Received: 7 June 2018 | Revised: 25 March 2020 | Accepted: 13 May 2020
DOI: 10.1002/joom.1103

RESEARCH ARTICLE

WILEY

Bilateral dependency and supplier performance ambiguity in supply chain contracting: Evidence from the railroad industry

Akhil Bhardwaj[1] | **Mikko Ketokivi**[2]

[1]Department of Management, Tilburg University, Tilburg, The Netherlands
[2]IE Business School, IE University, Madrid, Spain

Correspondence
Akhil Bhardwaj, University of Tilburg, Warandelaan 2, Koopmans Building K 933, 5037 AB Tilburg, The Netherlands.
Email: A.Bhardwaj@tilburguniversity.edu

Funding information
Ministerio de Economía y Empresa, España; European Regional Development Fund (ERDF), Grant/Award Number: ECO2016-80106-R

Handling Editor: Sriram Narayanan

Abstract

Research on supply chain relationships tends to focus on power and asymmetric dependency. Our objective is to complement this by examining contractual challenges in the context of bilateral dependency. Our specific empirical focus is on how Indian Railways, one of the largest railroads in the world, manages warranty claims related to engine failures. Warranty resolution is complex, because failures typically involve the simultaneous failure of several engine subsystems provided by different suppliers, giving rise to supplier performance ambiguity. Through a combination of theoretical reflection rooted in transaction cost economics and an empirical analysis of Indian Railways' supply chain, we argue that in the case of bilateral dependency, "the benign approach" is more efficient than "the muscular approach," to use Oliver Williamson's terminology. Specifically, while the latter is founded on unilateral decisions by the comparatively more powerful party, the former is based on voluntary long-term cooperation, and calls for mutual credible commitments and joint problem-solving. Additionally, we highlight that even if dependency is asymmetric at the outset, it can develop bilateral features over time. Theoretically, our research offers an enhanced understanding of trust in buyer–supplier relationships, emphasizing its organizational as opposed to personal basis.

KEYWORDS

bounded hindsight, contracting, outsourcing, supply chain management, system failure, transaction cost economics

1 | INTRODUCTION

Indian Railways, one of the world's biggest railroads, owns and operates some of the largest locomotive assembly plants in the world. Smooth operation of these plants hinges on stable long-term relationships with its key suppliers. To mitigate adverse selection problems and to ensure reliable supply of high-quality components and subsystems, Indian Railways maintains a rigorous supplier qualification process. Further, while Indian Railways relies on a competent first-tier supplier base, it seeks to avoid dependency on any *specific* supplier. To this end, it avoids single-source supply, and has incorporated various safeguards into its supply contracts. This approach is not only managerially sensible, but is also well established both in operations management (e.g., Grover & Malhotra, 2003) and in organization economics (e.g., Williamson, 1985).

Safeguards notwithstanding, conflicts with some of the major original equipment manufacturer (OEM) suppliers occur. Some of these conflicts have resulted in Indian Railways penalizing suppliers by reducing order placement and by deducting payments in cases of warranty disputes. At the same time, Indian Railways is fundamentally dependent on these suppliers, and seeks effective remedies to secure the continuation of the relationships. This is a challenging task: A former member of Indian Railways' board of directors noted that trying to manage the supply chain "eventually leads to poor performance" and "never exposes inefficiency" (personal communication, August 24, 2017). Another former board member noted that the implications of contractual clauses between Indian Railways and its suppliers aimed at addressing quality problems and warranty resolutions have not received adequate attention (personal communication, December 5, 2016). Finally, a former General Manager of Indian Railways' largest production facilities expressed a concern regarding top management's tendency to frame the problem in terms of moral hazard (personal communication, September 17, 2019).

The General Manager's comment in particular merits attention. Is the problem moral hazard and opportunism, or something else? Although Indian Railways is a large buyer, buyer–supplier dependency is bilateral for many crucial subsystems. Even in the case of having multiple suppliers for a given subsystem, the number of suppliers is never high enough to render any individual supplier inconsequential. In this exploratory study of Indian Railways' supply chain, we examine buyer–supplier relationships under conditions of such bilateral dependency (Williamson, 1985: 91), paying special attention to how these relationships evolve over time.

In our context, *supplier performance ambiguity* constitutes a critical complicating factor to bilateral dependency. As our analysis shows, Indian Railways consistently faces situations in which locomotive engines inevitably fail. Not only do most failures occur before the first maintenance event (within 1 year of commissioning the engine), it is often impossible to determine which subsystem triggered the failure. In this study, supplier performance ambiguity is specifically about the inability to determine the root cause of failure, which is analogous to Alchian and Demsetz's (1972: 779) *metering problem* that arises in team production where determining a single team member's contribution (e.g., marginal productivity) is impossible. Specifically, when failures occur, it is difficult to determine whether a subsystem failed due to a defect, improper assembly, or perhaps as a consequence of another subsystem failure. The diversity of potential causes of failure hinders traceability (Skilton & Robinson, 2009). Indeed, one of our key informants Dan Belisle (a Senior Field Engineer with 35 years of managerial and engineering experience) noted: "Failure analysis doesn't always lead us to the correct root cause, or even any cause. In fact, the most common failure identified and reported by the majority of railroads is NDF—short for No Defect Found." Similar practical challenges with the effectiveness of root cause analysis are well documented across several industries, including automobiles, aviation, and health care (Dekker, Cilliers, & Hofmeyr, 2011; MacDuffie, 1997; Wu, Lipshutz, & Pronovost, 2008).

When a supplier unilaterally depends on a comparatively larger buyer, the contracting implications of supplier performance ambiguity are a nonissue: The comparatively more powerful large buyer sets the terms, and the supplier either accepts them or walks away. The buyer can, for example, insert a general clause in the contract whereby failed subsystems and components will be considered de facto defective, effectively eliminating any contractual ambiguity. It is difficult for the smaller supplier to make any claims precisely due to the ambiguity of its own performance: How would the supplier demonstrate—to the buyer, an arbitrator, or the courts—that it has been treated unfairly? Further, the buyer would likely perceive arbitration and litigation as adversarial, which jeopardizes the contractual relationship (Williamson, 1975: 30). In sum, the comparatively weaker supplier is directly incentivized not to involve third parties in conflict resolution.

