Transaction Cost Theory: Past Progress, Current Challenges, and Suggestions for the Future

Article in The Academy of Management Annals - October 2020
DOI: 10.5465/annals.2019.0051

CITATIONS 30
READS 1,270

4 authors, including:

Jean-François Hennart
Tilburg University
130 PUBLICATIONS 12,912 CITATIONS

Brian Silverman
University of Toronto
80 PUBLICATIONS 11,919 CITATIONS

Some of the authors of this publication are also working on these related projects:

Biotechnology Startups View project
Asset Ownership and Incentives in Early Shareholder Capitalism: Liverpool Shipping in the 18th Century View project
TRANSACTION COST THEORY: PAST PROGRESS, CURRENT CHALLENGES, AND SUGGESTIONS FOR THE FUTURE

ILYA R. P. CUYPERS
Singapore Management University

JEAN-FRANÇOIS HENNART
Tilburg University
Politecnico di Milano
Aalborg University

BRIAN S. SILVERMAN
University of Toronto

GOKHAN ERTUG
Singapore Management University

Transaction cost theory (TCT) has been fruitfully applied to a wide range of organizational phenomena, as reflected in a vast and evolving body of research. However, in part due to the theory’s broad success, important advances in some fields have not diffused to other fields. In this essay, we lay out a path toward a pluralistic view of TCT that incorporates insights from multiple fields, primarily strategy and international business. In so doing, we critically assess the assumptions, key constructs, and evolving theoretical logic of TCT. We then propose an agenda for future research, highlighting opportunities for scholars to (a) expand and deepen the exchange of insights between strategy and international business, and further integrate TCT with the trust literature and with recent insights from behavioral economics and psychology, and (b) further apply TCT to study recent phenomena such as platforms and two-sided markets, the implications of artificial intelligence for governance decisions, and the pursuit of non-pecuniary objectives such as sustainability.

Building on the seminal work of two Nobel laureates (Coase, 1937; Williamson, 1975, 1985), transaction cost theory (TCT), or transaction cost economics, has become one of the most influential theories in management research. Originally applied to the “make-versus-buy” vertical integration decision, TCT has been used to shed light on a broad range of organizational phenomena, including horizontal diversification, the multinational enterprise, strategic alliances, supply chain relationships, and public–private partnerships (PPPs). TCT has also expanded to encompass an increasing roster of factors that predict governance choice, as well as the performance consequences of this choice. This expansion is evidence of the theory’s influence and success. However, that influence and success is a double-edged sword. Given the breadth of its application, the TCT literature risks becoming fragmented and difficult to navigate, as advances in one sphere are overlooked by others, and as definitions of key concepts evolve differently across these spheres.

The purpose of this article is to provide a roadmap for navigating this vast literature and to offer suggestions for conceptual and empirical future work on
TCT. Whereas prior surveys of TCT have summarized the broad empirical state of the field (e.g., David & Han, 2004) or focused on empirical phenomena such as vertical integration and entry mode choices (e.g., Geyskens, Steenkamp, & Kumar, 2006; Zhao, Luo, & Suh, 2004), our survey emphasizes the distinctive features of TCT research across different conceptual areas. We first provide an overview of the basic proposition of TCT as initially formulated by Williamson (e.g., Williamson, 1973, 1975), revisiting its underlying assumptions, key theoretical constructs, and logic. We then illustrate how the literature has evolved from this initial formulation, paying particular attention to contributions from the strategy and international business literatures. Of particular note is that whereas TCT initially emphasized asset specificity as the most important of the three attributes affecting governance, subsequent advances have highlighted the importance of behavioral uncertainty, and have also considered factors beyond that initial set of three attributes. To spur future studies in TCT, we conclude with a call for research that can apply these insights to three conceptual areas—further integration of research in strategy, international business, and institutional economics; greater engagement with research in sociology, in particular the literature on trust and the research on formal–informal organization; and greater links to the psychology and behavioral economics literatures for a better understanding of the behavioral processes that frame the governance of transactions. In addition, we highlight three phenomenological areas for future research—platform governance, technological advances such as artificial intelligence and machine learning, and the increased prevalence of nonpecuniary goals for a range of actors (e.g., the rise of nationalism, corporate social responsibility, and “grand challenges”). Given these conceptual and empirical opportunities for further development and application, we are confident that TCT’s future is as bright as its past.

THE BASIC PROPOSITION

TCT’s fundamental prediction is that organizational actors attempt to maximize the gains of interdependence by “assigning transactions (which differ in their attributes) to governance structures (the adaptive capacities and associated costs of which differ) in a discriminating way” (Williamson, 1985: 18). Building on the behavioral assumptions of bounded rationality and opportunism, Williamson’s TCT argues that transactions will be assigned to governance structures based on three key attributes—asset specificity, uncertainty, and frequency. According to Williamson, the most important of these three attributes is asset specificity, which is the degree of specific investment involved in a transaction. An asset is specific to a particular transaction if its value in its next-best use (i.e., in a transaction with a different party) is lower than in the present transaction. The greater the difference between the value of an asset in its first-best and next-best use, the greater the degree of asset specificity (Klein, Crawford, & Alchian, 1978; Williamson, 1979).

When exchange hazards are not significant or are negligible—at the limit, when the assets that underpin and facilitate a transaction are generic—spot markets offer the least costly form of governance. Markets provide strong incentives for effort and for autonomous adaptation, and parties incur few or no private setup costs. Since the exchange relies on generic assets, disputes that might arise between transacting parties can be settled at low cost by exiting the exchange. In contrast, when assets are transaction-specific, hierarchy is the least costly governance solution. Although hierarchy entails high private fixed setup costs and reduces incentives to maximize outputs, it facilitates a coordination of investments and activities that is challenging to manage through markets. Within a hierarchy, authority (“fiat”) can ultimately settle disagreements about the nature and allocation of tasks. These different governance arrangements are also supported by different legal regimes, which range from court enforced contract law for some market transactions to internal enforcement for hierarchy (Masten, 1988). In equilibrium, organizational actors are predicted to choose the appropriate organizational form to govern a given transaction. We summarize the basic proposition of Williamson’s TCT framework and its key components in Figure 1.

Over the years, this basic proposition has evolved in a number of ways, driven both by the application of TCT ideas to new types of transactions and by broad changes to modern economic activity. International business scholars, who had independently begun work on firm-boundary issues (Buckley & Casson, 1976; Rugman, 1981), were early movers in integrating and extending these ideas to inform scholarship on international expansion and the structure of the multinational enterprise (Hennart, 1982). Perhaps not surprisingly, given the importance of transferring knowledge (rather than specialized physical assets) across geographic locations, issues of behavioral uncertainty became particularly prominent in this literature. Related to this, and to deal with the growth of the knowledge-based economy, strategy scholars introduced appropriability—which refers to the degree to which an economic actor can protect its knowledge from leakage to other
parties (Teece, 1986a)—as another important factor that influences governance choices (e.g., Oxley, 1997; Silverman, 1999; Teece, 1986a). As the business world experienced an explosion in collaborative activity, TCT sharply increased its focus on governance modes that incorporate aspects of both market and hierarchy. This led to many insights, but also resulted in the generation of a number of alternative views. For example, some have viewed alliances and other “hybrids” as a distinct governance mode (Williamson, 1991). Others, however, have distinguished between governance mechanisms (the price system and hierarchy) and institutions (markets, firms, and hybrids), and viewed alliances and hybrids as institutions, rather than as a distinct governance mode, that use different mixes of prices and hierarchy (Hennart, 1993, 2013). We further discuss these advances and alternative views later in our review.

THE BEHAVIORAL ASSUMPTIONS

Bounded Rationality

The first assumption underlying TCT is that of bounded rationality. This assumption is broadly related to that of many perspectives rooted in economics—namely, that there are costs to the collection and analysis of information, and consequently different actors will have access to different sets of information (e.g., Arrow, 1974; Holmstrom & Milgrom, 1991). TCT’s assumption of bounded rationality goes beyond information costs to recognize that agents have limits to their analytical and data-processing abilities, and that therefore these agents experience constraints in processing information and in formulating and solving complex problems even when information is available.2 Put simply, economic actors are “intendedly rational, but only limitedly so” (Simon, 1957: xxiv). This concept of bounded rationality is distinct from both irrationality and hyperrationality. Of particular relevance to the theory is the idea that, if rationality were not bounded—that is, if economic actors could costlessly anticipate every future contingency—then

---

2 If the only thing that matters is the cost of information, then mechanisms that lead to information revelation can solve all problems. However, if there are also limits to people’s ability to process and analyze information, then such mechanisms alone will not be sufficient.
they could write complete contracts covering any potential outcome. In this situation, there would be no transaction costs associated with contracting. Consequently, all transactions could be effected through the market (Williamson, 1981).³

Opportunism

The second assumption underlying TCT is that of opportunism, which Williamson (1975: 255) famously defined as “self-interest seeking with guile.” In other words, actors do not always share full information, provide objective assessments of likely outcomes, or behave cooperatively during the execution of economic exchanges. As with bounded rationality, opportunism is consistent with many economists’ view of self-interested behavior, in which actors take actions that maximize their payoffs (e.g., Grossman & Hart, 1986; Hart & Moore, 1990; Holmstrom & Milgrom, 1991), but goes beyond this to recognize that actors may also make acts of omission or commission that tilt payoffs in their favor. This has a pervasive impact on economic organization. In a world without any risk of opportunism, actors could always contract through the market and simply agree to “work things out” as future events unfold. Governance would play no role in a world in which all people will honor their promises all the time. But in a world with opportunism (coupled with bounded rationality), market exchange can be fraught with hazards.

The assumption of opportunism has been more controversial than that of bounded rationality (Williamson, 1993), and has sometimes been construed to imply that most or all people are prepared to lie, cheat, and steal most or all of the time (e.g., Perrow, 1986). Several prominent scholars have argued that this assumption is actually damaging to the business world, because it encourages managers to assume the worst about exchange partners or, worse, seek to preemptively opportunistically themselves (Ghoshal, 2005; Ghoshal & Moran, 1996; Pfeffer, 2005). We note that, despite Williamson’s occasional use of a florid phrase (oftentimes a quotation from another social scientist), TCT does not propose that all actors behave opportunistically, but rather that some small proportion of actors will behave opportunistically some small proportion of the time, and that it is difficult to predict who will be opportunistic, and when. This small possibility of opportunistic behavior can be enough to cause market exchange to break down.⁴ Further, TCT scholars have not celebrated the existence of opportunism. As Ketokivi and Mahoney (2016: 129) pointed out, management scholars “spend a lot of time thinking about [and researching] insurable risks. . . . Does this mean that [they] condone fires and other calamities? Of course it does not.” In analogous fashion, TCT scholars think about transaction risks in order to devise safeguards that will allow economic exchange to

³ Absent bounded rationality, the cost of organizing activities in a hierarchical manner might also become negligible. Hence, in such a scenario, both hierarchy and the price system could efficiently organize interdependencies, making the choice of governance irrelevant (Hennart, 1982).

⁴ Prior work in economics had highlighted the problems of moral hazard and adverse selection with specific respect to insurance (e.g., Arrow, 1971); however, the field had not yet grappled with the underlying issue—that people might not always truthfully reveal their “type” (e.g., health) or honor pledges (e.g., to exercise or forgo smoking)—in other contexts. Williamson (1975: 5) was explicit that “the insurance problem [...] is really a paradigm for studying the employment relation, vertical integration, and competition in the capital market.” He defined opportunism thus: “Opportunism is an effort to realize individual gains through a lack of candor or honesty in transactions. It can take either of two forms. The most commonly recognized is the strategic disclosure of asymmetrically distributed information by (at least some) individuals to their advantage. Original negotiations may be impaired on this account. The second type manifests itself during contract execution and renewal. The impossibility of extracting what can be confidently regarded as self-enforcing promises to behave ‘responsibly’ requires that agreements be monitored and may pose problems, due to first-mover advantages, at the contract renewal interval” (Williamson, 1973: 317). This was later elaborated as “the strategic manipulation of information and misrepresentation of intentions.” Without the latter of these, “self-enforcing promises to the effect that ‘I solemnly pledge to execute this contract efficiently and to seek only fair returns at the contract renewal interval’” would be sufficient to eliminate transaction costs. However, “at least some of the agents who accede to such terms do it casually, in a self-disbelieved way. Since these types cannot be distinguished ex ante from sincere types,” contractual hazards arise (Williamson, 1975: 26–27; for a more complete discussion, see Williamson [1993]).
occur even in the face of these risks. Simply put, opportunism is important to this thought process.

In the international business literature, Verbeke and Greidanus (2009) offered a broader version of opportunism, which they called bounded reliability. They positioned this as an envelope concept, with two main components: (a) the Williamsonian dimension of strategic opportunism, but also (b) non-strategic preference reversals, which are _ex post_ reordering of commitments that were made in good faith _ex ante_, due to changes in priorities and over-commitment. The authors asserted that most cases of opportunism are due to the latter rather than to the former. In other words, opportunism arises because “most actors are reliable, but only boundedly so” (Kano & Verbeke, 2015: 98), and their failure to fulfill their commitment is more often due to reprioritization and overcommitment than to a deliberate intent to deceive and cheat. Verbeke and Greidanus (2009) argued that bounded reliability is a more accurate behavioral assumption than Williamsonian opportunism because it does not require making the assumption that human beings are strategically untrustworthy. Instead, “they have a propensity to make imperfect effort to make good on commitments” (Verbeke & Greidanus, 2009: 1490).

We believe that further exploration of bounded reliability is a productive direction to pursue. As discussed above, the concept is more general than strategic opportunism, which may allow it to explain a broader or different set of governance decisions. It may more faithfully reflect what happens in practice in some settings—particularly for a transaction that takes place over a long period where there is a risk that one or both of the parties will change their priorities. This is a transactional hazard, but it does not require _ex ante_ premeditation. One key issue for work in this area is under which circumstances bounded reliability will lead to predictions that differ from those of opportunism, and why.

