examples of scaling up TCE are Argyres and Liebeskind’s (1999) concept of governance inseparability, that is, how the governance mode choice of one transaction may constrain the options available for future transactions; and Kang et al.’s (2009) analysis of the inter-temporal and inter-project spillovers across transactions. In the spirit of these two attempts to “scale up TCE,” we explore in the following how operations and supply chain management researchers could provide useful contributions.

4.3. Contributing to Policy Conversations
The intellectual origins of TCE are in the policy and antitrust rulings of the 1960s and 1970s (Shapiro 2010, Williamson 1975). Many of the antitrust concerns related specifically to how firms had structured their supply chains: Was it anticompetitive for Schwinn, a manufacturer of bicycles, to impose constraints on who could sell its products, where, and at what price? Williamson (1985, pp. 183–189) disagreed with the Supreme Court’s decision to declare Schwinn’s actions anticompetitive. Williamson maintained that seeking monopoly power was not the only reason why firms integrate vertically. Vertical coordination could also be motivated by transaction efficiency.

Theoretical and empirical work in operations and supply chain management tends to ignore policy and antitrust issues (Cachon 2003). We suggest that because operations and supply chain management researchers have a unique and in-depth understanding of what happens in supply chains and why, they have the potential to contribute to policy conversations (Spring et al. 2017, de Treville et al. 2017). A better understanding of not only firm strategies but also supply chain strategies could serve discussions on policy. We should pay heed to Williamson’s (1996b) recommendation to seek an understanding of economic exchange in its entirety, which requires an inter-disciplinary approach.

Insofar as the explicit evaluation of firm action in supply chains is concerned, Amazon provides a useful illustration of current policy challenges. Amazon is being investigated both in the United States and in the European Union on its data collection from rival retailers in particular. This investigation requires a detailed understanding of Amazon’s (rapidly changing) business model and supply chain organization. The investigation must further explicitly consider competitive dynamics and potential harm to consumers. One of the key questions under examination is how Amazon uses the data it collects from smaller merchants on its site.
The policy and legal challenge associated with companies such as Amazon, Google, and Microsoft is that in addition to being massive corporations with undeniable de facto power in the supply chain, they are also in many ways unique companies. Consequently, there is no established industry practice against which actions of these corporate behemoths can be compared. More fundamentally, answering the simple question “To what industry should we classify Amazon?” turns out to be non-trivial. According to Investopedia.com (accessed November 2, 2018), Amazon competes with eBay, Netflix, Apple, Google, Walmart, Staples, Alibaba, Oracle, and Citrix Systems, just to mention a few examples. Further, to complicate matters, Amazon also collaborates with many of these competitors. When the characteristics of a unique and complex business model are considered, evaluating the efficiency of unfamiliar practices becomes more problematic, because instead of comparing to what others do, the company’s actions must be assessed on their own terms (Williamson 1996c, p. 151).

If management scholars are intimidated as they are trying to understand legislation and the court system, legal scholars and courts likely feel similarly about complexities of economic action and firm strategies. The courts must not only analyze these actions and strategies, but also rule on their legality. At least some of these analyses and rulings might be based on simplifying assumptions, lack of empirical rigor, and logical fallacies such as false equivalence and false dichotomies. “Easy answers to complex questions... relieved judges of the burden of wrestling with complicated facts” (Leslie 2014, p. 934).

In the academic management research community, operations and supply chain management researchers are particularly adept at examining supply chain action at the grassroots, with a variety of in-depth datasets and empirical methods that enable us to understand how supply chains function, who the actors are, and what they are trying to accomplish. A more detailed research of the grassroots of supply chains could be highly complementary to TCE-based research on industrial supply chains. Specifically, in contrast with operations and supply chain management researchers, general and strategic management researchers tend to focus on higher-level units of analysis and not map the key operational processes. A typical TCE-based study in general management, for example, looks at vertical integration as the dependent variable, and variables such as asset specificity, product complexity, and environmental uncertainty as the independent variables. Operations and supply chain management research can substantially complement these conversations by offering more detailed accounts of what specifically happens in these supply chains, how they are structured, who owns what, who controls what, and what kinds of contracts govern the transactions (Cachon 2003).