Under conditions of bilateral dependency, unilateral decisions are myopic: The key question is whether supplier performance ambiguity can be resolved in a manner that both parties find reasonable. The ability to address this question both *ex ante* (contract negotiation phase) and *ex post* (contract enforcement phase) constitutes an effective safeguard that protects the exchange relationship. Conversely, the inability to address it jeopardizes the exchange relationship and materially threatens the value-creation processes of the bilaterally dependent parties. These are the central concerns in this study.

This article is structured as follows. We begin with a review of the literature on supply chain dependency, and how our research question relates to the published literature. As part of this positioning, we propose that transaction cost economics (TCE) can offer a useful theoretical foundation and key terminology (Anand & Gray, 2017; Ketokivi & Mahoney, 2020). We conduct an empirical analysis of contractual dynamics in the context of engine failures at Indian Railways. Engine failures are an appealing option specifically because they are subject to both supplier performance ambiguity and bilateral dependency, which in combination cause contractual complications. Using a data set of 172 engine failures, we first examine supplier performance ambiguity. We then investigate how Indian Railways addresses the contractual problem associated with bilateral dependency and supplier performance ambiguity. Based on our analysis, we suggest courses of action to mitigate inefficiencies due to supplier performance ambiguity. Finally, we discuss the role of relational contracting and the organizational basis of trust for improving supply chain contracting, and discuss future research directions.

2 | BILATERAL DEPENDENCY IN SUPPLY CHAIN RELATIONSHIPS

Relationships that involve negligible dependency between the buyer and the seller are not vulnerable to the *hold-up problem*, a situation where one exchange partner holds the other "hostage" because the latter has made a unilateral relation-specific investment of some kind (Goldberg, 1976; Klein, Crawford, & Alchian, 1978). Absent dependency, the contracting relationship is likely an arm's-length market transaction where the number of actors is sufficiently large to render the identity of any individual buyer or seller insignificant (Williamson, 1985: 69).

Contracting in supply chains becomes more complex when identities matter and dependencies develop over time. In this vein, research on supply chain dependencies tends to focus on dependency that is *asymmetric*: Buyer–supplier relationships are approached from a perspective where "one of the parties, usually the large buyer, deals with smaller suppliers in a peremptory way, [and] often 'use up' their suppliers and discard them" (Williamson, 2008: 10). This "muscular approach" (Williamson, 2008: 10) is often adopted in empirical supply chain management research (Benton & Maloni, 2005; Handfield, 1993; Tsay, Gray, Noh, & Mahoney, 2018; Wang, Niu, & Guo, 2013). The organization-theoretic basis for such approaches lies in power-based theories of organizational boundaries (Santos & Eisenhardt, 2005). The empirical supply chain literature on the topic is as massive as it is diverse (e.g., Agrawal, Muthulingam, & Rajapakshe, 2017; Brinkhoff, Özer, & Sargut, 2015; Crook & Combs, 2007; Gray & Handley, 2015; Handley & Benton, 2012; Maloni & Benton, 2000; Nyaga, Whipple, & Lynch, 2010; Plambeck & Taylor, 2005; Villena & Craighead, 2017).

But what happens when dependency is bilateral and "the muscular approach" loses both descriptive and prescriptive power? How do contracting dynamics change when no single party can dictate general contractual clauses to serve its own purposes? What role does trust play in establishing a stable buyer–supplier relationship? What, exactly, is the basis of buyer–supplier trust (MacDuffie & Helper, 2006)?

As a counterpoint to "the muscular approach," Williamson (2008: 10) wrote about "the benign approach," which emphasizes cooperation and mutual gains in an attempt to contract efficiently. This alternative approach has economic theories of efficiency—most notably TCE—at its foundation. To clarify, even though TCE is often interpreted in the supply chain management literature as a theory of supply chain power (Tsay et al., 2018), it is in fact a theory of economic efficiency (Williamson, 1991)—it is crucial not to confound the two (Ketokivi & Mahoney, 2020). The premise in TCE is that efficient contracting works toward safeguarding the exchange relationship precisely because it focuses on mutual benefits. Importantly, TCE is applicable in situations where the dependency between the buyer and the supplier is indeed bilateral, which makes it a particularly suitable theoretical basis for our inquiry.

"The benign approach" emphasizes voluntary cooperation and trust. In our study, we seek a more nuanced understanding of this trust. The concern is that adopting "user-friendly" terms like trust may confuse rather than clarify (Williamson, 1994: 97). The specific challenge related to trust in the context of buyer–supplier relationships is that two individuals versus two legal entities (firms) trusting one another are not the same thing. Even though interpersonal trust may play a role in supply chain relationships (Cai, Jun, & Yang, 2010; Handfield & Bechtel, 2002), managers are not so much individuals as they are "repositories of organizational schemata" and behave in ways that reflect their organizations (Barden & Mitchell, 2007). Building on this logic, we submit that stable long-term buyer–supplier relationships have an interorganizational, not an interpersonal, basis.

2.1 | Buyer–supplier dependency: four contracting scenarios

It is crucial to distinguish between unilateral and bilateral dependency. Dependency is best thought of in terms of switching costs: Unilateral dependency means the switching cost is consequential for one of the exchange partners and inconsequential for the other, bilateral dependency means consequential switching costs for both parties. Operationally, we can think of switching costs as *consequential* when they have direct contractual implications. Here, we consider the unilateral–bilateral distinction in conjunction with supplier performance ambiguity. Incorporating these two dimensions, we arrive at four different contracting scenarios (Figure 1). Both theoretical and empirical work on unilateral dependency is well established, but the case of bilateral dependency merits more attention. When the buyer–supplier dependency is bilateral, market transactions give way to *relational contracting*: Contracting parties face a situation in which their ability to create value depends significantly on the continuity of the specific relationship