### Revisiting the Behavioral Assumptions

Many areas in management research have increasingly embraced a more behavioral focus. For example, scholars have looked at the micro-foundations of dynamic capabilities by focusing on the sensing and seizing of opportunities (Teece, 2007). Considering that the assumptions at the foundation of TCT are behavioral in nature, and that bounded rationality, in particular, clearly acknowledges limits on the rationality of decision making, it would appear that TCT is well-positioned to benefit from this trend. However, relatively few TCT scholars have turned their attention to fleshing out the behavioral aspects of the theory. A small body of work has explored the role of risk preferences and the perception of behavioral uncertainty (i.e., likelihood of opportunism) in refining the TCT logic (e.g., Buckley & Chapman, 1998; Chiles & McMackin, 1996; Weber & Mayer, 2014). One key insight from this work is that the way in which a contract is framed can influence perceptions of cooperativeness. For example, Weber, Mayer, and Macher (2011) noted that contractual clauses concerning early termination and extendability are essentially identical as safeguards—a five-year contract that allows for a second five-year term if both parties opt in is the same as a 10-year contract that allows for either party to opt out at the five-year mark. The authors theorized, however, that these clauses frame the safeguard in distinct ways that can create different psychological stances toward the exchange, and they uncovered empirical regularities in the use of each clause that are consistent with the notion that actors use renewal clauses when they seek to emphasize the beneficial prospect of future cooperation.

As we elaborate below, we envision significant potential for scholars to further integrate insights from decision making and psychology into TCT. While the consideration and incorporation of behavioral factors may not alter TCT’s prescriptions for how managers _should_ make decisions, conducting...

---

Contracts will be violated because of self-interest, and can be violated because of the costs and ineffectiveness of surveillance [...] Transaction-cost analysis, assumes that "human nature as we know it", as Oliver Williamson puts it, is prone to opportunism with guile." Perrow’s description ascribes to Williamson a view about the pervasiveness of opportunism that Williamson does not evince. Ghoshal and Moran (1996: 14) quote Williamson as follows: "[Opportunism] allows for 'strategic behavior,' that is, 'the making of false or empty, that is, self-disbelieved, threats and promises in the expectation that individual advantage will thereby be realized (Williamson, 1975: 26)" (Ghoshal & Moran 1996: 17, emphasis added). This may appear to be a more provocative definition due to some of the word choices in the italicized phrase. It is worth noting that Williamson did not coin this phrase—rather, it is a quotation from Erving Goffman (1969: 88) within Williamson’s sentence, an attribution that did not make it into the Ghoshal and Moran article.

Williamson’s later work (e.g., Williamson, 1991) increasingly emphasized the challenge of coordinated adaptation in the face of unanticipated events that could differentially influence the private benefits of parties to an exchange, leading to what Verbeke and Greidanus (2009) might call reprioritization.
work on heuristics, attention, risk preferences, and biases (e.g., Ariely, 2008; Gigerenzer & Gaissmaier, 2011; Kahneman, 2011; Kahneman, Slovic, Slovic, & Tversky, 1982) would improve our understanding of how behavioral uncertainty is perceived and incorporated into decision makers’ analyses. This could enable us to better explain what kind of decisions managers do make, and why.

Separately, there exists some tension in the TCT literature due to differing interpretations of the extent to which rationality is bounded. On the one hand, bounded rationality is assumed to constrain actors’ ex ante forethought sufficiently for transaction costs to emerge. On the other hand, it is assumed to be insufficiently constraining to systematically hamper their ability to make appropriate governance choices in the face of future unknowns. The TCT literature has accumulated evidence showing that firms, on average, do indeed “get it right,” in the sense of making governance decisions that accord with the prescriptions of TCT (e.g., Geyskens et al., 2006; Zenger, Felin, & Bigelow, 2011). However, several studies have also highlighted that some firms appear to make choices that are suboptimal from a TCT perspective (e.g., Sampson, 2004). The extant TCT literature has not offered systematic explanations for the sources of these deviations, yet their presence and patterns might be partially predictable given the bounded rationality of decision makers. As we will highlight below, a greater emphasis on such deviations can enable scholars to further understand the performance implications of firms’ governance choices.

There are also unexplored opportunities regarding opportunism to integrate insights into TCT from behavioral economics and from different strands of psychology. As noted above, many critics have taken issue with the somewhat pessimistic view of human nature implied in the opportunism assumption in TCT. Yet, there is substantial evidence from work in behavioral economics and psychology that people place different values on objects, and have systematically different perspectives on events, based on their position relative to the object (e.g., an asset that one already owns or is considering procuring) or event (e.g., what side of a disagreement one is on) that is being considered. For example, prospect theory has highlighted the salience of loss aversion, which, among other things, results in an “endowment effect” in which individuals tend to value items that they already have more highly than is “rational” (Kahneman, Knetsch, & Thaler, 1990).

Alternatively, “self-serving” biases may systematically lead individuals to overvalue their contribution, or the economic harm they face, and undervalue that of others—put differently, biases may systematically alter views of fairness (Roth & Murnighan, 1982). To illustrate, Loewenstein, Issacharoff, Camerer, and Babcock (1993) conducted an experiment in which they provided all participants with identical testimony from an actual court case, in which an injured motorcyclist sued the insurer of the automobile driver who had collided with him. The only difference was that half of the participants were told at the start of the experiment that they were the plaintiff, while the other half were told that they were the defendant. Participants were asked to submit their assessment of a fair settlement and their prediction of what the judge would rule in the case. Their expectations diverged dramatically based on their putative identities, with those assigned to be plaintiffs submitting “fair” and “expected” settlement figures that were nearly double those submitted by those assigned to be defendants.

Why does this matter? The presence of such systematic biases may provide an alternative route to predicting disagreements between economic actors as circumstances change, in a way that does not rely on the conceptualization of opportunism as self-interest-seeking with guile. In a world with self-serving biases, or, for example, endowment effects that distort the subjective value of already-owned assets, it is possible that what one transactor views as opportunistic behavior is sincerely seen as fair behavior by the other. This brief discussion shows that integrating insights from behavioral economics and psychology may yield alternative or additional ways of thinking about the opportunism assumption that underlies the TCT logic, thus potentially mitigating some of the concerns that have been raised about its original conceptualization.

Finally, opportunism has received substantial scrutiny in the literature on trust. Scholars have proposed that trust can reduce transaction costs, primarily by reducing concerns about opportunism (Anderson & Narus, 1990; Bromiley & Cummings, 1995). Trust can relate to a party’s ability or to its benevolence or integrity (Mayer, Davis, & Schoorman, 1995; Ring & Van de Ven, 1994), which reflects concerns about its bounded reliability or strategic opportunism, respectively. Given empirical evidence showing that trust reduces transaction costs (Dyer & Chu, 2003; Zaheer, McEvily, & Perrone, 1998), several scholars have proposed that trust—particularly of the benevolence or goodwill type—can serve as a substitute to formal governance safeguards by reducing or eliminating the threat of opportunism (e.g., Gulati, 1995; Gulati & Nickerson, 2008; Lui & Ngo, 2004; an alternative view...
is that trust can complement effective contracting under certain circumstances (Puranam & Vanneste, 2009). As we further elaborate later, we see substantial opportunity for continued engagement between the trust and TCT literatures. For example, when TCT scholars have acknowledged differences in the level of trust, typically at the societal level, this has been treated as a "shift parameter" that affects the optimal switching point between governance modes (Oxley, 1999; Williamson, 1991). Advances in our understanding of trust may allow scholars to unpack the "black box" of this shift parameter. At the same time, TCT may be able to inform trust scholarship. Whereas much of the trust literature has focused on conditions that enhance an individual's propensity to trust others, a recent stream of work has focused explicitly on "trust accuracy"—the proper placement of trust in those who warrant it and the withholding of trust from those who do not (Schilke & Huang, 2018). As this literature seeks to flesh out the situational factors that enhance trust accuracy (e.g., Schweitzer, Ho, & Zhang, 2018), TCT may provide useful insights.

In Table 1, we provide an overview of our discussion in this section.

### THE KEY WILLIAMSONIAN CONSTRUCTS

#### Characteristics of Transactions

**Asset specificity.** As we have noted, an asset is specific to a particular transaction if its value in its next-best use and user (i.e., in a transaction with a different party) is less than its use in the current transaction. The greater the difference between the value of an asset in its first-best and its next-best use, the more specific that asset is to the transaction. Williamson (1985) attached particular significance to the role of asset specificity, arguing that it is the most important determinant of governance. In the absence of asset specificity, economic actors can...
simply terminate a transaction that has gone awry and find alternative partners with whom to transact. Early empirical TCT research focused heavily on asset specificity, with numerous studies finding systematic evidence that higher levels of asset specificity are associated with more hierarchical governance (for an overview, see David & Han, 2004). The bulk of these studies focused on the vertical integration decision, where asset specificity may be particularly important due to how supply chain disruptions can harm a firm. As we describe below, other transaction characteristics have exhibited more prominent roles in studies of other governance decisions, such as horizontal expansion.

Given the early emphasis on asset specificity, it is not surprising that this concept is the most developed of the three key transaction characteristics. Williamson (1985) originally proposed four distinct types of asset specificity: site specificity, physical asset specificity, human asset specificity, and dedicated assets. In addition to these, several scholars have proposed other types of asset specificity. For example, Masten, Meehan, and Snyder (1991) introduced the concept of temporal specificity, which refers to investments in assets that require timely completion or delivery to retain their value. Zaheer and Venkatraman (1994) discussed procedural asset specificity, which refers to the degree to which workflows and processes are customized to the transaction.

**Revisiting asset specificity.** As Joskow (1988) pointed out, these different types of asset specificity essentially capture the same set of underlying issues. Namely, they all center on settings in which there are only a small number of actual or potential exchange partners, transactions involve lock-in or dependence that is often asymmetric, and there are high incentives for opportunistic behavior. Nevertheless, considering the variety of forms that asset specificity takes, further conceptual refinement of each individual type can be useful. For example, whereas a physical asset can be legally bound to an owner and—since it has no volition of its own—will work equally well regardless of governance, a human “asset” has the freedom to leave an employer (in most parts of the world) and has the discretion to work hard, shirk, or divert effort in various directions. As we note below, many TCT studies have presented a generic argument pertaining to asset specificity without considering the institutional details of their setting; we believe that future TCT studies can benefit from introducing greater precision about how particular types of asset specificity drive governance choices.

Furthermore, advances in the market design literature (e.g., Roth, 2008) may allow additional conceptual refinement of the asset specificity construct and identify circumstances in which hold-ups may arise even without dependence on conventional specific assets. A key feature of a well-functioning market is “market thickness,” where a thick market maps closely onto TCT’s notion of an exchange for which there are many potential partners. In this literature, thin markets may arise for reasons other than asset specificity. For example, the digital transformation has had a profound impact on competitive strategy and market structure. Of particular note is the idea that network effects in two-sided markets (usually based around technological platforms) frequently lead a market to “tip”; although many platforms may compete, it is often inevitable that only a small number survive (Arthur, 1994; Hagiu, 2014). Thus, a firm may initially have a large number of platforms to work with, but then find its position fundamentally transformed into a small-numbers situation, without itself having made investments in transaction-specific assets. As TCT scholars work to extend the theory to platforms and ecosystems (as we recommend in our discussion on the future of TCT, below), this is an interesting area for further research.

We conclude the discussion of asset specificity by highlighting three common empirical challenges in the literature. First, many TCT studies that have employed secondary-source data have relied on research and development (R&D) or marketing intensity to proxy for the level of asset specificity. There are two concerns arising from this. First, it is not clear that such intensity measures accurately reflect the level of asset specificity for a firm. Although these studies have implicitly or explicitly assumed that high R&D expenditure generates technological skills or assets that are specialized to a particular
transaction, or that high marketing expenditures reflect customer-facing skills or assets that are useful for a narrow set of transactions, it has rarely been clear why this is necessarily the case (Brouthers & Hennart, 2007). Second, even if we assume that these measures do indeed reflect specialized assets, many studies have used industry-average measures of R&D and marketing intensity, rather than firm-level measures. It is true that data at the individual firm level may not always be available; however, the use of aggregate industry measures reduces measurement precision at the very least, and may also lead to biased findings (e.g., Garrett, 2003). Hence, we strongly encourage scholars to collect data at a lower level of analysis when this is feasible. More generally, given the vast number of studies that have relied on firm-level accounting measures to proxy for asset specificity, further advances in empirical research will benefit greatly from measuring asset specificity directly (examples of this are Stuckey, 1983; Hennart, 1988a), which is ideally done at the transaction level rather than the firm or industry level.

Second, although TCT research has benefited tremendously from the many studies that have used primary or survey data to measure asset specificity, there is room to improve these measures as well. For example, in several studies (e.g., Klein, Frazier, & Roth, 1990), survey questions have assessed the amount of investment made to support a given transaction, but not the extent to which these investments are specific; that is, whether, and to what extent, the investment would have residual value outside of the current transaction. Although there may well be a positive correlation between the absolute level of investment and its asset specificity, scholars can assess this more precisely—and could even measure the amount of quasi-rents at risk, as has sometimes been done in labor economics (Abowd & Lemieux, 1993; Guertzgen, 2009)—by collecting data on the value of the investment outside its current use; that is, its residual value. Such efforts would get closer to the essence of the asset specificity construct. We thus encourage scholars to design surveys with the precise theoretical construct more firmly in mind.