In terms of understanding supply chain action, all theories and models are simplifications. Nevertheless, the more the theories and the models are empirically rigorous about grassroots action, the higher the probability of avoiding over-simplification. A detailed understanding and mapping of the key processes can be highly productive in developing a better understanding of why firms behave the way they do.

4.4. Contributing to General Management Conversations

Reading the research on organizational boundaries in the general management literature that links to management of supply chains (e.g., Grover and Malhotra 2003, pp. 463–464), one quickly notices that this research often ignores many micro-level processes and nuances that are the staple of operations and supply chain management researchers. An example of this lack of nuance is treating sourcing as a binary variable: the final assembler either makes the component in-house or outsources it to vendors. However, this assumption seldom does justice to actual sourcing decisions, which involve a more complex combination of concurrent in-house production and outsourcing (Parmigiani 2007, Wang et al. 2013). In service settings, various co-production arrangements between the buyer (customer) and the seller (service provider) are common (Xue and Field 2008). Further, in contract manufacturing and contract design settings, outsourcing tends to involve not just the outsourcing of components and subsystems but also entire business processes, such as product design and warehousing (Bardhan et al. 2007). Finally, sourcing decisions seldom occur in isolation from one another, instead, decisions regarding what to outsource tend to occur in bundles.

Elucidating the relevant micro-level processes and structures can complement the higher-level analyses that focus on examining statistical relationships among variables of interest. Adler’s (1995) study on inter-departmental interdependence (Thompson 1967) and coordination, for example, examined the micro-processes and nuances associated with managing the interface of design and manufacturing. One of Adler’s key arguments was that “in the course of a product development project, neither inter-departmental interdependencies nor coordination mechanisms are constant over time” (1995, p. 148). Understanding processes may also be instrumental in understanding competitive advantage: “[F]irms may possess competitive advantages at the level of...
business processes that are not reflected in a firm’s overall performance” (Ray et al. 2004, p. 34).

Operations management researchers are well trained and experienced in examining processes, disaggregating performance, and giving attention to nuance. Additionally, much of operations and supply chain management research is located specifically at the level of analysis where humans and technology turn inputs into outputs; investigating the productivity of an assembly line is a good example. Ray, Barney, and Muhanna described the advantages of focusing on this level of analysis: “By focusing on activities, routines, and business processes where resources are deployed and where their first-order effects are expected to be realized, managers might be in a better position to benchmark the resource endowment of their firms and identify critical resources that should be exploited, developed, and protected” (2004, p. 36). Many operations and supply chain management researchers spend their entire careers examining precisely these first-order effects. Often, shifting analysis from the level of the process to the entire organization glosses over important heterogeneity. For example, a factory may have seven production lines with drastically different productivities, and thus, aggregating productivity metrics to the factory level may mask important variance in productivity.

Insofar as TCE’s key concepts of asset specificity, uncertainty, and frequency are concerned, their treatment in the general management literature often seems uncomfortably abstract to the operations and supply chain management researcher. Invoking the notion of uncertainty immediately leads to questions such as “What type of uncertainty?” and “Which parts of which processes does it affect and how?” Operations and supply chain management scholars could participate in conversations by engaging in important disaggregation and empirically quantifying phenomena that general management scholars have addressed only qualitatively and theoretically. Walker and Weber’s study of the make-or-buy decision in the automobile final assembly is a good example: “When volume uncertainty is high, suppliers experience unexpected production costs or excess capacity and buyers experience stock-outs or excess inventory” (1984, p. 376). Operations and supply chain management researchers know how to estimate these costs, which can lead to a deeper understanding of the make-or-buy decision. Alternatively, consider the hypothesis that as volume uncertainty increases, the general propensity of making the component in-house rather than outsourcing it increases (Walker and Weber 1984). One could elaborate this hypothesis by introducing and quantifying the obvious mediating variable—the increasing transaction cost due to excess capacity or inventory—instead of offering it merely as a theoretical mechanism. Indeed, in their examination of transaction costs in operations management, Emery and Marques (2011, pp. 238–241) looked empirically at transaction costs by focusing specifically on inventory costs.