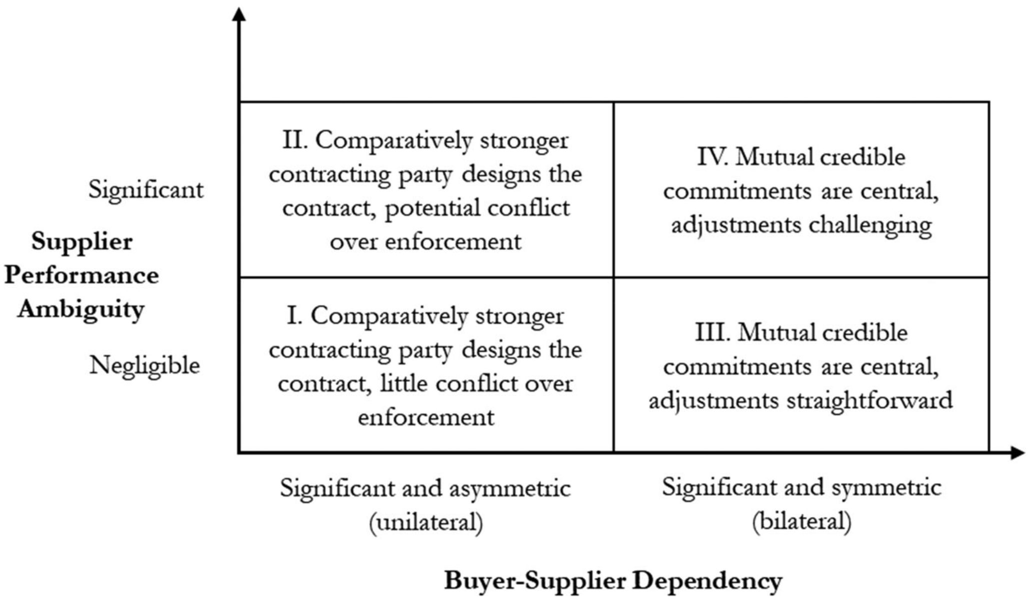

FIGURE 1 The four contracting scenarios

(Williamson, 1985: 71). "The muscular approach" is no longer feasible (Williamson, 2008).

An explicit theoretical examination of dependency is crucial, because this is where supply chain management researchers often misconstrue TCE's logic (Ketokivi & Mahoney, 2020). While TCE invites us to describe the world in realistic terms (however unflattering), it prescribes *mutual credible commitments* (Williamson, 1983) to safeguard a relationship on which both contracting parties depend and from which both benefit. TCE therefore pertains to Scenarios III and IV in Figure 1 (e.g., Jap & Anderson, 2003; Liu, Luo, & Liu, 2009; Poppo, Zhou, & Zenger, 2008). Power approaches to supply chain management (relevant in Scenarios I and II) serve an important purpose, but in this study, we highlight the equally important and somewhat overlooked aspect of understanding bilateral dependency that arises from *asset specificity* (Williamson, 1985: 52).

In order to understand Figure 1 in more depth, we review in the following section the theoretical arguments that focus on the drivers of bilateral dependency and supplier performance ambiguity. We approach supplier performance ambiguity specifically from the point of view of subsystem failure, because ambiguity seems most relevant in contexts when something unplanned and unexpected happens.

2.2 | The drivers of bilateral dependency: asset specificity and the fundamental transformation

Bilateral dependency and the need for relational contracting arise from relation-specific investments. In TCE, this is known as *asset specificity*, described by Williamson (1985: 55) as "durable investments that are undertaken in support of particular transactions." Asset specificity can take many forms: physical assets, human assets, location (site specificity), and so forth. Importantly, even if the relation-specific investment is made by only one of the parties, say, the supplier, the resulting dependency often develops bilateral features: "[T]he buyer cannot turn to alternative source of supply and obtain the item on favorable terms since the cost of supply from unspecialized capital is presumably great" (Williamson, 1979: 240). Under bilateral dependency (Scenarios III and IV), adopting "the muscular approach" is "myopic and inefficient" (Williamson, 2008: 10).

An important aspect of contractual dynamics in TCE is the *fundamental transformation*: "[A] large-numbers condition at the outset (*ex ante* competition) is transformed into a small-numbers condition during contract execution and at contract renewal intervals (*ex post* competition)" (Williamson, 1985: 12). The fundamental transformation occurs, because asset specificity develops over time through mutual learning and adjustment. Therefore, even if the exchange relationship starts as an arm's-length market transaction, it may develop relational features over time. Importantly, the fundamental transformation highlights the fact that specificity is not just about *ex ante* investment, it may also occur gradually *ex post*.

Despite having "pervasive contracting consequences" (Williamson, 1985: 61), the fundamental transformation remains empirically unexamined (Walker, 2007). The implications for contract design must be rigorously derived: What is considered *ex ante* versus *ex post*? When and how is it possible to become more efficient by "looking before one leaps," that is, addressing at least some *ex post* problems *ex ante*? To what extent do concerns of *ex post* hazards direct *ex ante* actions? Importantly, incorporating the time dimension into the analysis can lead to both theoretically and managerially relevant insights (Mahoney & Qian, 2013).

2.3 | The driver of supplier performance ambiguity: bounded rationality

Why and how do systems fail? Do failures have a root cause in that system failure occurs when one or more component or subsystem fails, and as a consequence, the failure propagates throughout the system? Do systems fail due to bad design, defective subsystems and components, faulty assembly, or exogenous shocks? Can the root cause be definitively identified?

Evidence suggests that even in high-hazard industries, failures often elude analysis (Carroll, Hatakenaka, & Rudolph, 2006; Tucker & Edmondson, 2003); despite our best efforts, analysis may at best lead to a list of "contributing factors." For example, Transportation Safety Board of Canada (2013) conducted a detailed investigation of the fatal Lac-Mégantic derailment that killed 47 people, and uncovered 18 "contributing causes" but no definitive root cause. This inability to identify a root cause is related to the fact that we tend to engage in *linear thinking*, specifically, "a chain of causal reasoning from a premise to a single outcome" (Dekker et al., 2011: 939). In reality, failures are often emergent outcomes that involve complex interactions and feedback loops (Perrow, 1984).