Third, the vast majority of TCT studies have focused on the asset-specific investments by one exchange partner, but have not considered the asset-specific investments made by the other partner. This can be a critical omission, since the other partner presumably also must agree to the governance mode selected for the transaction, and presumably has a preferred governance mode based on its own level of asset-specific investments. Further, recent TCT research has highlighted the role of “credible commitments,” or the reciprocal investment in transaction-specific assets by both partners (so that both have something to lose if the transaction falls apart), as a mechanism for supporting exchange (Ahmadjian & Oxley, 2005; Williamson, 1983). Given this, the focal partner’s commitment to the transaction, in terms of making asset-specific investments, might pave the way for the other partner to make similar investments. Failing to incorporate the asset specificity of the other partner—as a control variable, a hypothesized variable of interest, or otherwise, in the research design—may generate an omitted variable bias. While some studies have considered the asset specificity of both partners (e.g., Ganesan, 1994; McEvily, Zaheer, & Kamal, 2017; Poppo, Zhou, & Li, 2016), future research will benefit from more systematically using this comprehensive, integrated approach whenever data availability makes it feasible.

Uncertainty. Given Williamson’s emphasis on asset specificity, it is not surprising that other transaction characteristics received less attention in early TCT research. In particular, relatively little effort was made to conceptualize uncertainty or to distinguish among its different forms. Williamson himself offered an oscillating view of uncertainty. In early work, Williamson (1973: 318) focused on environmental uncertainty as a condition that would necessitate unforeseen adaptation:

The effects of uncertainty on economic behavior are extensive and pervasive. . . Of particular interest to us here is that, inasmuch as a full set of contingent claim markets is infeasible (by reason of bounded rationality), adaptive, sequential decision-making procedures need to be devised. Vulnerable as market exchange is to opportunism in these circumstances, hierarchical forms of organization are apt often to be favored.

By the next decade, however, Williamson (1981) took pains to distinguish between nonstrategic uncertainty and behavioral uncertainty. Building on Koopmans’s (1957) distinction between uncertainty due to a lack of knowledge about potential states of nature (“primary” uncertainty) and uncertainty due to a lack of knowledge about the actions of other actors (“secondary” uncertainty)—which he interpreted as

---

8 In fact, the resource-based view has typically looked at these same measures as proxies of a firm’s ability to diversify into new industries that benefit from modestly fungible technological or marketing resources (Montgomery & Hariharan, 1991; Silverman, 1999).
“nonstrategic” actions of others—Williamson proposed a third category, called behavioral uncertainty, which relates to an inability to predict the opportunistic actions of others. For Williamson (1981), behavioral uncertainty was the key characteristic that mattered. Yet, by the next decade, Williamson (1991) had returned to his original conception of environmental uncertainty as the relevant transaction characteristic. Why this retrenchment? One interpretation might be that Williamson’s formulation of behavioral uncertainty largely overlaps with the assumption of opportunism itself. If there is a general distribution of propensity for opportunistic behavior among potential transaction partners, then there is a latent level of behavioral uncertainty in all transactions. In such a case, the relevant variation across transactions is not behavioral uncertainty (which is always lurking) but environmental uncertainty, since greater environmental uncertainty increases the likelihood that an opportunistic actor will find a pretext to engage in unsavory behavior. Another interpretation might be that, given its focus on vertical integration, traditional TCT accordingly privileged the role of asset specificity (which may be particularly relevant for supply chain transactions), such that uncertainty’s main role was to exacerbate the contracting problems associated with asset specificity. As we discuss below, scholars who have extended TCT to horizontal expansion and the multinational enterprise have reinvigorated and further dimensioned the analysis of behavioral uncertainty as a stand-alone concept.

**Revisiting environmental uncertainty.** Numerous scholars have followed the path of assuming a background level of behavioral uncertainty, and consequently focused on the explication and measurement of environmental uncertainty. These studies have typically followed Williamson’s proposition that environmental uncertainty has a conditional effect on governance, rather than a direct one: “an increase in parametric uncertainty is a matter of little consequence for transactions that are nonspecific. . .[but] whenever assets are specific in nontrivial degree, increasing the degree of uncertainty makes it more imperative that the parties devise a machinery to ‘work things out’” (Williamson, 1985: 59–60). This view has been adopted and corroborated empirically by several scholars (e.g., Cuypers & Martin, 2007; Leiblein, Rouer, & Dalsace, 2002; Lu & Hebert, 2005). Leiblein (2003) concluded that empirical findings in the TCT literature are generally consistent with this view.

At the same time, some scholars have postulated a direct effect of environmental uncertainty on governance choice, and argued that such uncertainty makes it more difficult, if not impossible, to contractually specify *ex ante* the circumstances surrounding an exchange. For example, Walker and Weber (1984, 1987) looked at the impact of technological and demand uncertainty on make-versus-buy decisions. However, the results of studies that have investigated a direct effect of environmental uncertainty have been mixed and inconclusive (Cuypers & Martin, 2007). In addition, it has often been unclear whether the direct effects of environmental uncertainty on governance choice—when such effects are indeed found—are attributable to differences in transaction costs, or rather can be explained by other theoretical approaches in which environmental uncertainty plays a prominent role (e.g., real option theory). In sum, the literature has clearly established that environmental uncertainty has the predicted joint effect with asset specificity. In contrast, there is little systematic evidence that environmental uncertainty also has a direct effect on transaction costs, and consequently about how it might affect governance choice. This is not surprising considering that Williamson (1985) explicitly stated that environmental uncertainty should have no direct effect on governance choices. Therefore, we suggest that researchers who wish to continue this line of research, first and foremost, aim to extend and refine the TCT logic in order to theoretically derive a direct effect of environmental uncertainty; or, alternatively, seek to integrate TCT with complementary theories that predict such a direct effect.

**Revisiting behavioral uncertainty and appropriability.** Some scholars have grappled explicitly with the notion of behavioral uncertainty, often extending it beyond Williamson’s somewhat narrow emphasis on strategic behavior alone. For scholars who have placed more emphasis on bounded rationality and less on opportunism, Koopmans’s (1957) “secondary uncertainty”—the inability to predict others’ nonstrategic behavior—may also affect forms of organization. For example, the presence of irreducible knowledge differences between exchange partners (e.g., Conner & Prahalad, 1996) or the existence of bounded reliability (e.g., Verbeke & Greidanus, 2012) might be logically sufficient on their own to influence governance choice (although Foss [1996a, 1996b] offers a contrary view).

---

9 However, others have noted that even though behavioral uncertainty and opportunism are related, there can also be sources of behavioral uncertainty that are less welded to opportunism. We detail this view and the corresponding sources of behavioral uncertainty in the next section.
Our review of the literature indicates that many scholars have conceptualized and operationalized behavioral uncertainty in ways that encompass both Koopmans’s (1957) secondary uncertainty concept and Williamson’s (1985) narrower focus on uncertainty of a strategic kind. In addition, some scholars have conceptualized behavioral uncertainty in alternative ways. Rindfleisch and Heide (1997) and Geyskens, Steenkamp, and Kumar (2006) both viewed behavioral uncertainty as a performance evaluation problem, “that is, difficulty in ascertaining ex post whether contractual compliance has taken place” (Geyskens et al., 2006: 521).

The international business literature on the multinational enterprise has been especially innovative at highlighting and dimensionalizing behavioral uncertainty. This stems largely from this literature’s focus on horizontal integration, rather than vertical integration (Dunning & Lundan, 2008). Horizontal integration occurs when a firm integrates into production of a broader range of goods and services (i.e., expansion into new product markets, or expansion into new geographies with existing products). Whereas vertical integration largely turns on the dependence across different stages of production, which highlights challenges of transaction-specific assets, horizontal integration is more commonly motivated by contracting difficulties caused by information asymmetry, measuring costs, or other sources of behavioral uncertainty. For example, information asymmetry plays a crucial role in the market for intangible assets (Hennart, 1982, 2015a). For markets to be efficient, buyers and sellers must possess the same information. In the case of knowledge exchange, this is frequently not the case, since buyers often do not have precise knowledge of what they are buying (Arrow, 1962). Similarly, although in principle a high-reputation firm can license its brand to a firm in another country, it is sometimes difficult for the firm to monitor the care with which the licensee treats the brand. As a result, the market for intangible assets will sometimes break down, and firms will opt for a hierarchical solution rather than a market-based one (e.g., Davidson & McFetridge, 1982; Hennart & Park, 1994; Teece, 1985). We will revisit this issue below when discussing work on appropriability. Our brief discussion in this paragraph underscores the importance of factors other than asset specificity, such as information asymmetry.

Further, the focus on multinational activity in international business research also highlights cross-national differences that introduce variation in behavioral uncertainty. Thus, the different phenomenological focus in international business research invites a more central role for behavioral uncertainty, including the possibility of a direct effect on governance, rather than only a joint effect with asset specificity. Consequently, this literature has explored a wide range of sources of behavioral uncertainty. For example, numerous studies have examined cultural distance across countries, arguing that such distance increases behavioral uncertainty and thereby impacts governance choices (for an overview, see Beugelsdijl, Kostova, Kunst, Spadafora, & van Essen, 2018; Tihanyi, Griffith, & Russell, 2005). Some of these studies have proposed that cultural distance makes it harder ex ante to understand and predict an exchange partner’s behavior (e.g., Anderson & Gatignon, 1986; Maseland, Dow, & Steel, 2018), while others have argued that cultural distance increases ex post monitoring costs (e.g., Fladmoe-Lindquist & Jacque, 1995). However, as Maseland et al. (2018) pointed out, most studies have not been precise about how cultural distance creates behavioral uncertainty.

In addition to cultural distance, several other factors have also been considered. Cuypers, Ertug, and Hennart (2015) looked at ownership levels in acquisitions, and argued that linguistic distance between exchange partners and their level of English proficiency creates information asymmetry and thereby behavioral uncertainty, and that this results in acquirers’ taking a larger equity stake in targets. While Maseland and colleagues’ (2018) call for greater precision regarding behavioral uncertainty was made with specific reference to cultural distance, it can also be applied to the other factors that have been proposed. As examples of such precision, Henisz (2000) and Henisz and Macher (2004) demonstrated that variation in countries’ political institutions affects the level of “political hazards” faced by a multinational firm, and consequently influences the decisions of whether and by what mode to invest in a country.

It is not only international business scholars who have provided valuable insights into the factors that might drive behavioral uncertainty. As the knowledge-based economy has grown, a cadre of (mostly strategy) scholars have devoted increased attention to innovation, technological knowledge, and other intangible assets. This has led to the introduction of another important factor that drives behavioral uncertainty, and thus governance: appropriability. Appropriability has been discussed in the context of intellectual property and defined as
the degree to which an economic actor can protect its knowledge from leakage to other parties (Teece, 1986a). This is a special case of a more general requirement for market exchange, which is that market coordination—which works through the exchange of outputs—is only possible if these outputs are protected by property rights (Hennart, 1982). While some types of intangibles benefit from strong property rights and hence can be traded on markets (Arora & Gambardella, 2010; Gans & Stern, 2010; Silverman, 2019), this is not always the case, and the exchange of knowledge with poorly defined and protected property rights will be afflicted by appropriability hazards. As foreshadowed earlier, a key challenge stems from Arrow’s (1962) paradox of information: a seller of knowledge cannot credibly convey its value unless they reveal the knowledge to the potential buyer, but the buyer has no need to pay to acquire the knowledge once it has been revealed.

The patent system is meant to solve this problem by making knowledge public, while giving patentees property rights in the patented knowledge. Although such a system results in strong appropriability for some technologies and in some countries, therefore making market exchange of knowledge possible, the protection afforded by patents is imperfect—because patents can only cover explicit knowledge, because patentees must shoulder the costs of enforcing their own patents, and because the efficiency of public patent enforcement varies across countries (Levin, Klevorick, Nelson, & Winter, 1987; Veugelers & Cassiman, 1999). Consequently, a strong “appropriability regime” for intellectual property—for example, a patent and trademark system that establishes and records property rights and courts that provide their effective enforcement—can address this concern (Teece, 1986a), allowing markets for knowledge and reputation embedded in trademarks to function (Hennart, 1982, 2015a). Put differently, strong intellectual property protection can reduce the risk of behavioral uncertainty in much the same way that reduced environmental uncertainty reduces the risk of behavioral uncertainty.

Studies that have incorporated appropriability have typically found that the stronger the appropriability regime, the less likely hierarchy will be used to govern a transaction, as predicted by TCT. For example, in a study of technology-driven diversification, Silverman (1999) found that firms are more likely to exploit their technological assets by diversifying into a given industry when that industry is characterized by weak patent protection, and inferred that these firms rely on licensing contracts to leverage these assets in industries characterized by strong protection. Oxley (1999) found that when U.S. firms engage in international R&D alliances, they employ more hierarchical governance mechanisms when their partners come from countries with weaker patent protection. Oxley’s work highlighted that resource characteristics, other than transaction-specificity, might also be important drivers of behavioral uncertainty.

Based on these insights from the international business literature and the literature on innovation and technological knowledge, we see potential for substantial benefits from a systematic categorization of the factors that have been linked to behavioral uncertainty. If such a categorization is done in a conceptually coherent manner, it is likely to encourage future research to be more precise in its theorizing, improve our understanding of the behavioral uncertainty concept itself, as well as its effects, and allow for better integration between TCT and other theoretical perspectives. For example, the information economics literature (e.g., Arrow, 1984; Gibbons, Holden, & Powell, 2012; Spence, 1973, 2002) could inform us about the factors that create information asymmetry ex ante or ex post, and thereby also influence behavioral uncertainty. More sociological perspectives might help to explain how dissimilarity or homophily (i.e., the tendency to associate with similar others [Lawrence & Shah, 2020]) between exchange partners can affect the perception of behavioral uncertainty (how predictable the other party’s behavior is [e.g., Ertug, Gargiulo, Galunic, & Zou, 2018: 913]), and institutional theory might explain how contextual factors influence such uncertainty. Each of these ideas would have to be substantially developed, but we believe that these brief examples illustrate that a categorization of theoretical mechanisms is both feasible and valuable.