Disaggregation and measurement of transaction costs is useful in establishing managerial relevance in particular. Nevertheless, the problem is also relevant in empirical tests of TCE (Shelanski and Klein 1995, p. 339): “[E]mpirical research in TCE is often hampered by confusion about the definitions, and therefore the empirical parameterization, of key variables. The primary conceptual problem that we have found lies in the treatment of uncertainty as a factor that raises transaction costs and increases the probability of [vertical] integration. The confusion may explain some seemingly contradictory results on the effects of [uncertainty] on the vertical integration decision.”

An additional challenge with empirical research in the general management literature on TCE is the reliance on proxy variables (Macher and Richman 2008). In operations and supply chain management research, use of direct measures—operational definitions and measurement scales, not proxies—is more common (Roth et al. 2008). Operations and supply chain management researchers are thus in a unique position to participate in academic conversations by offering “greater micro-analytic measurement” (Macher and Richman 2008, p. 40) of the key concepts, and consequently, a more fine-grained and nuanced understanding of governance decisions.

Finally, operations and supply chain management researchers probably need not be reminded that “[the location of facilities, the adoption of specialized designs or equipment, and the scale of investments should all, by rights, be treated as endogenous variables” (Masten 1995, p. 60).29 The reminder is, however, important for general management researchers who have largely treated asset specificity as an exogenous variable (cf. Macher and Richman 2008). For operations and supply chain management researchers, process choice is a decision variable driven by factors such as product life cycles (Hayes and Wheelwright 1979). Similarly, the drivers of production location decisions—a potential source of site specificity—are well documented in the operations management literature (Brush et al. 1999, Ferdows 1989, Ketokivi et al. 2017b; Schmenner 1982). Herein can be found another contribution opportunity for operations and supply chain management researchers: providing tools and insights on how to incorporate asset specificity as an endogenous variable.
4.5. Conclusion
In this study, we have described and clarified the logic of transaction cost economics and examined its applicability to operations and supply chain management research. Considerable overlap between the central problems in the two domains makes the ground fertile for “give and take.” Indeed, the conclusion that TCE can be applied directly as a theory of supply chain efficiency is clearly warranted. The efficiency approach further offers important contrasts to views that emphasize competence and power, both of which are in many ways more established perspectives in supply chain research. What is more, supply chain efficiency is arguably more important than examination of supply chain competence or of power: whereas only some companies have sufficient power to engage in genuinely strategic actions in the supply chain or to exercise control over assets owned by other companies, economizing has actionable implications for all actors in the supply chain, large or small. Avoidance of waste has always been, and continues to be, everyone’s business.

We submit that becoming a participant to TCE conversations offers a possible avenue for operations and supply chain management researchers to broaden their role in the academic management community. Due to the focus on micro-analytics, operations and supply chain management researchers can contribute to conversations on organizational boundaries through detailed examinations of supply chain governance structures and processes, and direct operationalizations of key constructs and variables. This approach offers an important complement to more general-level theories that rely more on abstractions and proxy variables. Operations and supply chain management researchers can offer greater precision in describing and explaining supply chain action.