Another cognitive shortcut used is *reductionism*: The goal of failure analysis is to identify "the trigger, the original culprit…[that everybody agrees] could carry the explanatory load of [the failure]" (Dekker et al., 2011: 940). However, in many failures, this "eureka part" (Dekker et al., 2011: 940) is never found. In TCE terminology, reductionism is a manifestation of bounded rationality, where economic actors are "*intendedly* rational, but only *limitedly* so" (Simon, 1997: 88). To clarify, reductionism should not be thought of as a form of self-deception, but rather, an honest attempt to solve a problem that may be too complex to address without some form of simplification.

Some causes of failures can be remedied by redesign (e.g., Baldwin & Clark, 2000; Cui, Loch, Grossmann, & He, 2012). However, tracing a design flaw is challenging in fragmented supply chains (Hora, Bapuji, & Roth, 2011: 769; Skilton & Robinson, 2009), particularly since admitting product design flaws can have severe financial consequences for the designer or the OEM (e.g., Tang, 2008). Further, redesign may be ineffective, because it hinges on the assumption that faulty design can be confirmed as the cause of the failure (Agrawal et al., 2017; Peerally, Carr, Waring, & Dixon-Woods, 2017). In systems that interact with the environment, failures may be caused by exogenous shocks, such as a foreign substance entering the system, making the root cause difficult to identify (Starbuck & Farjoun, 2005). Finally, from the organizational point of view, root cause analyses are often not aimed at solving the technical problem, but rather, "assigning financial responsibility" (MacDuffie, 1997: 486).

2.4 | What is missing?

The implications of buyer–supplier interdependence for both supply chain relationships (e.g., Hoetker, Swaminathan, & Mitchell, 2007; Mahapatra, Narasimhan, & Barbieri, 2010) and firm boundaries (Novak & Eppinger, 2001; Sosa, Eppinger, & Rowles, 2004; Ulrich & Ellison, 2005) are well established. The role of supply chain contracting in the context of quality failures (e.g., Zu & Kaynak, 2012) and supplier performance ambiguity (Agrawal et al., 2017; Mellewigt, Hoetker, & Lütkewitte, 2018) has also been scrutinized in detail. There are many authors who write about the importance of avoiding lock-in by either using multiple suppliers for the same component, or maintaining component production capabilities in-house as well (e.g., Gray & Handley, 2015; Mahapatra et al., 2010). However, little attention has been paid to relationships where bilateral dependency is unavoidable and the relationship is subject to supplier performance ambiguity. In this study, we explore the contractual implications in such cases.

The general prescription in the face of uncertainty, particularly in situations that involve asset specificity, is vertical integration (Williamson, 1985). At the same time, sometimes vertical integration is simply infeasible (Williamson, 1985: 96). Here, contractual mechanisms such as *performance-based contracting* have been found to be improve product reliability (Guajardo, Cohen, Kim, & Netessine, 2012). But performance-based contracting faces obvious problems, because the premise that "a supplier is paid based on the realized outcome of customer value" (Guajardo et al., 2012: 961) hinges precisely on the assumption of no performance ambiguity. Performance-based contracts are particularly difficult to apply in fragmented supply chains where specialized individual suppliers deliver components, and at the same time, performance and customer value are *system-level* outcomes.

We also need a more thorough empirical examination of the essence of supplier performance ambiguity. Most empirical studies, as well as theories such as TCE and agency theory, focus on *behavioral uncertainty*, that is, the difficulty of one contracting party being able to anticipate how the other transacting party will behave in unforeseen circumstances (Crook & Combs, 2007). Further, supplier performance is often addressed not directly but by soliciting managerial perceptions (Gray & Handley, 2015; Mellewigt et al., 2018). Supplier performance ambiguity may of course have a behavioral basis, but the basis can also be technical: It is not the behavior of the transacting parties but that of the technical system that is unpredictable.

When the problem is technical as opposed to behavioral, monitoring suppliers (as agency theory would prescribe) or vertical integration (as TCE would prescribe) are both less effective. Specifically, if the essence of the problem is our inability to understand how a technical system fails (i.e., bounded rationality), conventional governance mechanisms are not the remedy: “Any ‘shortfalls’ due to misperception or mistake will *not be remediable* by supplanting a [buyer–supplier contract] with vertical integration” (Williamson, 1985: 66). Similarly, efforts to merely align incentives are a blunt instrument due to performance ambiguity. There are a number of cross-sectional studies that examine the issue (e.g., Bai, Sheng, & Li, 2016; Nyaga et al., 2010), but much of the relevant action is in the dynamics. Particularly buyer–supplier relationships that undergo the fundamental transformation have not received research attention (Walker, 2007). Additionally, how contract enforcement evolves over time remains under-researched (Terpend, Tyler, Krause, & Handfield, 2008).

In the empirical part that follows, we seek to fill some of the gaps in our current understanding by examining contracting, bilateral dependency, and supplier performance ambiguity in the context of locomotive engine failures at Indian Railways. All the complicating characteristics mentioned apply to Indian Railways, an example of Scenario IV in Figure 1. In addition to understanding failures, our central focus is on the contractual dynamics.

3 | THE CASE OF INDIAN RAILWAYS

In our empirical analysis, we focus on supply chain relationships between Indian Railways and its locomotive engine subsystem suppliers. There are several theoretical reasons for this choice. First, many of the central buyer–supplier relationships exhibit bilateral dependency: Indian Railways is indeed dependent on particular suppliers, which are in turn reciprocally dependent on Indian Railways. Second, the relationships are characterized by significant technological asset specificity on the supplier side, because production of specific engine components requires specific special-purpose equipment. Even though these investments are unilateral (they appear on the supplier's balance sheet), they give rise to bilateral dependency. Third, locomotive engines fail with a frequency that requires systematic organizational attention in contracting, not simply *ad hoc* exception management: Direct warranty costs alone are significant enough to merit attention. Fourth, an engine failure tends to be associated with the failure of several subsystems, making it difficult to identify the root cause. Because the subsystems come from external suppliers and any given supplier typically supplies only one subsystem, the relationship is characterized by significant supplier performance ambiguity. Finally, many of the buyer–supplier relationships are long-term, which enables an examination of the fundamental transformation.