Finally, another important issue in recent theoretical and empirical work has been whether scholars should emphasize behavioral uncertainty in an objective sense, or as it is perceived by decision makers. Perhaps because of TCT’s roots in economics, it has generally been assumed (at least implicitly) that decision makers are able to accurately assess the level of behavioral uncertainty in a transaction. Accordingly, behavioral uncertainty is usually conceptualized and measured as an objective construct. However, in reality, managers make governance decisions based on their perceptions of behavioral uncertainty. Accordingly, rather than assuming that these subjective perceptions accord with objective reality, recent studies have increasingly discussed the relevance of perceived behavioral uncertainty (e.g., Boersma, Buckley, & Ghauri, 2003; Buckley & Chapman, 1998; Tsang, 2006). Such an approach is not inherently at odds with TCT, given that the theory’s assumption of bounded rationality allows
for decision makers’ perceptions to deviate from an objective assessment. Nevertheless, this does raise questions about which approach is more suitable, or when one approach is more appropriate than the other.

We see room for both approaches. When the objective of a study is to establish how firms should make decisions from a TCT perspective, and to verify whether, in fact, decisions are made in the predicted manner, we see objective measures as more appropriate. In contrast, when the objective of a study is to explain deviations from TCT predictions, or to improve our understanding of how managers make decisions that are based on TCT logic, perceptual measures might be more useful. Specifically, although behavioral factors per se do not have to alter—or yield direct implications about—how firms should make decisions, the work on heuristics, attention, risk preferences, and biases (e.g., Ariely, 2008; Gigerenzer & Gaissmaier, 2011; Kahneman, 2011; Kahneman et al., 1982) could improve our understanding of how behavioral uncertainty is perceived and then incorporated into decision makers’ transaction cost analyses and the decisions they in fact make. This would better explain the kinds of decisions that are actually made, and why. Hence, both approaches have their place in TCT research. We encourage scholars to devote more care and effort to justifying and explaining their choice between perceived or objective behavioral uncertainty, both conceptually and empirically. Although the distinction between objective and subjective assessment is relevant for all transaction characteristics, we believe that it is most salient for behavioral uncertainty, because this concept is more challenging to quantify and to objectively assess than is asset specificity or frequency.

In sum, although most early TCT work emphasized asset specificity, more recently behavioral uncertainty has been shown to be equally relevant, especially when TCT is applied beyond its original vertical integration context. For research on the effects of behavioral uncertainty to achieve its potential, it is important to further refine the concept and to be more precise in terms of the mechanisms that link specific factors to behavioral uncertainty, and consequently to firms’ governance choices.

**Frequency.** The third transaction characteristic is frequency, which refers to the extent to which transactions recur. Williamson (1985) proposed that the overhead cost of hierarchical governance should be easier to recover in the case of more frequently recurring transactions. As a result, the likelihood of hierarchical governance should be higher for more frequently recurring transactions than for less frequently recurring ones, all else being equal. Several observers have noted that, compared to asset specificity and uncertainty, frequency has received little attention in the TCT literature (e.g., Rindfleisch & Heide, 1997). The extent of this dearth becomes clear when we consider that two recent meta-analyses of TCT (David & Han, 2004; Geyskens et al., 2006) were not even able to include frequency in their analysis, owing to the lack of studies that include this construct. Furthermore, the few studies that have examined frequency have yielded mixed results. For example, Klein (1989) found a positive association between transaction frequency and the degree of vertical control. However, Anderson and Schmittlein (1984) did not find an effect of frequency on sales force integration (which represents a governance mode that is more hierarchical than contracting with independent sales representatives). Given the paucity of studies that have investigated this construct, the mixed results in the small body of literature that does exist, and the more skeletal theoretical rationale for this attribute, it may be time to redirect future efforts to better understand the frequency–governance relationship.

Notably, it may be feasible to reframe an obstacle to the study of frequency—the challenge of empirically distinguishing among Williamson’s (1985) proposed cost-amortization and other mechanisms that relate to the recurrence of transactions—as a fruitful way to disentangle alternative mechanisms. For example, repeated transactions are also associated with the development of trust (e.g., Gulati, 1995), and with reductions in information asymmetry and the development of routines (e.g., Reuer, Zollo, & Singh, 2002). Consequently, it is likely that these mechanisms are confounded with the TCT mechanism in studies that have relied on secondary data and traditional regression approaches. One avenue forward might be to refine the empirical methods to better identify the effects of frequency. For example, researchers might more directly capture the cost implications of more-frequent transactions, which have typically been measured as a latent construct, or control more rigorously for the other mechanisms that are likely at play. Alternatively, future research might consider exploring the effects of frequency in more controlled settings (e.g., Pilling, Crosby, & Jackson, 1994), where frequency can be manipulated.

---

10 Richman and Macher (2008: 6) noted that studies that employ a continuous measure of frequency tend to find no relationship between frequency and governance, while those that dichotomize frequency into one-time versus recurring exchange tend to find the predicted relationship.
to allow for a more precise examination of its effects, and to rule out alternative explanations.

To conclude, although transaction frequency is an inherent part of TCT, it has been largely overlooked in both theoretical and empirical work. As a result, it is not known whether this factor is determinant of transaction costs to the same degree as the other transaction characteristics in the TCT framework, which have been shown empirically to be important drivers of transaction cost and governance choices. Therefore, to the extent that future research grapples with frequency, we encourage more systematic attempts to theoretically distinguish among cost-, trust-, and routine-based effects, and to also explore novel empirical approaches to do so. It is possible that such future research will find no relationship between frequency and governance; however, this, too, is useful information, and we encourage researchers to report (and journals and editors to publish) nonsignificant findings to avoid the “file-drawer” problem (Rosenthal, 1979), and, if applicable, to implement appropriate methods to test for the existence of such null effects (e.g., Cuypers & Martin, 2010; Quintana & Williams, 2018).

**Governance Modes and Mechanisms**

*Markets, hierarchies, and hybrids.* Williamson (1975) initially proposed two discrete alternative governance modes that can be used to organize economic activity: markets and hierarchies. These modes differ in terms of the mechanisms that enable them to govern exchange and adaptation, with markets relying primarily on the price mechanism and hierarchies relying on administrative control. In subsequent work, Williamson fleshed out intermediate modes within these two polar forms, including long-term contracts (Williamson, 1979), relational contracting (Williamson, 1979, 1981), and alliances or “hybrids” (Williamson, 1991). The consideration of these other modes paralleled empirical TCT research that revealed a broad array of governance mechanisms, including the lending of specialized assets to suppliers (Monteverde & Teece, 1982), the matching of contract duration to the degree of asset specificity (Joskow, 1985), the signing of contracts that were demonstrably nonenforceable (Palay, 1985), reliance on strategic alliances (Oxley, 1997; Pisano, 1990), and the exchange of credible commitments or “hostages” to facilitate continuation of relationships (Ahmadjian & Oxley, 2005; Anderson & Weitz, 1992; Heide & John, 1988; Hennart, 1989). It also responded to criticism from non-TCT scholars that much economic activity does not take place at either of the extreme poles of spot-market and within-firm (e.g., Powell, 1987; Stinchcombe, 1990). As a result, by the late 1990s many TCT scholars recognized a wide range of discrete governance modes along the spectrum from market to hierarchy, with each mode being matched to a different set of transaction attributes (most commonly to different levels of asset specificity or appropriability).

An alternative view of governance modes was provided by Hennart (1993), who noted that the price mechanism rewards agents based on their outputs, while administrative control rewards agents on their behavior, or inputs. The price mechanism thus elicits high effort but encourages agents to cheat through quality-shading or other means, while administrative control elicits more harmonious behavior but encourages agents to shirk.11 These methods thus incur different costs, with the price mechanism requiring measurement of output to reduce quality-shading or other forms of cheating, and administrative control requiring specification and enforcement of behavioral constraints to reduce shirking.12 The relative cost of each organizing method will vary with the specific characteristics of the transaction. With this background, Hennart (1993) then proposed that the spectrum between market and hierarchy can include a large variety of governing institutions that embody different combinations of price and administrative control. On the horns of the dilemma between cheating and shirking, economic actors will judiciously choose an appropriate combination of price- and administration-based mechanisms to govern their transaction most efficiently. This approach suggests that a far greater set of transactions are governed by institutions that are different from pure market and pure hierarchy than has been traditionally recognized. Such an approach also implies that there might be considerable variation within modes, in terms of the bundles of governance mechanisms used, and that these bundles might exhibit richer variation in governance modes than has been recognized by the traditional market–hierarchy continuum. For example, in a study of the design of contractual alliances, Reuer and Devarakonda (2016) found that in some cases—that is, when contractual alliances rely heavily on administrative committees—contractual alliances

---

11 Shirking is defined here as a failure to abide by the letter and spirit of one’s employment contract. Shirking can also be defined as the difference in behavior between how employees behave and how they would behave if they were running their own business (Hennart, 1993: 535).

12 Since it is impossible to perfectly monitor either behavior or output, there is also a residual cost of some irreducible shirking under hierarchy and some irreducible cheating under the price system.
might actually have more hierarchical features than equity joint ventures, despite the fact that equity joint ventures have typically been considered more “hierarchical” than contractual alliances in the literature. This indicates the promise of more careful theorizing and more in-depth empirical investigation of the properties of different governance alternatives (or “institutions” in Hennart’s terminology). It is also worth highlighting that this view does not see markets as the default option, to be replaced by hierarchy only when markets fail. Instead, it proposes that a full understanding of the choice of governance requires one to consider simultaneously the factors that lead to market failure and those that lead to firm failure (Hennart, 1982, 1993).

**Relational governance.** In addition to the governance mechanisms we discussed above, scholars have proposed another form of governance—namely, relational governance. We will briefly discuss the origins of the relational governance view and touch upon the alternative views on how relational governance might fit within the TCT framework.

Drawing on legal work by MacNeil (1974, 1978), among other inspirations, Williamson (1979) introduced the concept of relational contracting into TCT and considered it as a distinct form of governance that relies on “private ordering”—resolution of disputes by the two parties to an exchange, rather than by pursuing legal recourse via third-party courts. Relational contracting emphasizes the ongoing nature of transactions and recognizes that these ongoing transactions are embedded in relationships (for a sociological interpretation, see Granovetter, 1985). This view has drawn the attention of numerous scholars, and substantial literature on relational governance has since emerged (e.g., Carson, Madhok, & Wu, 2006; Dyer & Singh, 1998; Zaheer & Venkatraman, 1995). This has also been a beneficial point of contact with non-TCT literatures. For example, research on relational embeddedness has offered valuable insights by demonstrating the ability of relational embeddedness to enhance formal governance (e.g., Hoetker & Mellewigt, 2009; Poppo & Zenger, 2002) and the ability of overall network structure to moderate the need for hierarchical mechanisms in an alliance (Robinson & Stuart, 2007).

As the literature on relational governance has expanded, some diverging views have emerged that challenge TCT in ways that have substantial implications for work that seeks to combine the logics of TCT and relational governance. Some scholars have proposed that network embeddedness or trust can even obviate the need for TCT-style governance entirely, and have suggested that networks provide a way of organizing economic activities that is altogether distinct from markets and hierarchies (e.g., Powell, 1990). Hence, they have suggested that relational governance is a distinct mechanism that does not lie on the continuum of governance structures proposed in TCT.

Alternatively, other scholars have taken a rather different, and less radical, view that allows for easier theoretical integration of social mechanisms into TCT. They have viewed the phenomena that relational governance scholars have focused on (for example, trust, embeddedness, or reputational concerns) as mechanisms that reduce behavioral uncertainty and thereby affect governance choices (e.g., Argyres & Mayer, 2007; Malhotra & Lumineau, 2011). For example, Robinson and Stuart (2007) studied the network of alliances among U.S. biotechnology firms and found that alliances between firms that are more proximate in the network, or that involve firms that are central in the network, entail less hierarchical governance, *ceteris paribus*. They interpreted this as evidence that the ability to send and receive information throughout the network facilitates the use of a firm’s cooperative reputation as a disciplining device, reducing the need for formal hierarchical control. More generally, such variation in social mechanisms can act as a “shift parameter” that affects the relative costs of governance modes (Williamson, 1991). The control-mechanism view has been shared, for example, by Hennart (2015b), who used *guanxi* in China as an example to demonstrate that TCT can accommodate relation-based (or relational) governance, arguing that this kind of understanding and discussion of relational and social mechanisms—that is, within the TCT logic—also makes it possible to evaluate the pros and cons of different types of social enforcement in an objective manner.13 This view also highlights the relevance

---

13 *Guanxi* is the “the cultivating of personal relationships through the exchange of favors and gifts for the purpose of obtaining goods and services, developing networks of mutual dependence, and creating a sense of obligation and indebtedness” (Standiford & Marshall, 2000: 21). Relatedly, a firm can set up internal teams so that peer pressure can be marshaled to curb employee opportunism, which is one well-known feature of Japanese management (Kenney & Florida, 1993). In addition, market actors may unilaterally internalize generalized morality—inculcated through schooling and the family—and professional standards, taught through training and education (Hennart, 1991; Ouchi, 1981). Differences in the extent of such internalization relate to the shift parameters that will alter the comparative costs of different governance structures (Williamson, 1991), such that, for example, nonfirm organization might be used for higher levels of asset specificity in Japan than in the United States.
Many TCT studies have presented a generic argument pertaining to asset specificity, without considering the details of their setting and the characteristics of the investment that is made.

Studies using secondary data to capture asset specificity have often used measures that are potentially problematic. Namely, R&D and marketing intensity data, often measured at the industry level, have been commonly used to operationalize asset specificity, which might have led to inaccurate findings.

Studies using primary or survey data have often failed to directly capture the essence of the asset specificity construct, since they have assessed the amount of investment that is made but not the extent to which these investments would have residual value outside of the current transaction.

The vast majority of TCT studies have focused on the asset-specific investments made by one exchange party only, ignoring the asset-specific investments made by the other party. Empirically, this may have generated an omitted variable bias and theoretically may have constrained researchers from addressing more dyadic questions.

Digital transformation and other economic trends have a profound impact on the types of investments and commitments firms make. TCT has insufficiently explored how this might affect existing conceptualizations of the asset specificity construct.