What is next? As TCE continues to develop, scholars find novel, puzzling phenomena that call for interpretation and explanation. What implications does blockchain technology have for economic organization and governance (e.g., Feng and Shankikumar 2018)? Regarding the topic of probity hazards, are private prisons an organizationally appropriate alternative to public prisons? How should policy-makers evaluate the legality of idiosyncratic business models, such as that of Amazon’s? TCE can offer insight to these emerging questions, and many others. TCE’s strength is that it has always been inter-disciplinary (Williamson 2005). We encourage future developments to strengthen this inter-disciplinary basis by an ever-broadening array of specialists coalescing around common research problems to ask the central question that kept Williamson intrigued and busy for over half a century: What is going on here? (1996b, p. 25).

The world is changing so rapidly that the academic community struggles to keep up with the pace of development. The strategic, operational, and policy challenges of new forms of organizing are as massive as they are diverse, and the fields of strategic management and operations and supply chain management have much to contribute in the evolving science of organization. In particular, the next generation of research should strive for an even stronger synthesis of the governance and competence perspectives to incorporate not only costs but also revenue, profit, strategy, and competitive advantage (Williamson 1999b). To this end, operations and supply chain management researchers have much to offer, which is why we invite our colleagues to join forces with the broader academic management community and embrace Williamson’s optimistic, forward-looking attitude: “Research in transaction cost economics faces an interesting, challenging future” (2010b, p. 224).

Acknowledgments
We thank the six anonymous reviewers, Editor Kalyan Singhhal, and Jackson Nickerson for their counsel. The usual disclaimer applies.

Notes
1We are indebted to a reviewer who importantly pointed out that whereas the make-or-buy decision is a salient example of a TCE topic that is relevant to operations and supply chain management research, TCE addresses a broad range of contract design issues that are equally relevant to operations and supply chain management research as well: administered (long-term) contracting in regulatory settings (Goldberg 1976), contract duration in general (Joskow 1987), pricing in long-term contracts (Crocker and Masten 1991), long-term procurement contracting (Bajari and Tadelis 2001), hybrid contracts (Kalnins and Mayer 2004), strategic alliance contracts (Reuer and Ariño 2007), and contract design dynamics (Argyres and Mayer 2007) are but a few examples. In the case of automobile assembly, TCE can shed light not only on the (binary) make-or-buy decision, but also on the specific details of buyer–supplier contracts, such as incentive structures, dispute resolution, intellectual property rights, and so forth. In addition to these topics relevant to supply chain management, TCE also addresses topics that are relevant to management practice more generally: firm capital structure, the composition of the board of directors, and employment contracts are good examples (Williamson 1996b).

2Williamson maintains that, “alternative modes of governance—markets, hybrids, hierarchies, bureaus—differ from each other in structural ways” (1996b, p. 7). Each mode of governance is supported by and in significant ways is defined by a distinctive form of contract law, and hence constitutes a discrete structural alternative (Williamson 1991a), which is at the heart of the TCE framework.
Williamson challenges “the muscular approach,” which posits that, “one of the parties, usually the large buyer, deals with smaller suppliers in a peremptory way, [and] often ‘use up’ their suppliers and discard them” (2008, p. 10). He further submits that such an approach is “myopic and inefficient,” particularly when “investments in specific assets are made” (Williamson 2008, p. 10). In such situations, Williamson prescribes (2008, p. 10) “the benign approach” and calls for attention to cooperation and mutual gains. TCE and theories of power are fundamentally different in that theories of power adopt “the muscular approach,” TCE represents “the benign approach.”

Hodgson (1998) strongly champions the competence perspective to explain the existence, organizational structure, and organizational boundaries of the firm. Williamson (1999b, pp. 1096–1097), however, notes that while Hodgson (1998) provides “a series of interesting remarks about formal and informal relations, tacit knowledge, mental models, organizational learning, trust, dynamic corporate culture, and the like, we are never told why these effects work better (or worse) in a unified firm (AB) rather than in autonomous firms (A and B).”