In sum, Indian Railways provides an opportunity to examine empirically Scenario IV (Figure 1) in particular. As a practical matter, Indian Railways is appealing from a data availability perspective. Specifically, the first author of this article is an Indian national who—by the Right to Information Act of 2005—has access to the detailed records relating to state-owned companies such as Indian Railways. This provided us with access to data on engine failures. The first author has also worked for several years in diesel locomotive engine maintenance, troubleshooting, warranty fulfillment, and contracting. Making sense of failure data requires technical expertise.

3.1 | The empirical context: the locomotive engine and its supply chain

Indian Railways is one of the largest railroads in the world. Its fleet of 6,000 diesel locomotives (and roughly a similar number of electric locomotives) transports daily 23 million passengers and 3 million tons of freight in its 12,500 trains traveling across 70,000 miles of track. Majority of the diesel locomotives were designed by an external supplier (American Locomotive Company [ALCO]), but all are assembled, owned, and operated by Indian Railways. ALCO engines are analogous to the Boeing 737 aircraft in the sense that they represent a dominant design that has not been subjected to significant redesign in several decades. Indeed, the ALCO is considered “the workhorse” of Indian Railways (Dutta &

Agarwal, 2017: 11). Dutta and Agarwal (2017: 11) further noted that the ALCO "[was] adapted to operate in our [-Indian] dusty conditions and was very versatile in operating freights and cross-country mail/express trains.... [it] had [also] operated with amazing reliability on the Andes mountain route to Cuzco from Lima, which was in the 1960s, the highest altitude railroad in the world." A former General Manager of one of the production plants echoed this sentiment regarding the reliability of the ALCO engine. For this reason, we focus in our empirical analysis on the design and redesign of *contracts*, not the *engine*.

The 15-metric-ton, 12- or 16-cylinder turbocharged diesel engine is the heart of the modern 132-metric-ton locomotive. The engine is attached to a 6-metric-ton generator, which in turn drives individual motors connected to six axles. The engine can usefully be partitioned into five subsystems: top deck, cylinder head, piston, cylinder liner, and engine block (the locomotive experts with whom we talked during data collection validated this partitioning). In the remainder of this article, these five engine parts are referred to as *subsystems*. Indian Railways uses different suppliers for different subsystems, and in addition, there are multiple suppliers for each subsystem. At the same time, no individual subsystem has so many suppliers that an individual supplier would not matter. In fact, there may be only one supplier for a particular *type* of a given subsystem, for example, a particular type of cylinder liner (Indian Railways internal minutes of meeting, distributed on March 15, 2019). The general architecture of the engine is illustrated in Figure 2.

Some of the subsystems are physically *coupled* in that there is a direct physical interface between the two. The interface of the piston and the cylinder liner is a good example: In an internal combustion engine, the piston moves back and forth inside the cylinder covered by a thin metallic sleeve, the liner. In addition to six coupled interfaces, there are also four *decoupled* ones (Figure 2). For example, the piston and the engine block are physically decoupled from one another in that they share no direct interface. However, even the decoupled subsystems are ultimately coupled in the following way: For any given pair of decoupled subsystems, there exists at least one other subsystem with which the two are physically coupled. For example, piston and engine block are decoupled, but both are tightly coupled with the liner and the cylinder head. Given the system architecture, failure in any of the five subsystems can propagate to any other subsystem. At the same time, if the engine block fails and the entire system stops functioning immediately, it is possible that the pistons (with which the engine block is decoupled) do not suffer damage.

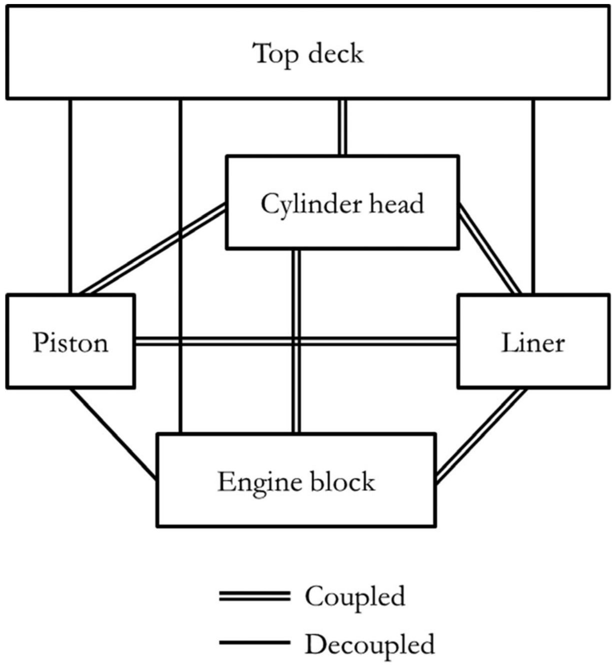

FIGURE 2 The five subsystems and their interrelationships

All Indian Railways locomotives and engines are assembled at either Diesel Locomotive Works (in the city of Varanasi in Northeastern India) or at Diesel Loco Maintenance Works (in the city of Patiala in Northern India). These massive assembly plants produce approximately 350 locomotives each year. At these plants, there are roughly as many administrative personnel as there are manufacturing personnel, which speaks to the scale of the coordination efforts required: Administrative personnel are primarily responsible for managing the upstream supply chain, including inbound logistics, processing of warranty claims, and other coordination tasks. The assembly plants further operate in collaboration with Indian Railways' Research Design and Standards Organization, which is responsible for supplier qualification for all components and subsystems. The qualification process includes visits to the suppliers' manufacturing facilities by Indian Railways experts, locomotive field-testing, and the like. Upon approval, suppliers start participating in competitive bidding for orders in a two-tier qualification system: Eighty percent of the orders are awarded to Tier 1 suppliers, twenty percent are awarded to Tier 2. All suppliers initially get qualified at Tier 2, and upon demonstration of capability and capacity, can be promoted to Tier 1. This process can take in excess of 5 years and is associated with the emergence of bilateral dependency.