The international business literature has proposed various sources of behavioral uncertainty that might matter for governance choices. However, there has been a lack of systematic categorization of these factors.

In empirical work there has been a lack of clarity about when and whether researchers should use objective or subjective measures of behavioral uncertainty.

Frequency has received little attention, even though it is an inherent part of classic TCT. Empirical operationalization has varied between continuous and binary (one-time vs. recurrent transactions) measures of frequency, with different patterns of results.

It is difficult to disentangle Williamson’s (1985) proposed cost-amortization mechanism, which would favor hierarchical governance of recurrent transactions, from alternative mechanisms (e.g., trust, the development of routines) that would favor nonhierarchical governance of transactions.

Future work should place more emphasis on identifying what type of asset specificity is relevant in the specific context that is studied and be more precise about how that particular type of asset specificity drives governance choices.

Future empirical research on asset specificity would benefit considerably from using finer-grained data (i.e., at the firm level or the transaction level itself) and more precise and direct proxies for asset specificity (e.g., Stuckey, 1983).

Researchers collecting primary data should design surveys with the precise theoretical construct more firmly in mind, and also collect data on the value of the investment outside the current use—i.e., its residual value—or even measure the amount of quasi-rents at risk.

Where possible, future studies should aim to collect and use data on all the asset-specific investments of all exchange parties. This would reduce the likelihood of biased findings, and would also allow the study of additional theoretical issues such as “credible commitments” and the dynamics of the commitments made by all parties in the exchange.

Scholars should assess more systematically whether new types of investments fall within the existing categories of asset specificity, or whether such investments constitute a new category altogether and require conceptual refinements of the asset specificity construct.

Future research could theorize more precisely about, and improve our understanding of, the behavioral uncertainty concept by categorizing, in a theoretically coherent manner, the various mechanisms that have been linked to behavioral uncertainty.

In deciding whether to use subjective or objective measures, researchers should carefully consider the aim of the particular study: When the aim is to establish how firms should make decisions from a TCT perspective, objective measures might be more appropriate; when the aim of a study is to explain deviations from TCT predictions, or to improve our understanding of how decision makers actually end up making decisions that are based on TCT logic, perceptual measures might be more useful.

More work is needed that systematically considers the effect of frequency.

Researchers should try to use more refined empirical methods to support better identification of the effects of frequency. For example, they should aim to (a) directly capture the cost implications of more-frequent transactions, which are typically measured as a latent construct; (b) control more systematically for other mechanisms that are likely at play; and (c) use more controlled settings where frequency can be manipulated to allow for a more precise examination of its effect and to rule out alternative explanations.
of TCT to the study of governance in parts of the world where transactions are less commonly governed by contracts enforced by courts.

In Table 2, we provide an overview of our discussion in this section.

### THE LOGIC

Above, we discussed the essential building blocks of TCT. This discussion allows us now to move on to examine the issues that pertain to the logic, and how the logic can be extended.

### A Pluralistic View of TCT

Although TCT is typically most closely associated with Williamson (1975, 1985, 1991, 1996), the sheer amount of research inspired by the theory has generated extensions and alternative concepts that are not always recognized or used across different topic areas. Within the strategy field, the most prominent extension has been the introduction of appropriability as a relevant transaction characteristic, particularly for knowledge-based transactions (Oxley, 1997; Teece, 1986a). Important work has also included deeper recognition of the importance of credible commitments, usually in the form of mutual investments in specific assets (Ahmadjian & Oxley, 2006 [although early work in this area came from marketing; e.g., Heide & John, 1988]), and a rigorous assessment of interactions between formal and informal governance mechanisms (e.g., Poppo & Zenger, 2002).

In the international business field, key advances have emphasized the role of behavioral uncertainty as a standalone transaction characteristic of consequence even in the absence of asset specificity (Anderson & Gatignon, 1986; Hennart, 1982, 1993), as well as the range of intermediate governance forms and the degree to which price and administrative mechanisms may be judiciously combined to govern transactions (Gatignon & Anderson, 1988), and the question of how a firm will choose to internalize a transaction—by building its own in-house operation or by acquiring an existing operation (Anderson & Gatignon, 1986; Hennart & Park, 1993). The version of TCT that has been more commonly used in the international business literature has also been less willing to assume that markets are the default governance structure, and has paid more attention to outlining the ways in which firms might “fail,” perhaps because the well-known challenges to multinational expansion (such as the liability of foreignness) highlight the challenges of expanding firm boundaries. Presumably, the TCT research in this field highlighted these issues because of the classic international business-related phenomena under study: how to best exploit competitive-advantage-generating assets in a new geographic market that has its own distinct institutional background and culture. Given the wide range of mechanisms that multinational enterprises have used to manage subsidiaries or overseas relations, especially when faced with high information costs in monitoring those operations and with the prospect of negative

### TABLE 2 (Continued)

<table>
<thead>
<tr>
<th>Governance Modes and Mechanisms—Markets, Hierarchies, and Hybrids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issues</td>
</tr>
<tr>
<td>A number of alternative views on governance have emerged that have implications for “hybrids” and the market–hierarchy continuum.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Governance Modes and Mechanisms—Relational Governance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issues</td>
</tr>
<tr>
<td>There are diverging views on whether relational governance is a third type of governance mechanism alongside markets and hierarchies, or whether it refers to a set of social mechanisms that influence behavioral uncertainty within the TCT framework. Many studies in the TCT literature have focused on the Anglo-Saxon world, where contracts and courts play a pivotal role in governing transactions.</td>
</tr>
</tbody>
</table>

2021 Cuypers, Hennart, Silverman and Ertug
spillovers to operations in other countries if, say, a brand is damaged in one host country, it is not surprising that international business scholars have devoted attention to these issues as they have worked to extend TCT (Hennart, 1991). At the same time, strategy scholars could also benefit from more thoroughly grappling with these ideas, particularly with respect to product-market diversification and to reputational spillovers across product lines, which are relevant for strategy research.

The above extensions and alternative concepts have yielded valuable insights that have substantially improved our understanding of the boundaries of the firm and firms’ governance choices. However, several of these contributions have not widely diffused beyond their own fields and streams of research. Considering such extensions allows us to think about opportunities regarding how TCT, as a whole—with respect to its use in different areas of research—can be advanced. Integrating and reconciling the logics underlying these extensions and alternative views would lead to a more complete version of TCT. Figure 2 provides our sense of promising extant and future approaches to TCT.

A More Dynamic TCT

TCT has been criticized for being static (e.g., Langlois, 1992; Zenger, Felin, & Bigelow, 2011). Even though there has been progress in conceptually clarifying how TCT can be used in a more dynamic way (e.g., Williamson, 1999), empirical research has rarely taken a dynamic approach. This is an increasingly salient shortcoming of the TCT literature, since the contexts in which many firms make decisions have become more dynamic and turbulent, and also given the general trend toward more longitudinal, dynamics-sensitive research in the management field. Accordingly, we discuss how TCT can be used in a more dynamic way and provide our perspective of how future empirical research can address a number of questions that are particularly relevant in today’s competitive landscape by taking a more dynamic approach.

We will start by elaborating on two dynamic, or intertemporal, issues within the existing TCT framework. First, Williamson (1975) highlighted the importance of the “fundamental transformation” in governance decisions. This transformation refers to the fact that the buildup of asset-specific investments

---

**FIGURE 2**

Current and Future Approaches to TCT

<table>
<thead>
<tr>
<th>BEHAVIORAL ASSUMPTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bounded Rationality</td>
</tr>
<tr>
<td>Opportunistic Behavior</td>
</tr>
<tr>
<td>Bounded Reliability</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TRANSACTION CHARACTERISTICS AND CHARACTERISTICS OF INTERDEPENDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Market thickness</td>
</tr>
<tr>
<td>o Asset specificity</td>
</tr>
<tr>
<td>o Network effects</td>
</tr>
<tr>
<td>o Other sources of thin markets</td>
</tr>
<tr>
<td>• Behavioral uncertainty</td>
</tr>
<tr>
<td>o Information asymmetry</td>
</tr>
<tr>
<td>o Appropriability</td>
</tr>
<tr>
<td>o Output measurement costs</td>
</tr>
<tr>
<td>o Behavior measurement costs</td>
</tr>
<tr>
<td>• Environmental uncertainty</td>
</tr>
<tr>
<td>• Frequency</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ORGANIZATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Governance Mechanisms and Institutions</td>
</tr>
<tr>
<td>Alternative Views</td>
</tr>
<tr>
<td>• Markets</td>
</tr>
<tr>
<td>• Hierarchies</td>
</tr>
<tr>
<td>• Hybrids</td>
</tr>
<tr>
<td>o Contractual provisions—Control vs. coordination</td>
</tr>
<tr>
<td>• Price system (output control)</td>
</tr>
<tr>
<td>• Hierarchy (behavior control)</td>
</tr>
<tr>
<td>o Pure markets</td>
</tr>
<tr>
<td>o Pure hierarchies</td>
</tr>
<tr>
<td>o Institutions that mix markets and hierarchies</td>
</tr>
<tr>
<td>(e.g., hybrids, relational governance)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PERFORMANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A function of appropriate alignment of governance mechanism(s) with transaction characteristics given the micro (property rights, psychological framing) and macro (polity, culture) contexts</td>
</tr>
</tbody>
</table>
during contract implementation results in a shift from large-numbers to small-numbers conditions, and to the possibility of one party holding up the other at contract renewal. An implication of this is that ex ante small-number conditions are not a necessary condition when a governance choice is made, and the anticipation of this fundamental transformation ex post should be driving governance choices ex ante. Hence, this is clearly an inherently dynamic aspect in the TCT logic as outlined by Williamson. Related to this point, we see three potential research opportunities. Incorporating the implications of this fundamental transformation requires foresight from managers, as well as an accurate understanding of the process, for them to make governance choices that are consistent with the TCT logic and appropriate in the long run. Extant research has not explicitly investigated the role played by managers’ foresight and their understanding of the fundamental transformation in their governance choices. In addition, this fundamental transformation is likely to be an endogenous process in some contexts. Hence, it would be useful to explore how firms’ ability to influence this process then affects their own governance choices, as well as the governance choices of other firms. Finally, a similar transformation might occur within hierarchies as well, which is a topic that has received only limited empirical or theoretical attention (Argyres & Liebeskind, 1999). For example, employees’ incentives might be transformed over time in similarly systematic ways because of their accumulation of firm-specific skills and knowledge. Thus, there are several research opportunities by which scholars can refine and extend the dynamism inherent in Williamson’s fundamental transformation.

Alternatively, several scholars have incorporated dynamics by studying the adaptation of governance modes over time. Williamson (1991) argued that TCT can be the basis of a comparative analysis that explains the adaptation of governance modes to changing circumstances. Following this line of reasoning, Nickerson and Silverman (2003) found that U.S. trucking firms adapted their governance choices in response to an exogenous shock to the conditions surrounding their transactions, but at varying rates. Reuer and Ariño (2002) examined whether changes in the environment affect the decision to renegotiate the governance structure in alliances, finding evidence that contractual terms are restructured to reduce misalignment caused by environmental changes. Brahm and Tarzijan (2014) demonstrated that Chilean construction firms adapted their governance structures and increased their extent of vertical integration after a legal shock that increased contracting hazards.

This small body of work presents a number of opportunities for future research. First, most studies have focused on discrete and substantial changes in governance mode (e.g., shifts from a market-based or hybrid governance mode to hierarchy-based governance). However, firms might also make finer-grained adaptations to their governance modes to minimize transaction costs that arise from changes in the conditions surrounding the transaction. For example, in response to contextual changes, firms might continue with a joint venture rather than move to a wholly owned subsidiary, but they might increase the levels of hierarchical control within the joint venture by altering the ownership stakes or reshaping the board of directors (e.g., McQuade & Gomes-Casseres, 1992). It would be worthwhile to investigate such finer-grained governance adaptations. Second, most studies have focused on governance adaptations in response to external changes, such as regulatory changes (e.g., Nickerson & Silverman, 2003). One reason for this is that an external change simplifies research design: with an exogenous shock, one has a natural way to address concerns about the endogeneity of governance choice. However, governance adaptations might also occur due to internal changes. For example, the nature of investments might change over time and become less specific, thereby triggering a need for governance adaptation (e.g., Hennart, 2009). Accordingly, future research could also investigate how such internal factors lead to governance adaptations. Finally, the adaptation process itself remains inadequately understood. Combining insights from the literature on organizational adaptation (e.g., Ring & Van de Ven, 1994) with the TCT literature might help us better understand when and how firms adapt their (misaligned) governance modes. For example, Reuer, Zollo, and Singh (2002) showed that prior experience can facilitate governance adaptation in general. Accordingly, further studies could investigate how experience in general, as well as different kinds of experience more specifically, could help to realign governance modes that are misaligned from a TCT perspective. The identification of other kinds of factors that facilitate or hamper realignment would also be relevant, both theoretically and practically; as an example, Argyres, Mahoney, and Nickerson (2019) proposed an approach that integrates adjustment costs, transaction costs, and opportunity costs.

In sum, we believe that TCT is by no means inherently static and that it can accommodate more dynamic approaches. Even though only a modest number of studies have applied TCT in a more dynamic way, our discussion and the examples above regarding future research indicate the considerable payoffs in continuing these efforts and taking them further.
Performance Implications

Management scholars have an inherent interest in understanding what drives firm performance. While TCT explicitly discusses performance implications (e.g., Williamson, 1985, 1999), the literature has predominantly focused on explaining strategic choices, and relatively few studies have looked directly at the performance implications of these choices (David & Han, 2004). There is a clear opportunity for more TCT work that studies performance and that clarifies how the theoretical logic pertains to performance.