Although headway has been made, Williamson’s (1996b, pp. 227–228) suggests a need to unpack the construct of competence. Key features are communication codes that a firm develops (Arrow 1974, Williamson 1975), routines that a firm uses (Cyert and March 1963, Nelson and Winter 1982), and firm-level capabilities that develop over time (Penrose 1959, Teece et al. 1997). Implementing the competence-based view is a work in progress. Studies joining TCE and capabilities include Argyres (1996), Hoetker (2005), Novak and Stern (2008), and Argyres and Zenker (2012), among others.

Williamson (1996b, pp. 238–239) notes that the construct of power is diffuse and multifaceted: the power of capital over labor (Bowles and Gintis 1993), strategic power over rivals (Shaprio 1989), power in the political process (Moe 1990), and power based on resource dependency (Pfeffer and Salanick 1978). All of these aspects of power are relevant to organizations, but resource dependency in particular is relevant to TCE. However, even within discussions of resource dependency, there are important nuances: “The strong version of resource dependency assumes myopia, [in which] myopic parties to contracts are victims of unanticipated and unwanted dependency. Because myopic parties do not perceive the hazards, safeguards will not be provided…” (Williamson 1996b, p. 239). Williamson (1996b) contrasts myopia with TCE’s more farsighted view, which holds that not only are managers able to consider asset redeployability at the time of the initial contract (Masten 1984), but also that the exchange parties can introduce various economic safeguards as asset specificity builds up over time (Joskow 1988). Therefore, in TCE, dependency is neither unanticipated nor necessarily unwanted, but it does give rise to various safeguards.

On this point, TCE and agency theory provide complementary efficiency views. Agency theory has two forms: (i) The mathematical principal-agent model (Holmstrom 1979) is concerned with efficient risk bearing under uncertainty and asymmetric information; (ii) Positive agency theory is concerned with monitoring and economic bonding to reduce agency costs (Fama 1980). There are several differences between both variants of agency theory and TCE. First, the unit of analysis of agency theory is the individual (Jensen 1983, p. 327), while TCE regards the transaction as the basic unit of analysis (Commons 1934). Second, agency theory assumes fungible assets (Alchian and Demsetz 1972), while TCE incorporates non-redeployability as a central characteristic of assets. Third, agency theory focuses on minimizing the sum of monitoring costs, bonding costs, and “residual loss” (Jensen and Meckling 1976), while TCE focuses on the costs of maladaptation (Williamson 1996b, p. 179). Fourth, agency theory views the firm as a “nexus of contracts” while TCE regards the firm as a “governance structure.” Fifth, the principal-agent model considers risk-averse agents, while TCE posits risk-neutrality. Sixth, agency theory examines ex ante incentive alignment, while TCE focuses on designing ex post governance structures (market, hierarchy, or hybrid) (Hennart 1993, Oxley 1997, Williamson 1991b) to safeguard against contractual disputes (e.g., by use of mediation and arbitration clauses). Seventh, agency theory assumes an evolutionary process of strong-form selection of “survival of the fittest,” (Jensen 1983, p. 331), while TCE subscribes to weak-form selection of “survival of the fitter” (Williamson 1996b, p. 177). For a detailed comparison of these theories see Kim and Mahoney (2005), Lajili and Mahoney (2006), and Tan and Mahoney (2006).

The TCE literature often uses labels transaction cost and governance cost synonymously. In this study, we use the latter term as an inclusive label for all relevant costs. For example, employment contract negotiations are a relevant cost category, however, to think of the employment relationship as a transaction may sound somewhat awkward; thinking of it as governance may be less confusing.

TCE specifies that the ownership of an asset consists of “(a) the right to use the asset, […] (b) the right to appropriate returns from the asset; [and] (c) the right to change the asset’s form and/or substance” (Williamson 1996b, p. 112). Ownership further involves both ex ante residual income rights (i.e., residual claimancy) and ex post residual control or decision rights (Alchian and Demsetz 1972, Grossman and Hart 1986). TCE emphasizes that all the relevant costs and benefits of these property rights—both ex ante and ex post—be incorporated into the analysis.