3.2 | Data

In our research, we used multiple sources of data collected over a five-year period (2015–2019). We conducted site visits, interviewed key engineering and general management representatives, and analyzed buyer–supplier contracts, warranty procedures, and engine failures (see Table 1 for details). In the spirit of TCE, we wanted to understand Indian Railways' buyer–supplier relationships *in their entirety* (Williamson, 1985: 29), hence multiple sources of data relating to different aspects of the economic exchange were required. Crucially, our data collection also included interviews with Indian Railways' suppliers.

3.3 | First analysis: establishing supplier performance ambiguity

In the first analysis, we examined engine failures in a fleet of 919 diesel locomotives engines between September 2007 and August 2015. In this eight-year period, there were a total of 172 engine failures that involved failures of subsystems supplied by 17 different

TABLE 1 Data sources

Data source	Relevance for current study
Production plant site visits and follow-up correspondence (2015–2016)	
Archival data: technical documents	Technical documents such as blueprints, assembly procedures, and case histories are critical for understanding and partitioning engine structure, understanding assembly and maintenance procedures, identifying failure patterns, and understanding supply chain structure.
Archival data: warranty records	Warranty records include information on failure modes, days to failure, and supplier data. Failure case histories enable an understanding of how and when locomotive engines fail.
Management interviews: discussions with and feedback from managers, engineers, and administrative staff	*General managers*: understanding supply chain and tier decision-making processes, management of supplier relationships, institutional factors, policy constraints, and performance management. *Engineers and production experts*: understanding specific supplier relationships, technical engine details not covered in the archival data, root cause analysis, performance data, and technical details pertaining to supplier qualification decisions. *Warranty experts*: understanding the contractual and warranty claim processes, coding scheme information, administrative procedures in warranty resolution, contract enforcement.
Discussions with Railway Board Members (2016–2017)	
Interviews and correspondence with former board members	Understanding the decision-making context, strategy, institutional constraints, performance management, buyer–supplier relationships, and contract design.
Correspondence with Dan Belisle, an experienced industry expert (2015–2019)	
Personal communication with Mr. Dan Belisle	Mr. Belisle served as our fact checker and "sounding board" throughout the research project. Mr. Belisle, an experienced engineer with 35 years of industry experience, provided technical insights and information regarding industry practice in North America.
Discussions with Locomotive Technology Company, LTC (10/2019)[a]	
E-mail correspondence with one of Indian Railways' suppliers	Getting the supplier's perspective on the complexity of the buyer–supplier relationship. It was particularly important to hear from a supplier that had made a unilateral investment in asset specificity.
Discussions with Indian Railways (10/2019)[a]	
Data source	*Relevance for current study*
Discussions with Indian Railways regarding LTC	Getting the buyer's perspective on the complexity of the buyer–supplier relationship.
Discussions with former Indian Railways' general managers (10/2019)[a]	
Discussions with former general managers	Details on conflict resolution in buyer–supplier relationships, contract enforcement, and contract renegotiation/modification.

[a]Additional data collection efforts undertaken based on feedback during the review process.

suppliers. Critically, the dataset contains every relevant engine failure that occurred within the warranty period (typically one year.) The only failures missing from the dataset are the few failures where time to failure exceeded two years; for all practical purposes, our sample constitutes the entire population of failures.

All 172 failures occurred within three years of operation, which is considerably less than the rated 6-year engine life. These unusually early failures obviously have a significant impact both on warranties and contracting more generally. To clarify, even though the 172 failures involved the failure of one or more subsystems, the failures are characteristically system failures: The *entire system* is brought an abrupt shutdown. The majority of these failures occur when the locomotive is in operation and the train is moving (there were a few failures that occurred during testing). While these failures do not typically constitute a safety hazard, the economic consequences are serious. A key factor complicating failure analysis is that when an engine fails, one does not simply "pull over to the curb" and switch off the engine, as one would do when driving a car. Even if the locomotive driver observes an engine problem, the engine cannot be simply turned off immediately. Due to subsystem coupling (Figure 2), the failure can propagate to subsystems that had nothing to do with the immediate cause of the failure. Consequently, a subsystem that failed during a system failure may have been fully intact, and could be damaged in the several minutes following a failure elsewhere in the system while the engine was still running and the train was moving.

In order to analyze the failures in detail, we partitioned the engine into five physically distinct subsystems: top deck, cylinder head, piston, liner, and engine block. Failure of any combination of these five subsystems can induce a system failure, making the total of potential failure modes 32–1 = 31. The number of actual failure modes observed was 21. In 16 of these failure modes (114 out of the 172 failures), more than one subsystem failed. Further, not only did the failures occur early (consider the rated engine life of 6 years), 120 of them occurred in the first year of operation, that is, within the warranty period (see Appendix A for details on time to failure).

Some of the failure modes were puzzling even to industry experts. For example, the failure mode that involved the top deck, piston, and engine block is puzzling, because these three subsystems are not only physically decoupled from one another, they are tightly coupled with subsystems that did *not* fail. With other failure modes, arriving at plausible interpretations were more straightforward. However, establishing that something is merely *plausible* is insufficient for warranty decisions. In many cases, it is simply impossible to determine with requisite certainty what caused the failure, and whether the subsystems that failed had been defective.

3.3.1 | Joint failure patterns

Which subsystems are more likely to fail together? Can the failure of one subsystem function as a "fuse" and preempt the failure of another? In addition to establishing that system failures involve multiple subsystems, it is important to examine whether specific

TABLE 2 Two-way contingency tables of joint failures

		Cylinder Head		Piston		Cylinder Liner		Engine Block	
		No Fail	Fail	No Fail	Fail	No Fail	Fail	No Fail	Fail
Top Deck	No Fail	**58**	**14**	46	26	49	23	**46**	**26**
	Fail	**42**	**58**	57	43	77	23	**79**	**21**
	Type	Concomitant		Independent		Independent		Preemptive	
Cylinder Head	No Fail			**51**	**49**	**67**	**33**	**57**	**43**
	Fail			**52**	**20**	**59**	**13**	**68**	**4**
	Type			Preemptive		Preemptive		Preemptive	
Piston	No Fail					**91**	**12**	**63**	**40**
	Fail					**35**	**34**	**62**	**7**
	Type					Concomitant		Preemptive	
Cylinder Liner	No Fail							**86**	**40**
	Fail							**39**	**7**
	Type							Preemptive	