TCT presumes that economic actors will enjoy performance benefits when they govern transactions appropriately—or, put differently, firms whose transactions are not properly aligned with appropriate governance structures will suffer performance consequences. The presumption that firms with misaligned governance structures will suffer performance consequences relies “in a general, background way on the efficacy of competition to perform a sort between more and less efficient models and to shift resources in favor of the former” (Williamson, 1985: 22). As a consequence of such competitive pressures, firms whose transactions are inappropriately governed should perform worse, compared to those firms whose transactions are appropriately governed, such that the former set of firms are pressured to adapt or risk being forced to exit a market.

The above implies that studying performance from a TCT perspective is both conceptually and empirically more complex than is the case for many other theories. Namely, rather than thinking in terms of how a given factor affects performance, studies that look at performance from a TCT perspective need to first conceptually clarify and empirically identify misaligned and properly aligned governance structures, and only then compare these in terms of their performance implications. However, such research comes with several empirical challenges (as we elaborate below). We observe two broad approaches that researchers have used to empirically investigate the performance implications of governance modes. One involves employing more sophisticated empirical approaches, such as two-stage regression approaches (e.g., Anderson, Dekker, & Van den Abbeele, 2017), often also taking advantage of external shocks (e.g., Hamilton & Nickerson, 2003). The other is to focus on predicted tradeoffs in performance that are associated with a governance choice, and then to seek evidence supporting or refuting these tradeoffs (without the presumption that one mode is universally better than another). We consider these two approaches in turn.

Regarding the first broad approach, which employs more sophisticated empirical approaches, Masten and colleagues (1991) used a two-stage-least-squares approach to measure transaction costs for sea-vessel construction. Through interviews and surveys within one shipbuilder, they collected information on the degree to which each of 74 components exhibited high asset specificity and complexity, whether it was made in-house or outsourced, and, for the in-house components, the amount of money spent governing the component’s production. Through the features of this econometric approach, they were able to infer the transaction costs associated with the outsourced components, and the expected costs had the outsourced parts been made in-house, or vice versa. Their results indicated that transaction costs comprised 14% of total production costs, and misalignment would have doubled these costs.

More commonly, scholars have explored the effect of misalignment on profits or survival. Nickerson and Silverman (2003) found that the over- or under-use of company drivers and company trucks was associated with lower return on assets and lower likelihood of survival for motor carriers in the post-deregulation U.S. trucking industry. Relatedly, Argyres and Bigelow (2007) found that misalignment in production of engines was associated with lower survival for automakers—but only after the “shakeout” stage of the industry life-cycle had begun. Gartenberg and Pierce (2017) leveraged the 2008 financial crisis to find that the vertical integration–performance relationship for banks was moderated by the presence of strong corporate governance practices. Bruce, de Figueiredo, and Silverman (2019) took a different approach, relying on environmental stickiness rather than shock. In a study of the governance of private R&D contracts funded by the U.S. federal government, they noted that close project oversight is feasible for the government only
when there are suitably expert government personnel available. Using this factor as an instrument, they used two-stage methods to measure the performance consequences of misaligned governance for projects, finding a substantial penalty in terms of patents generated.

A number of scholars (e.g., Anderson et al., 2017; Powell, 2014; Sampson, 2004) have relied on another type of two-stage approach. They first determined whether a governance choice is aligned with the context that surrounds the transaction, thereby arriving at a measure of misalignment, and then regressed their measure of misalignment against performance in a second stage. The measure of misalignment in these studies has typically been the residual from the first-stage regression (i.e., the difference between the observed governance choice and the estimated—what would have been the properly aligned—choice). While this approach is more broadly applicable, as it does not require a conveniently timed shock and sidesteps the often challenging task of finding a reliable instrument (e.g., French & Popovici, 2011), it does assume that researchers are able to identify the appropriate governance choice. This highlights the importance of ensuring that the first-stage model is properly specified (among other things, to limit the potential of omitted-variable bias) and to carefully consider the validity of the misalignment measure.

Regarding the second broad approach, which focuses on tradeoffs associated with governance choice, Novak and Stern (2008) noted that a close reading of TCT implies that markets should be particularly good at getting an arrangement right early on in an exchange, since parties will spend time to specify important matters in a contract, whereas hierarchy should be particularly good at managing change over time. They then collected longitudinal consumer report data on 112 automobile models, as well as data on whether each of nine major components for each model was outsourced or made in-house. Of particular note, once an automaker committed to internal (or external) production for a model, it retained that mode for the life of the model. Consistent with the implication of TCT, the authors found that models with more outsourced components outsized models with more in-house components at the beginning of their lives, but that, over time, the score for heavily outsourced models deteriorated, whereas the score for inourced models did not, falling below the insourced models after the third year. Thus, by looking at tradeoffs in a creative way, Novak and Stern (2008) offered a valuable corroboration of the TCT—performance relationship without wading into the endogeneity morass.

The above discussion highlights two important issues for researchers to consider when deciding on a research design: endogeneity of the governance choice and survival bias. As Shaver (1998) explained, it is reasonable to assume that a firm’s management team tries to make the best choice given a firm’s idiosyncratic strength, weaknesses, opportunities, and constraints. As a result, a researcher cannot easily examine the effect of a choice on performance, without having an adequate strategy to deal with such endogeneity issues. Of the approaches we discussed above, the one that relies on leveraging an exogenous shock might be particularly well-suited to deal with this issue. In terms of the second issue—that is, survival bias—TCT scholars have argued and shown that misaligned firms or subsidiaries might be forced to exit a market (Argyres & Bigelow, 2007). This highlights the importance of considering survival issues when sampling, or correcting for this bias empirically via a selection stage. Despite these challenges, which are also relevant for other theories that relate choices to performance, we are optimistic, given the growing set of econometric tools available to deal with these issues.

In sum, there remains much exciting work to be done on the performance implications of transaction alignment. For example, Forbes and Lederman (2010), Sampson (2004), and Anderson and colleagues (2017) looked at particular facets of performance, rather than overall profitability, and found evidence consistent with TCT predictions in their investigation of operational performance in the airline industry, innovative performance in R&D

15 As we also note below, a reasonable default assumption is that the firm’s decision makers will make the best alignment choice for their unique situation, so that what looks like misalignment to the researcher is plausibly due to information that is unavailable to the researcher. At the same time, given the assumption of bounded rationality, some managerial mistakes are to be expected. Recent empirical advances, such as propensity-score matching, that match firms based on a wide range of observables, arguably go a long way to address this concern.

16 Shaver (1998) compared the simple ordinary least squares approach to a two-stage approach that addresses endogeneity to demonstrate that the “conventional wisdom” in the early 1990s—that foreign-market entry by greenfield plant performs better than foreign-market entry by acquisition—was attributable to the fact that strong firms chose to enter by greenfield while weak firms chose to enter by acquisition, and thus the choice–performance relationship was a spurious correlation.
alliances, and residual risk in IT investments, respectively. Hence, while the bulk of the work on performance in the TCT literature has focused on profitability, these studies have highlighted the potential rewards of considering other types of performance as well. Related to this, and inspired by Argyres and Bigelow’s (2007) work, future research could investigate more thoroughly the conditions under which misalignment might lead to firm mortality, versus those for which the performance consequences of misalignment are more forgiving; as de Figueiredo and Silverman (2012) showed, relevant conditions may stem from governance decisions in related industries. Sparked by Gartenberg and Pierce (2017), future research could also focus on complementarities among governance features that jointly influence performance. Additionally, continuing in the direction explored by Novak and Stern (2008), scholars could focus on subtle differences in performance, not only to confirm or refute the predictions of TCT, but perhaps also to test its explanatory power compared to that of other theories (e.g., if property rights theory does not suggest the same pattern of early versus late relative performance, then this would be a way to distinguish between the implications of these theories). More generally, we see potential for more work that improves our understanding of how firm and contextual factors might aggravate or alleviate the negative performance implications of misalignment.

In Table 3, we provide an overview of our discussion in this section.

**THE FUTURE OF TCT**

Above, we provided an overview of the basic proposition of TCT, and its underlying assumptions, key theoretical constructs, and logic. We used this set of parameters to provide a roadmap for navigating the vast TCT literature and to critically assess it. Throughout these sections, we brought up a number of specific suggestions for future research. In this section, we identify three areas where TCT can be conceptually extended, and three areas where it can be expanded to the study of important and recently emerging phenomena. We elaborate on these extensions below, and provide an overview of this discussion in Table 4.

### TABLE 3

<table>
<thead>
<tr>
<th>Issues</th>
<th>Opportunities for future research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different versions of TCT have evolved: Work in strategy has focused largely on market failure and appropriability, while work in international business has put more emphasis on behavioral uncertainty and why firms might fail.</td>
<td>There is an opportunity to integrate and reconcile the logics underlying these extensions and alternative views to develop a more pluralistic version of TCT.</td>
</tr>
<tr>
<td>Theoretical and empirical research has not fully exploited opportunities to employ a dynamic approach to TCT.</td>
<td>Future research can address a number of questions that are particularly relevant in today’s competitive landscape by taking a more dynamic approach. For example, future research could:</td>
</tr>
<tr>
<td>- Investigate the role played by managers’ foresight and their understanding of the “fundamental transformation” in their governance choices.</td>
<td></td>
</tr>
<tr>
<td>- Explore the adaptation of misaligned governance modes over time, specifically looking at finer-grained governance adaptations, and at the effect of internal factors that yield such changes.</td>
<td></td>
</tr>
<tr>
<td>- Study the actual adaptation process to illuminate how firms adapt their governance modes.</td>
<td></td>
</tr>
<tr>
<td>The TCT literature has predominantly focused on explaining strategic choices, with less effort devoted to the performance implications of these choices.</td>
<td>Scholars can explore the performance implications of governance choices using a two-step approach:</td>
</tr>
<tr>
<td>- Step 1 consists of careful theorizing about what the appropriate choice for a firm is, and then incorporating this theorizing and identification in a research design, so that one can reliably predict what a firm’s governance choice should be in the first place. This will allow researchers to assess how well-aligned or misaligned the firm’s actual choice is.</td>
<td></td>
</tr>
<tr>
<td>- In step 2, the identified level of (mis)alignment can be used to predict performance differences.</td>
<td></td>
</tr>
<tr>
<td>Future research could also investigate the conditions under which the performance implications of misaligned governance choices for firms are more or less severe.</td>
<td></td>
</tr>
<tr>
<td>Panel A: Conceptual Extensions</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Theory and literature</strong></td>
<td><strong>Opportunities for future research</strong></td>
</tr>
</tbody>
</table>
| More engagement with the international business literature | - How do cultural and institutional factors influence the choice between the enforcement of transactions through relational governance versus through contracts?  
- How do cultural and institutional factors influence contract design?  
- How does institutional variation across countries influence the enforceability of contracts?  
- How do cultural and institutional differences across countries affect the reliance on contracts versus relational governance mechanisms?  
- How do linguistic factors influence the design and effectiveness of contracts?  
- How and when does prior experience influence governance choices?  
- How do capabilities and adjustment costs influence the incentives to adjust misaligned governance structures?  
- How do the control and coordination aspects of contracts, or other governance mechanisms, interact? |
| More engagement with the strategy literature | - How can apparent contradictions between alternative views in strategy and international business be resolved? |
| Reconciliation of the strategy and international business literatures | - How can TCT illuminate mechanisms that can enhance trust accuracy?  
- To what extent is competence trust strategic?  
- To what extent can insights from trust open up the black box of TCT’s “shift parameters”? |
| Engagement with information economics and signaling theory | - How do different types and sources of information asymmetry drive behavioral uncertainty? |

<table>
<thead>
<tr>
<th>2. Link to sociology—trust and opportunism; formal and informal organization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theory and literature</strong></td>
</tr>
</tbody>
</table>
| Engagement with sociology: | - To what extent, and where, is trust misplaced?  
- Can an understanding of opportunism help tease apart justified from unjustified trust?  
- Can TCT illuminate mechanisms that can enhance trust accuracy?  
- To what extent is competence trust strategic?  
- To what extent can insights from trust open up the black box of TCT’s “shift parameters”? |
| Trust and opportunism, behavioral reliability | |
| Formal and informal organization | |

<table>
<thead>
<tr>
<th>3. Link to psychology and behavioral economics—context and framing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theory and literature</strong></td>
</tr>
</tbody>
</table>
| Engagement with psychology: | - How does the framing of an exchange affect governance and execution?  
- How does the framing of a governance structure affect the way that it is perceived by actors?  
- How does the process by which an agreement is reached affect its subsequent execution?  
- How do heuristics, attention, risk preferences, and biases affect the extent to which decision makers evaluate the potential for opportunistic behavior? |
| Framing | |
| Biases and heuristics | |
| Attention | |

**Panel B: Extensions to New Phenomena**

<table>
<thead>
<tr>
<th>1. Technological advance—Platforms and governance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phenomenon or topic</strong></td>
</tr>
</tbody>
</table>
| Platforms, ecosystems, two-sided markets | - How does a platform’s ability to exclude parties affect the governance of exchanges that occur?  
- Does a technology platform represent an alternative form of governance?  
- How does the threat of market tipping affect market thickness and governance choices? |

<table>
<thead>
<tr>
<th>2. Technological advance—Artificial intelligence, machine learning, and transaction costs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phenomenon or topic</strong></td>
</tr>
<tr>
<td>Artificial intelligence and machine learning; the economics of prediction; technology and work</td>
</tr>
</tbody>
</table>
Conceptual Extensions

1. Further integration of strategy, international business, and institutional economics. As described throughout the previous sections, strategy and international business researchers have advanced TCT in ways that are relevant to their respective fields, but that can also be productively combined. We believe that there are many more opportunities for such a combination, which have the potential to lead to extensions in both fields. For example, consider contracts that span national borders. A large body of TCT research has explored domestic contracts, demonstrating that contract terms are tailored to the key characteristics of transactions (e.g., Argyres, Bercovitz, & Mayer, 2007). For example, Reuer and Ariño (2007) found that the level of asset specificity is related to the complexity of contracts and the adoption of enforcement contractual provisions. This research has devoted much less attention to how international factors influence contract design, yet such work offers substantial opportunity to dimensionalize contracting. Zhou and Poppo (2010) found that changes in the perceived legal enforceability of contracts over time within China have an impact on the relationship between asset specificity (among other factors) and contractual terms. Luo (2005) found that institutional variation within China influences the terms of contracts in international joint ventures. These studies have highlighted the potential for future research to explore how heterogeneity in legal frameworks and in contract enforceability across countries affect transaction costs and the design of contracts. In addition, recent TCT work has shown that linguistic differences between exchange partners influence governance choices; for example, how much ownership acquirers take in targets (e.g., Cuypers et al., 2015). This underscores the potential value in investigating how language differences, both between exchange partners and across countries, might influence contract design and effectiveness.