Further, to simplify analysis, many TCE studies focus exclusively on governance costs (Williamson 1991a, 1996b). Analogously, researchers in production economics may be interested in modeling total factor productivity (Van Beveren 2012), but typically focus on a subset of its drivers, such as production technology.

When a firm vertically integrates, there are changes not only in ownership but also in incentives and monitoring (Mahoney 1992, Mahoney and Qian 2013, Williamson 1985).

Williamson (1996b, p. 60) distinguishes between environmental and behavioral uncertainty, or essentially, between uncertainty exogenous to the relationship (e.g., changes in the institutional environment) and uncertainty endogenous to the relationship (e.g., uncertainty about how the transacting parties will behave, and whether this behavior
takes on opportunistic features). Demand and technological uncertainty are two—among many others—sources of environmental uncertainty.

Williamson (1996b, p. 26) suggests that the central problem of economic organization is adaptation, and that different governance modes have different adaptation mechanisms associated with them. Generally, discrete market transactions allow for autonomous (Hayek 1945) adaptation. In internal (within-firm) transactions, adaptation is cooperative (Barnard 1938). Adaptation mechanisms in hybrids (alliances, franchising, and joint ventures) are typically a combination of autonomous and cooperative adaptation (Dyer and Singh 1998, Williamson 1996b, pp. 102–103). The costs and benefits of adaptation mechanisms have implications for transaction cost efficiency and must be incorporated into the analysis of governance decisions (Nickerson and Silverman 2003).

Williamson’s (1979) main logic concerning the role of frequency was that it would not pay to incur the costs of an elaborate governance structure (e.g., creating a firm) for one-time or occasional transactions. In this sense, an increase in frequency leads to the prediction of greater vertical integration. However, a plausible counter-argument exists when there is a long “shadow of the future,” in which maintaining an ongoing relationship is in the economic interest of the exchange parties, and the discounted value of anticipated future transaction outweighs the economic benefit of a one-time opportunistic act that runs the risk of terminating the exchange relationship (Klein and Leffler 1981). In this sense, transactional frequency reduces the likelihood of opportunism and therefore reduces the need for vertical integration.

The incomplete contracting setup in TCE builds on bounded rationality, and consists of the following. Contracts are effectively incomplete if (i) not all relevant future contingencies can be imagined; (ii) the details of some of the future contingencies are obscure; (iii) a common understanding of the nature of the future contingencies cannot be reached; (iv) a common and complete understanding of the appropriate adaptations to future contingencies cannot be reached; (v) the parties are unable to agree on what contingent event has materialized; (vi) the parties are unable to agree on whether actual adaptations to realized contingencies correspond to those specified in the contract; and (vii) even though both parties may be fully apprised of the realized contingency and the actual adaptations that have been made, third parties (e.g., the courts) may be fully apprised of neither, in which event costly haggling between bilaterally dependent parties may ensue” (Williamson 1996b, p. 378).

In TCE, opportunism is “self-interest seeking with guile, to include calculated efforts to mislead, deceive, obfuscate, or confuse” (Williamson 1996b, p. 378), and can take many forms. One is adverse selection, which is a derivative condition of ex ante opportunism and asymmetric information (Williamson 1975, p. 14). For example, in the context of insurance, individuals may not candidly disclose their objective risk attributes (Williamson 1996b, p. 15). Another form is moral hazard, which is a derivative condition of ex post opportunism and asymmetric information (Williamson 1996b, p. 15); in the case of insurance, individuals covered by insurance may not exercise the same degree of due care (Williamson 1985, p. 51). A third form is the economic holdup problem, another form of ex post opportunism: an attempt by some contractual parties to renegotiate the contract to their advantage because there are appropriable quasi-rents. “The quasi-rent value of the asset is the excess of its value over its salvage value, that is, its value in its next best use to another renter. The potentially appropriable specialized portion of the quasi rent is that portion, if any, in excess of its value to the second highest-valuing user” (Klein et al. 1978, p. 298).