Note: Cells that appear in boldface indicate a statistically significant association at $p = .05$; gray shading indicates physically coupled subsystems; white cells indicate decoupling.

subsystems tend to fail jointly. To this end, we conducted a statistical analysis of all the subsystem pairs to uncover potential systematic associations; the results are presented in Table 2. The analysis reveals three joint failure patterns:

1. *Preemptive failure*: The failure of one subsystem is associated with a lower failure rate in another. Such fuse effect was the most common type of relationship: Six subsystem pairs exhibited preemptive failures.
2. *Concomitant failure*: The failure of one subsystem is associated with a higher failure probability in another. There were two subsystem pairs that exhibited concomitant failures. An obvious potential cause of concomitant failure is the failure of one subsystem propagating to another to which it is directly physically coupled. Of course, it is also possible that two subsystem failures have a common cause.
3. *Independent failure*: The most counterintuitive joint failure mode is when two subsystems may fail jointly, but the failures are statistically independent in that the failure of one does not change the probability of the other failing.

Importantly, the first two failure modes are matters of degree: No preemptive failure is such that the failure of one subsystem *always* prevents the other from failing, it merely lowers the probability. The same applies to concomitant failure.

The joint analysis focused on two-way associations: Does the failure of one subsystem change the frequency of another failing? Of course, higher-order interactions are also possible: In a three-way interaction, we would ask whether the failure of a third subsystem changes the strength of the interaction of two other subsystems. In a system consisting of five subsystems, there could be 10 different three-way interactions, 5 four-way interactions, and 1 five-way interaction. The empirical challenge in separating these interactions is the fact that they are strongly confounded with one another. This confounding is relevant but not crucial to the analyses, therefore, we present the details in Appendix B. The primary conclusion is that with these failure data, it is impossible to examine the failures by specifying statistical models where the failure modes constitute the independent variables. We read this as yet another manifestation of supplier performance ambiguity: Because the interactions are confounded with one another, we cannot determine with sufficient certainty (for the warranty process) how and why an engine failed.

3.3.2 | Supplier performance ambiguity, summary

In summary, we present three empirical manifestations of supplier performance ambiguity. The first is that about two thirds of engine failures involve the failure of at least two subsystems, typically supplied by different suppliers. A failure analysis tends to be ineffective in uncovering the root cause of the failure, causing a *failure metering problem*. This is supported by the second observation that there are concomitant failures: Failure of one subsystem is associated with the higher probability of failure in another. Of the three joint failure modes, concomitant failures in particular are the most troublesome from a warranty point of view. The third observation lends further support to the first two observations: Failure interactions are confounded with one another, which makes statistical analysis subject to considerable collinearity. In an attempt to use the failure modes to predict time to failure using Cox regression (Cox, 1972), we indeed ran into serious collinearity problems (for details, see Appendix B). From the failure data, it is therefore impossible to determine whether subsystem interactions predict time to failure. This is a perfect example of bounded rationality and supplier performance ambiguity in action.

These failure patterns would not cause contracting problems if the supply chain were not fragmented. Specifically, if a single supplier provided the top decks, the pistons, and the cylinder liners, concomitant failures and confounding of the interactions would not create contractual complications: Problems would be isolated to within the supplier, who would have a *high-powered incentive* (Williamson, 1985: 90–91) to do everything it can to address the problem. However, because these subsystems are supplied by different suppliers, performance ambiguity is unavoidable. This brings us to our second analysis: How are warranties managed under supplier performance ambiguity? This is a recurring problem at Indian Railways: Largely due to supplier performance ambiguity, the data indicate that Indian Railways encountered contractual conflicts in 11% of the failure cases we analyzed.

3.4 | Second analysis: contracting under supplier performance ambiguity

In this section, we explore how Indian Railways manages its supply chain under conditions of bilateral dependency and supplier performance ambiguity. Indian Railways' approach can be described as *pragmatic* and *uniform*. It is *pragmatic* in that the overarching objective is to minimize unplanned locomotive downtime. Specifically, the focus is less on technical failure analysis and more on rapid recommissioning. This requires effective management of both spare parts and the first-tier supplier base. The pragmatic approach in understandable: Losses in revenue for an unavailable locomotive are estimated to be $6,000 per day per unavailable locomotive (to interpret this in the U.S. context, the figure should be multiplied by a factor of twenty to adjust for purchasing power parity). Indeed, a former General Manager of one of the manufacturing plants noted that the overall cost of system failure and its ripple effects (e.g., train delays) across the railroad exceed the replacement cost of the failed subsystems by an order of magnitude. This is why Indian Railways must maintain a swift supply of subsystems to its assembly facilities, and maintain a buffer inventory of subsystems.

Indian Railways' approach is *uniform* in that even though the subsystems provided by suppliers are heterogeneous with respect to function, complexity, and price, a standard contract template is used for all suppliers. The standard template contains, among other things, a warranty clause that the supplier will not be held responsible for potential consequential damages. Specifically, if subsystem *A* fails and it is obvious that subsystem *B* failed as a consequence, Indian Railways makes the warranty claim only to the supplier of subsystem A and absorbs the cost of the consequential failure of subsystem *B*. Indemnifying suppliers against consequential damages is standard industry practice, and signals a credible commitment on the part of Indian Railways. A former General Manager in charge of both production units noted that the ability of many suppliers to pay for consequential damages is limited. Being held responsible for consequential damages would create at least two problems. First, it would create a disincentive to contract with Indian Railways, effectively making the already limited supplier pool even more limited. Second, it would lead to suppliers charging higher prices, because the supplier would not agree to bear the full cost of what would effectively constitute Indian Railways' insurance policy against consequential damages.

Because of the pragmatic and uniform approach, the intricacies of multiple subsystem failures and interdependencies have been considered contractually irrelevant: The goal of the failure analysis is simply to identify the subsystems that need to be replaced, and replacement is done as quickly as possible. This is reminiscent of MacDuffie's (1997) observation in automobile assembly: A root cause analysis is not necessarily conducted in order to prevent failures from occurring, but rather, as a matter of organizational expediency.