We note, however, that not all of the above-discussed views and extensions are complementary, and in fact some of them even appear to be contradictory. For example, Williamson’s (1991) and Hennart’s (1993) views on governance modes, and specifically on how to conceptualize hybrids, appear difficult to reconcile. Hence, another important area for future research is to juxtapose such alternative views to see whether such apparent contradictions can be resolved, and, if not, which conceptualization is preferable when. To be fruitful, these efforts need to clearly lay out and then rigorously investigate the arguments and boundary conditions with respect to each side. Such investigations would yield a better calibration (to the phenomena under consideration) and more precise exposition (of the logic, constructs, and boundary conditions) of each viewpoint, thereby making it more likely that the resolving of apparent contradictions and proposed syntheses or juxtapositions in fact do yield both theoretical and empirical advances. Overall, we strongly encourage more dialogue between different fields and streams of TCT research.

<table>
<thead>
<tr>
<th>Phenomenon or topic</th>
<th>Opportunities for future research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nationalism</td>
<td>How does nationalism affect opportunistic behavior and (the perception of) behavioral uncertainty?</td>
</tr>
<tr>
<td>CSR</td>
<td>What are the most suitable types of monitoring for different types of CSR activity?</td>
</tr>
<tr>
<td></td>
<td>What are the tradeoffs in encouraging CSR behavior?</td>
</tr>
<tr>
<td></td>
<td>Which organization form—for-profit firms, not-for-profit firms, nongovernmental organizations, etc.—is more effective for which types of social effort?</td>
</tr>
<tr>
<td>Public–private partnerships</td>
<td>When or how can government work effectively with private-sector organizations?</td>
</tr>
<tr>
<td></td>
<td>How does contract rigidity in the public sector affect the utility of PPPs?</td>
</tr>
<tr>
<td></td>
<td>How do rigid civil-service rules impact the effectiveness of “fiat” in the public sector?</td>
</tr>
<tr>
<td>Grand challenges</td>
<td>How should collaborations to tackle grand challenges, which are often more diverse and complex, be governed?</td>
</tr>
</tbody>
</table>
We also see potential to incorporate insights from other literatures in strategy, such as the resource-based view and organizational learning. For example, Mayer and Argyres (2004) emphasized the importance of experience in making optimal governance choices, and Argyres and colleagues (2012) referred to firms having a governance capability. This small but growing literature has suggested that capabilities and firms’ experience play a positive role in terms of firms opting for governance structures that are better aligned with the transactional characteristics. We see opportunities in continuing this line of work by bringing in additional nuances from these literatures. The work that has linked experience and capabilities with TCT has largely taken a static approach so far. Therefore, introducing the notion of dynamic capabilities might facilitate a more dynamic perspective, as such capabilities might help firms to identify the need to adjust misaligned governance structures and to mobilize the necessary resources to do so. Moreover, most work has implicitly assumed that the effects of experience and capabilities on governance choices are uniformly positive. However, in the organizational learning literature it has been well-established that experience can also have negative effects (e.g., Halebian & Finkelstein, 1999). Hence, we see opportunities for future research to explore the possibility of such negative effects in the context of governance choices as well.

Moving slightly farther afield, TCT has long drawn on economics, but can benefit from further integration with two related theoretical lenses, namely information economics and signaling theory. The appropriability literature in strategy, and work on behavioral uncertainty in international business, have clearly highlighted the relevance of information asymmetry in TCT. Incorporating additional insights from information economics will likely improve our understanding of the different types and sources of information asymmetry that might drive behavioral uncertainty. In addition, signaling theory (for a review, see Connelly, Certo, Ireland, & Reutzel, 2011) can offer valuable insights on how behavioral uncertainty can be reduced. Specifically, it might improve our understanding of the mechanisms that firms can use to reduce information asymmetry and thereby also the level of (perceived) behavioral uncertainty their transaction partners face. Hence, these two theoretical lenses have the potential to lead to a better conceptualization and understanding of the behavioral uncertainty construct.

2. Link to sociology: Trust—opportunism and formal—informal organization. Vibrant literature on trust has explored the ways in which interorganizational trust develops and the ways in which it can influence organizational arrangements. Definitions of trust have commonly included a willingness to “accept vulnerability based upon positive expectations of the intention or behaviors of another” (Rousseau, Sitkin, Burt, & Camerer, 1998: 395). Studies of trust in interorganizational arrangements have found that it is positively associated with collaborative performance (Dyer & Chu, 2003, McEvily, Perrone, & Zaheer, 2003; Zaheer et al., 1998), and can serve to alter the degree of formal governance used in the relationship (Gulati, 1995; Zaheer & Venkatraman, 1995). Gulati and Nickerson (2008) conceived of trust as a shift parameter that allows actors to reduce their concern about opportunism, and consequently use a less formal governance structure than would otherwise be necessary. Puranam and Vanneste (2009) proposed that different aspects of trust may either substitute for or complement formal governance, but the expectation either way is that greater trust facilitates relations and enhances performance.

A number of studies in this area have implicitly assumed that trust is well-placed (Gargiulo & Ertug, 2006). Consequently, a common assumption in the literature has been that trust increases with repeated interactions (e.g., Gulati, Lavie, & Singh, 2009). Theoretical models (e.g., Puranam & Vanneste, 2009) have tended to focus on the consistent building of trust (rather than its dissipation), perhaps through a series of modest interactions that build confidence, among the partners, allowing them to shift from calculative trust to a more faith-based, heuristic trust (McEvily, 2011). In such a world, it is not surprising...
that trust would appear to be a strong antidote to opportunism.

However, a small number of studies that have explored trust in conflictual settings have found evidence that the trust—opportunism relationship is more nuanced than conventional wisdom might suggest. Graebner (2009) examined the role of trust in acquisitions through eight case studies in which technology-based entrepreneurial firms were acquired by larger suitors. She presented noteworthy results: almost all of the executives from the acquiring firms engaged in deceptive practices, ranging from making empty promises of bridge financing to dissuading targets from pursuing venture financing or seeking alternative suitors (which would have raised the ultimate price of acquisition), through misrepresenting their outside options and reservation prices when haggling over acquisition price to making false commitments regarding post-acquisition behavior (e.g., promising not to fire or relocate employees when they fully intended to do so post-acquisition). Of particular note is the finding that these executives engaged in more deception when they believed that the target executives were more trusting. Finally, some of the target executives never realized that they had been deceived, even after the acquisition was consummated. This study provided striking evidence of opportunism in action, and of the inability of accomplished executives to always distinguish good faith from opportunistic behavior. It thus suggested that trust may not always overcome opportunism, and reminded us that trust may open the door for more severe opportunism (as noted by Granovetter, 1985: 491–492). This is buttressed by laboratory-experiment evidence that high trust may damage performance by constraining people’s willingness to monitor behavior (Langfred, 2004), and it may contribute to explaining why more experienced parties capture more value in acquisitions (Cuypers & Martin, 2017). An intriguing line of trust research has begun to explore “trust accuracy” (Feltchenauer & Dunning, 2010; Schilke & Huang, 2018; Schweitzer et al., 2018), and specifically the situational features—such as interpersonal communication and the availability of monitoring mechanisms—that lead parties to more accurately assess each other’s trustworthiness.

Relatedly, a few scholars have explored whether trust can repair relationships when they break down. Jap and Anderson (2003) studied buyer–supplier relationships and found that, after one side suspects the other of opportunism, trust is a less useful lever through which to rectify the relationship than is goal congruence or the presence of bilateral investments. Lucineau and Malhotra (2011) found that the reservoir of trust between two parties can enable them to resolve a dispute rather than severing the relationship, and that this reservoir of trust is itself a function of the contractual terms governing their collaboration.

For us, these studies open up a host of fascinating questions at the boundary of TCT and trust. To what degree can parties reliably count on their perceptions of each other’s trustworthiness to govern their relationship? How frequently is opportunistic behavior actually detected by a trusting party? Does trust of another elicit reciprocal trusting (and trustworthy) behavior, or does it create more temptation for opportunistic behavior? How do contracts affect these dynamics? Lucineau and Malhotra (2011) found that control-related contractual provisions tend to reduce goodwill trust, while coordination-related contractual provisions do not, and that this reduction in goodwill trust increases the likelihood of ending a relationship after a dispute. To what extent can we further understand the role of specific contractual terms in jointly affecting trust and performance? Do coordination-related contractual provisions support the development of the goal congruence that Jap and Anderson (2003) highlighted? Can TCT inform the use of “situational factors” that facilitate more accurate assessment of others’ trustworthiness? Finally, when conflict arises, what affects a party’s perception of its partner’s benevolence and competence? In a laboratory experiment, Harmon, Kim, and Mayer (2015) found evidence that subjects are more forgiving of perceived incompetence than of perceived deliberate shirking. It is worth noting that their experiment exposed subjects to identical protestations of “I didn’t realize” from their partners, while a specific contractual provision was varied so that the violation would be perceived to be either of the letter or the spirit of the contract. Subjects perceived violations of the letter to be more deliberate than violations of the spirit. Future work that further unpacks the link between formal stipulations and trust in the case of conflict in collaborations would enhance our understanding of trust and opportunism, and, like the study by Harmon, Kim, and Mayer (2015), may also speak to issues of framing (discussed in greater detail in the next subsection).

More generally, TCT has tended to focus on formal organization mechanisms (e.g., Argyres & Silverman, 2004). At the same time, literature exists on the role of informal organization (Dyer & Singh, 1998; Gulati, 1995). The last two decades have witnessed much work on whether these are complements or substitutes.
Extensions to New Phenomena

1. Technological advance: Platforms and governance. Recent years have witnessed the application of TCT insights to understand new phenomena that leverage advances in communications technology including e-business (e.g., Amit & Zott, 2001; Hennart, 2019), alliance constellations (e.g., Gomes-Casseres, 2003), innovation ecosystems (e.g., Adner & Kapoor, 2010), and technology platforms (e.g., Lehdonvirta, Kässi, Hjorth, Barnard, & Graham, 2019). This modest body of work has suggested that TCT is relevant and can indeed offer value in explaining some of the important aspects of these phenomena. In addition to using TCT to understand these new phenomena, conversely, one can also use these phenomena to extend and refine TCT. Notably, new business models such as platforms and ecosystems raise questions pertaining to governance in a multiparty context, which encourage an extension of the TCT logic beyond the usual dyadic level of analysis. Hence, there might be substantial theoretical rewards to exploring new contexts and phenomena.

Hennart (2019) argued that a good starting point is understanding that most forms of e-commerce have analogs in the pre-IT era. Hence, Consumer-to-consumer (C2C) platforms, such as eBay and eharmony, are brokers, putting in contact individuals in search of used household furniture or relationships, respectively, analogous to classified advertisements, flea markets, and marriage bureaus. Consumer-to-business (C2B) sites are electronic analogs of traditional brokers. Business-to-business (B2B) sites are electronic analogs of businesses. Business-to-consumer (B2C) sites, such as Booking.com and Expedia, provide the same services as travel agents. Uber, Lyft, and Ola offer taxi rides, just like traditional taxi companies, and customers can choose between ordering from online retailers such as Amazon or buying from brick-and-mortar retailers.

Similarly, platform-based ecosystems have analogs in the pre-Internet world, in situations where scale and switching costs matter. Platforms such as Amazon or Apple emerged from differences in optimal scale between the upstream and the downstream stages of a business. The optimal scale of designing, manufacturing, and retailing a mobile phone is larger than that of designing apps. At the same time, the attributes of apps are such that contracting between app developers and phone makers is feasible; hence,
Apple can contract with app developers rather than vertically integrate into app production. This is similar to the ethical drug business, where the optimal size of drug discovery is smaller than that of drug marketing, with the result that the industry is characterized by a larger number of drug discovery firms that contract with a smaller number of drug distributors (Big Pharma). The power that mobile phone companies have over app developers (and Big Pharma has over drug discovery firms) derives from the fact that, because of scale economies, there are fewer platforms than there are app developers. The same logic applies to large retailing platforms such as Amazon. Because of first-mover advantages and customer switching costs, these firms have obtained market power vis-à-vis the firms using their platforms. This is very similar to the case of Sears when it was the dominant mail order distributor.