Unfortunately, those less familiar with TCE often misinterpret or exaggerate the opportunism assumption. Contrary to what some TCE critics (e.g., Ferraro et al. 2005) propose, TCE merely acknowledges the possibility of opportunism, it does not make any claims about its prevalence. Theoretically, positing that opportunism constitutes a non-trivial exchange hazard is sufficient (Ketokivi and Mahoney 2016). Williamson captures the sentiment: “[M]ost people will do what they say (and some will do more) most of the time without self-consciously asking whether the effort is justified by expected discounted net gains... But while accurate descriptions of what is going on ‘most of the time’ are plainly essential, much of what is interesting about human behavior in general and organizations in particular has reference not to routines but to exceptions” (2010a, pp. 678–679). Both critiques (Ferraro et al. 2005, Ghoshal and Moran 1996) and their rebuttals (Ketokivi and Mahoney 2016, Williamson 1996a) are well-documented in the literature, and need not be revisited here. Suffice it to state that scholars have taken equivocal positions on whether the assumption of opportunism is a necessary condition for transaction costs to occur: some submit that it is necessary (Foss 1996, Mahoney 2001, Williamson 1985), others maintain that it is not (Alchian and Woodward 1988, Conner 1991, Kogut and Zander 1992). Opportunism certainly is a sufficient condition for substantial ex post governance costs to occur when appropriate income is available (Klein et al. 1978, Williamson 1979).

Whereas the courts routinely hear disputes over prices, delivery, quality, and the like in transactions between firms, these same courts refuse to be drawn into identical disputes between divisions within a single firm. In effect, hierarchy becomes its own court of ultimate appeal. It is largely because of forbearance law that firms are able to exercise fiat, whereas markets cannot, to manage transactions.” (Williamson 1996b, p. 27). Further, “[t]he underlying rationale for forbearance is twofold: (i) parties to an internal dispute have deep knowledge—both about the circumstances surrounding a dispute as well as the efficiency properties of alternative solutions—that can be communicated to the court only at great cost, and (ii) permitting the internal disputes to be appealed to the court would undermine the efficacy and integrity of hierarchy” (Williamson 1996b, p. 150).

TCE predicts that discrete market transactions will work efficiently when the exchange relationships involve no specificity. Indeed, absent specificity, “rivalry among large numbers of bidders will render opportunistic inclinations ineffectual” (Williamson 1975, p. 27). Some type of specificity is required to introduce contractual hazards, which,
in turn, call for an explicit examination of the comparative efficiency of governance mode alternatives in the exchange process.

This introduces the idea of incomplete contracting in its entirety (Williamson 1996b, p. 9) “in which [all] complex contracts are unavoidably incomplete because of bounds on rationality” (Williamson 1996b, p. 377). However, “even boundedly rational decision-makers have the capacities both to learn and to look ahead, perceive hazards, and factor those back into the contractual relation” (Williamson 1996b, p. 9). Bound rationality does not imply myopia.

A credible commitment is the irreversible and non-salvageable part of an investment that “can have both ex ante (screening) and ex post (bonding) effects” (Williamson 1996b, p. 123). For example, when the mini-mill steel company, Nucor, made a pioneering investment with SMS Company in Germany for the new technology of thin-slab casting, the initial investment was $340 million (Ghemawat 1993). If $250 million of Nucor’s investment is redeployable under the circumstance of complete failure of the new technology, then the economic commitment of Nucor is $90 million, that is, the $90 million is the irreversible part of the $340 million investment. Nucor’s willingness to make an irreversible investment of $90 million is a credible signal to SMS of Nucor’s capabilities and thus serves the ex ante screening function. Ex post, the $90 million irreversible commitment by Nucor aligns economic incentives with SMS. Thus, in general, ex post bonding effects are particularly important as they pave the way toward self-enforcing contracts in which the economic incentives of the exchange parties are aligned. Mutual commitments or “mutual hostages” (Ahmadjian and Oxley 2005, Kim and Mahoney 2006, Schelling 1960, Williamson 1983) support exchange and cooperation.