So, what is different? At least three items. First, network effects often magnify the scale economies and first-mover advantages associated with a platform business (Arthur, 1994). This has implications for the strategies that young platform businesses pursue to gain scale (Boudreau, 2017). Of particular note for TCT, to the extent that these network effects create “winner-takes-all” markets that are ultimately dominated by a handful of entrenched incumbents, the difficulty of unseating an incumbent platform owner can also magnify hold-up risks later in the platform’s life. TCT is likely to be particularly useful for understanding how a new platform can try to attract partners, and how those partners should anticipate future problems when contracting with the young platform. Second, technology platforms typically have more information on both the profitability of the business generated by the firms using their platforms and the behavior of its partners. For example, Amazon appears to have used profit-related information to expropriate the profits of the partners using its platform by replacing their products with its own (Zhu & Liu, 2018). In addition, although there are no fundamental differences between the business model used by Uber and that used by franchisors such as McDonald’s, Uber has more power over its “franchisees” than does McDonald’s. Whereas McDonald’s must send people to intermittently inspect its franchisees, Uber can do this remotely and consistently through in-car cameras. Uber can also control driver behavior through the navigation system drivers must use and keep them from cheating passengers through its payment system. TCT should be particularly useful at analyzing the implications for the governance of platform-based exchange of lower-cost monitoring of both behavior and output. Third, technology platforms have an unusually high ability to bundle and to exclude (Boudreau, 2010; Hagiu & Wright, 2015). Thus, it is much easier for Uber to cut off franchised drivers than it is for McDonald’s to terminate franchised outlets; or for Amazon to add its own, or third-party, certification information than for Walmart (in their brick-and-mortar stores) to do so. From a TCT perspective, this is a double-edged sword. On the one hand, platform owners can use these advantages to elicit desired behaviors from partners (Rietveld, Seamans, & Meggiorin, 2020), including enhanced coordination with other partners. At the limit, the platform owner may even provide stronger institutions than the national government (Liu & Weingast, 2017). On the other hand, potential partners should anticipate future dependence on the platform owner, and thus demand strong contractual protections before joining. There are many opportunities to apply and extend TCT analysis to these situations; our sense is that this extension will explain much of e-business dynamics without the need to develop an entirely new theory of platform governance.

2. Technological advance: Artificial intelligence, machine learning, and transaction costs. Here, we illustrate how TCT can be applied to new phenomena by focusing on recent technological trends, such as advancements in machine learning, artificial intelligence, big data, and blockchain technology. We believe that each of these trends has the potential to affect the level of behavioral uncertainty firms face, and the relative costs of different governance mechanisms. For example, law scholars, legal practitioners, and management scholars who focus on contracting issues are starting to recognize that machine learning and artificial intelligence, combined with the availability of big data on contracts, might pave the way for the partial, or even full, automation of contract drafting, as well as of contract review and analysis (e.g., Betts & Jaep, 2017; Betts & Liu, 2017).

20 In both cases, the minimum efficient scale at the upstream stage (format and product design and advertising for McDonald’s, platform for Uber) is greater than that at the downstream stage (restaurants for McDonald’s, taxi service for Uber). Subcontracting to independent agents (franchisees for McDonald’s, drivers for Uber) makes sense in this case, because the gains in effort and flexibility achieved by using independent contractors are greater than the potential loss in quality shading that will inevitably result.

21 Note, however, that the introduction of new technology has not totally eliminated quality shading by drivers, which, in the case of Uber, manifests itself in dangerous driving and assaults on customers.
Mills, 2016; Rich, Weber, & Bauman, 2020). Early applications of these ideas suggest that they could lead to substantial efficiency gains and cost reduction in the enforcement of contracting, thereby making market-based transactions less costly. At the same time, machine learning and artificial intelligence might lead to significant enhancements and expansion in the automation of tasks and decision-making within firms (e.g., Jarrahi, 2018; Schneider & Leyer, 2019). Human resources scholars and practitioners have acknowledged that these technologies have important implications for monitoring and evaluating employees (e.g., Kellogg, Valentine, & Christin, 2019; Tambe, Cappelli, & Yakubovich, 2019). Thus, advances in machine learning and artificial intelligence might reduce shirking and the need for and cost of monitoring employees, as well as the potential for opportunistic behavior more generally. This implies that the same technological trends that have the potential to make market transactions more efficient also have the potential to have a similar effect on hierarchies. Therefore, it is important not only that future research investigates the cost implications of these technological trends for one particular governance mechanism or mode, but also that such research considers more broadly how such technologies affect the relative cost of different governance mechanisms (since it is these relative, comparative, assessments that allow us to form predictions that are properly grounded in the TCT logic).

Another technological advance that might be relevant for TCT is blockchain technology. While successful blockchain applications might be scarce at this moment, scholars have already started to explore their potential (e.g., Davidson, De Filippi, & Potts, 2018; Schmidt & Wagner, 2019). Blockchains allow for immutable decentralized public ledgers and thus enable the keeping and sharing of records of past behavior in a distributed and decentralized way. Catalini and Gans (2016) argued that this technology has the potential to lower transaction costs and improve the efficiency of markets through costless verification and by reducing the need for costly intermediation. Furthermore, as blockchain technology might make past opportunistic behavior available as searchable public information, it also has the potential to reduce behavioral uncertainty and to deter or reduce future opportunistic behavior.

### 3. Nonpecuniary phenomena: Nationalism; sustainability and corporate social responsibility; government action via public–private partnerships; and “grand challenges.”

The rise of nationalism. In recent times, several areas in the world have experienced a surge in nationalism (Luce, 2019). Research in management has remained mostly silent on this important phenomenon (Ertug, Cuypers, & Dow 2018). However, insights from decades of work in political science and social psychology (e.g., Davidson, 2009; Druckman, 1994; Mead, 1929) have shown that nationalism is linked to behaviors that are likely to be relevant for research in management, organizations, and strategy. For example, and as particularly relevant for TCT and governance choices, higher levels of nationalism have been linked to a lower tendency to view foreigners as being trustworthy (e.g., Yzerbyt & Demoulin, 2010), and to favoritism toward domestic actors (e.g., Shi & Wright, 2003). Among other matters, this might have implications for global supply chains and buyer-supplier relationships between firms from different countries. A buyer from a more nationalistic country, as compared to one from a less nationalistic country, might have a tendency to view foreign suppliers as less trustworthy. As a result, a buyer from a more nationalistic country might perceive a greater threat of opportunistic behavior from its partners, and therefore be more likely to try to mitigate this by internalizing the transaction (i.e., make rather than buy), rather than to exchange with a foreign supplier. Nationalism might also matter on the supplier’s side in this relationship. As we mentioned, nationalism has been linked to a tendency to favor domestic actors over foreign ones. Hence, a buyer might be more concerned that a foreign supplier from a more nationalistic country will be more likely to behave opportunistically, in the sense of favoring domestic firms or buyers at the expense of the focal buyer, than when dealing with a foreign supplier from a less nationalistic country. This suggests that a buyer might be more likely to internalize a transaction when a supplier is from a more nationalistic country. These two examples point to opportunities to investigate the impact of nationalism on the strategic and governance decisions of firms from a TCT perspective.

**Corporate social responsibility.** Corporate social responsibility (CSR) has received dramatically increased attention in recent years. Key issues in this stream of work include the extent to which ostensible commitment to CSR is reflected in actual change, whether engagement in CSR affects economic performance or competitive advantage, the effective monitoring of firms’ CSR promises, and the relative efficacy of different actors in undertaking socially beneficial efforts. TCT has direct implications for all of these questions.

In a series of papers, Flammer and colleagues provided evidence that strategic investment in CSR can yield bottom-line benefits to firms. The thread
through many of these studies is that a firm’s CSR investment can credibly signal its intention to behave responsibly in the future, in part because its investment in a reputation for responsibility will be eroded otherwise (Flammer, 2020). Such investments can also create favorable impressions on employees, customers, or bestowers of government contracts, and these give the firm an advantage in the face of competitive pressures (Flammer, 2015; Flammer & Luo, 2017). Relatedly, Henisz, Dorobantu, and Nartey (2014) found that judicious stakeholder engagement enhances the market value of gold mining firms, even if their underlying assets remain unchanged. In a cross-national study, Ioannou and Serafeim (2012) found that national institutions affect firms’ likelihood of engaging in CSR, presumably because of differences in the expected returns on these actions.

Other scholars have explored the degree to which firms may decouple their actual operations from their professed claims to pursue CSR, and the ways in which such claims can be monitored (e.g., Cuypers, Koh, & Wang, 2016; Luo, Wang, & Zhang, 2017). Of particular note here is the fact that many CSR initiatives require concomitant efforts by other actors throughout a supply chain, which makes CSR initiatives require concomitant efforts by other employees, customers, or bestowers of government contracts, and these give the firm an advantage in the face of competitive pressures (Flammer, 2015; Flammer & Luo, 2017). Relatedly, Henisz, Dorobantu, and Nartey (2014) found that judicious stakeholder engagement enhances the market value of gold mining firms, even if their underlying assets remain unchanged. In a cross-national study, Ioannou and Serafeim (2012) found that national institutions affect firms’ likelihood of engaging in CSR, presumably because of differences in the expected returns on these actions.

These studies point to yet more fascinating questions illustrating how TCT can be fruitfully applied to CSR research. Given the feasibility of public, private, or nongovernmental organization monitoring of CSR efforts, which form should be used for which type of CSR activity? Analogous to research on industry-wide versus firm-specific lobbying (e.g., de Figueiredo & Tiller, 2001; Jia, 2018), when are industry-wide certification efforts preferable to firm-specific efforts? And what are the tradeoffs in encouraging CSR behavior? Flammer, Hong, and Minor (2019) found that, not surprisingly, making CSR a formal part of the CEO’s compensation criteria yields greater corporate attention to CSR. But is this costless? As the CEO devotes more effort to CSR, what other aspects of the firm’s strategy get less attention and resources, and how does this affect firm performance? As agency theory tells us, the wider the range of performance targets managers must meet, the more difficult it is to monitor their performance (Jensen & Meckling, 1976; Williamson, 1963). Therefore, what are the implications of moving to a more expansive set of performance criteria, which now encompass CSR?

Finally, two papers distinguished between the private and social benefits of CSR through a comparative governance lens. Kaul and Luo (2018) explored for-profit and nonprofit provision of social goods, demonstrating conditions under which a for-profit firm’s CSR effort will generate private returns but no social benefit. They developed a “comparative efficiency” typology that matches organization form to CSR activity; they then expanded this (Luo & Kaul, 2019) to consider the comparative efficiency of a wider set of market frictions and organization types. Drawing directly on ideas of discrete comparative analysis (Williamson, 1991), these studies have introduced TCT to a new research field and laid out a roadmap to address key questions about the governance of CSR efforts.

Public-private partnerships. There has also been an upsurge of interest in PPPs as vehicles through which government can address important public-good problems (Mahoney, McGahan, & Pitelis, 2009). We find TCT to be well-suited for contributing to further addressing important questions in this area as well (Chong, Saussier, & Silverman, 2015; Kivleniece & Quelin, 2012; Rangan, Samii, & Van Wassenhove, 2006), just as it has been fundamental to understanding collaboration between private firms (Hennart, 1988b; Oxley, 1999; Sampson, 2004). There are two particularly interesting aspects to PPPs from a TCT perspective. First, public contracting is complicated because out-of-office political parties have an incentive to accuse the in-office party of exercising poor judgment or excessive discretion in these contracts. Consequently, public–private contracts tend to be more rigid than private contracts (Beuve, Moszoro, & Saussier, 2019; Moszoro, Spiller, & Stolorz, 2016). This rigidity constrains the value that can be created by PPPs. Second, whereas TCT typically assumes that within-organization governance is more flexible than interorganizational governance (because of
the possibility of exercising managerial fiat), in the public sector this is frequently not the case—civil service rules often constrain in-house activity more than public contracting will. Thus, the costs and benefits of PPPs are qualitatively different from those of private-sector collaborations. Not only can TCT continue to contribute to our understanding of PPPs, but a systematic study of public–private contracting can also serve to refine the boundary conditions of TCT.

Yet another set of phenomena that has gained importance in recent years is related to what are referred to as “grand challenges” (e.g., George, Howard-Grenville, Joshi, & Tihanyi, 2016). Addressing these grand challenges often requires collaboration between diverse types of organizations (e.g., for profit firms, not-for-profit firms, governments), which might require different ways of organizing compared to typical collaborations. Hence, we see potential for TCT to inform the managers and decision makers, both of firms and of other types of organizations, about how these more diverse and complex collaborations should be governed.

In Table 4, we provide an overview of our discussion in this section.

**CONCLUSION**

Transaction cost theory has enjoyed substantial influence across a wide swath of management research and cognate disciplines. The theory has proven resilient due to its ability to evolve effectively—to extend to new phenomena and to incorporate a broader set of theoretically relevant factors, while retaining a consistent perspective on the primary motivations for economic organization. Yet, as a theory progresses, especially across a range of fields and phenomena, it is important to intermittently devote attention to consolidating and integrating its advances and challenges. This review is our attempt to provide such a consolidation, and to suggest a roadmap for future research—notably, encouraging greater engagement with recent advances in relevant cognate disciplines and highlighting how TCT can be applied to some of the most important novel phenomena of our time. In sum, in addition to its proud history, we are confident that TCT scholarship has a promising future awaiting it.

**REFERENCES**


Contractor, F., Kundu, S., & Hsu, C. 2003. A three-stage theory of international expansion: The link between


Klein, S. 1989. A transaction cost explanation of vertical


Tambe, P., Cappell, P., & Yakubovich, V. 2019. Artificial intelligence in human resources management:


Zhao, H., Luo, Y., & Suh, T. 2004. Transaction cost determinants and ownership-based entry mode choice: A


Ilya Cuypers (ilyacuypers@smu.edu.sg) is an associate professor of strategy at Lee Kong Chian School of Business, Singapore Management University. He received his PhD in strategy and international business from Tilburg University. His research focuses on governance, dynamics, and performance implications of cross-border external corporate development activities, investment decisions under uncertainty, and issues regarding interorganizational ties.

Jean-François Hennart (j.f.hennart@uvt.nl) is professor of international management emeritus at Tilburg University and visiting professor at Politecnico di Milano and Aalborg University. His research focuses on the comparative study of international economic institutions such as multinational firms and their contractual alternatives, family firms, Born Globals, joint ventures and alliances, and modes of foreign market entry and exit.

Brian S. Silverman (silverman@rotman.utoronto.ca) is the J. R. S. Prichard Chair in management, and professor of strategic management, at the University of Toronto’s Rotman School of Management. His research focuses on the ways in which firms’ strategies and organizational structures interact to affect their performance.

Gokhan Ertug (gokhanertug@smu.edu.sg) is a professor of strategic management and Lee Kong Chain Fellow at the Lee Kong Chian School of Business, Singapore Management University. He received his PhD from INSEAD.