If it is redundant at best and can be misleading to use the term ‘trust’ to describe commercial exchange for which the cost-effective safeguards have been devised in support of more efficient exchange” (Williamson 1996b, p. 256); . . . “[U]ser-friendly terms [such as trust] do not encourage us to examine the deep structure of the organization. Rather, we need to understand when credible commitments add value and how to create them, when reputation effects work well, when poorly, and why. Trust glosses over, rather than helps unpack, the relevant micro-analytic features and mechanisms” (Williamson 1996b, p. 216).

The extreme form of small numbers bargaining is the bilateral monopoly, a market with just one buyer and one seller. A bilateral monopoly can be very problematic because the transacting parties are “inclined to expend considerable resources bargaining over the price at which the exchange is to take place” (Williamson 1975, p. 28).

Williamson (1985, p. 161) asked, “Why can’t a large firm do everything that a collection of small firms can do and more?” The answer is because selective intervention within the firm is impossible. In the case of a buyer–supplier relationship, selective intervention would mean that the supplier would make all of its decisions independently, unless compelled by the buyer’s intervention to adapt to new circumstances. Symmetrically, the supplier could impose a similar intervention on the buyer. Williamson (1985) maintained that even when the buyer and supplier are part of the same firm, selective intervention—by the buyer, the supplier, or corporate management—is infeasible. What is worse, “the transfer of a transaction out of the market into the firm is regularly attended by an impairment of incentives. It is especially severe in circumstances where innovation (and rewards for innovation) are important” (1985, p. 161). It should be noted that the weaker incentives of integration (and bureaucracy, more generally) still have their place (Williamson 1999a). Here, as elsewhere, comparative assessment of imperfect governance mode alternatives is needed.

Organizations suffer from many bureaucratic distortions: (i) internal procurement bias in which “the existence of an internal source of supply tends to distort procurement decisions . . . A norm of reciprocity easily develops, [for example], I buy from your division, you support my project proposal or job promotion” (Williamson 1975, pp. 119, 120); (ii) internal expansion bias where the organization adopts “a compromise solution by which concessions are made to subsystems rather than require them to give up essential functions or resources” (Williamson 1975, p. 120); and (iii) persistence bias, where the “sunk costs in programs and facilities of ongoing projects insulate existing projects from displacement by alternatives which, were the current program not already in place, might otherwise be preferred” (Williamson 1975, p. 121).

More generally, due to various probity hazards (Williamson 1999a), lower-powered incentives are preferred in some contexts. Foreign affairs, the military, and the prison system are examples in which high-powered incentives may be downright detrimental. In these “sovereign transactions” (Williamson 1999a, p. 307) contexts, performance-based-pay (a high-powered incentive) should be supplanted by a simple hourly pay or basic salary (a low-powered incentive). Williamson (1999a, p. 318) elaborated the logic: “To denounce public agencies because they have lower-powered incentives, more rules and regulations, and greater job security than are associated with a counterpart private bureau completely misses the point if those features have been deliberately crafted into the public bureau, thereby mitigating contractual hazards, albeit at a cost.”

Williamson (1996b, p. 2010) comments on using the label inefficient in reference to a governance mode that falls short of the criterion of Pareto optimality: “Claims of inefficiency that can be recognized only after the fact and/or cannot be implemented with net gains have no operational importance.”

Even if Amazon is highly competent for certain supply chain activities, and further assuming that Amazon holds all of the bargaining power in their supply chain relationships, Amazon would (and should) still prefer to set up a comparatively efficient governance arrangement (Silverman 2002).

Nickerson et al. (2001), provide a supply chain analysis of international courier and small package services, which predicts a fit among three strategic choices: market position, resource profile, and organizational structure, and shows an endogenous self-selection of strategies across the entire supply chain.