The question of consequential damages is largely irrelevant, however, because as our data demonstrate, situations in which the failure of one subsystem unambiguously led to the failure of another are rare: Concomitant failures are matters of degree, and often, there is considerable residual ambiguity about the direction of causality. Under such ambiguity, Indian Railways uses a simple principle: If a subsystem fails, it will be considered, for contractual interpretation purposes, a defective subsystem. Whether the subsystem was in fact defective is irrelevant, it is used as a *factual premise* (Simon, 1997: 60) in contract enforcement. If the supplier contests the decision, Indian Railways can exert even more pressure by unilaterally adjusting payments to the supplier. If the supplier disagrees, an arbitration process is initiated. However, such processes are risky as they may jeopardize the bilaterally dependent buyer–supplier relationship. Finally, Indian Railways may also unilaterally decide to downgrade a supplier, which has a significant impact on the order volume the supplier receives. The mere threat of downgrading may operate as an effective incentive for Tier 1 suppliers not to challenge Indian Railways' decisions.

In summary, Indian Railways has the option of adopting "the muscular approach" if it so chooses, effectively pushing the cost of failures upstream to its Tier 1 suppliers. However, we became convinced that Indian Railways' management readily understood this to be myopic, because dependency was bilateral. Consequently, Indian Railways is more inclined, and well advised, to adopt "the benign approach." We saw two kinds of evidence of this in our analysis, one pertains to the warranty resolution process, and the other, to the fundamental transformation. These analyses are discussed in the following sections.

3.4.1 | Warranty resolution

The locomotive engine is remarkably robust to minor failures, for example, the failure of a single piston ring will not induce a system failure. However, the failure of a major subsystem will almost certainly cause a system failure. After such a failure, the locomotive no longer supplies the requisite power to move the train, and must be towed to one of Indian Railways' fifty maintenance depots. Engineers then conduct a teardown inspection to identify the failed subsystems, and if possible, the root cause. In cases with extensive damage, the evidence to conduct an effective root cause analysis of failure may not be available. Often, engineers rely on their vast experience with similar failures to draw conclusions regarding root causes. In practice however, the warranty department for Indian Railways tends to treat all failed subsystems and components within the warranty period as defective, and files warranty claims for all of them.

Disassembled subsystems can be shipped back to suppliers so that they can conduct their own analyses. However, in the case of multiple subsystem failures, the suppliers are obviously provided only with the subsystems that they had supplied. Suppliers seldom have dedicated engineers at Indian Railways' maintenance depots, so they have no empirical access to system failures, only to the failures of their own subsystems. However, if a particular failure pattern is identified, Indian Railways does invite suppliers to conduct joint inspections and analysis to facilitate root cause identification. These joint inspections, meant as a forum for joint problem-solving, can be interpreted as applying "the benign approach."

Conflicts do occur. Indeed, Indian Railways has downgraded critical suppliers when there has been a reason to believe they have not met contractual specifications. In one case, failures in the subsystems provided by a supplier unexpectedly spiked. Indian Railways interpreted this as an indication of poor supplier quality, which the supplier denied. Indian Railways employed "the muscular approach" and unilaterally downgraded the supplier to Tier 2, and upgraded an alternative supplier to the subsystem to Tier 1. But Indian Railways quickly realized that the problem persisted. Indeed, the failed subsystem was one that exhibited concomitant failure with another subsystem, therefore, it was plausible that the subsystem supplied was not initially defective, but it failed for systemic reasons. Indian Railways was eventually able to uncover the actual cause of the failures, and took corrective (maintenance) action, which included reinstating the supplier to its Tier 1 status.

While it may be tempting to invoke opportunism to make sense of Indian Railways' application of "the muscular approach" in the case of the downgrading decision, we saw no evidence of anything but an honest disagreement between two boundedly rational contracting parties. Specifically, the supplier in question was a Tier 1 supplier, which meant that buyer–supplier dependency had developed bilateral features. It would have been myopic for Indian Railways to exploit its supplier by engaging in opportunistic action. The conclusion that the subsystems supplied were defective may have been misplaced, but given Indian Railways' understanding of failures, it was nonetheless a (boundedly) rational decision. Note that opportunism is not merely about self-interest, it implies intentional distortion and self-disbelieved promises (Williamson, 1975: 255). We saw no evidence of this in the downgrading decision, particularly since the downgrading decision had adverse economic consequences not only to the supplier but also to the bilaterally dependent buyer: Indian Railways did not benefit from the downgrading decision in any way, just the opposite. Finally, making the decision to reinstate the supplier's Tier 1 status worked toward increasing Indian Railways' credibility as a supply chain partner. We interpret this decision as a clear indication that the downgrading decision was not opportunistic.

Most fundamentally, we find the *assumption* of opportunism problematic. In our analyses, we found it both useful and impartial to turn opportunism into something to be established empirically instead of being assumed (cf. John, 1984; Lumineau & Oliveira, 2020). In the context of Indian Railways in particular, and risky long-term contracting in general, we find it unfathomable that exchange partners would enter into a potentially long-term relationship, but at the same time, assume that the other party is inclined to intentionally distort data and make self-disbelieved promises. It is, however, important to understand that honest disagreements can be sufficient to bring about friction and increased transaction costs (Alchian & Woodward, 1988: 66).

3.4.2 | The fundamental transformation and the benign approach: the case of LTC

To gain further insight into how a mutually beneficial long-term contracting relationship can be developed and maintained under conditions of supplier performance ambiguity, we now examine the relationship between Indian Railways and Locomotive Technology Corporation, LTC (a pseudonym), spanning ten years. The goal is to shed light on the temporal dynamics of contracting, and illustrate how the benign approach may be efficient in the long run. LTC is currently a Tier 1 supplier of one of the five subsystems we analyzed, and therefore, its failures are included in our dataset. LTC initially approached Indian Railways in an attempt to obtain qualification as Tier 2 supplier. Several years after obtaining Tier 2 status, LTC was upgraded to Tier 1. Crucially, our data covers the period over which the fundamental transformation between Indian Railways and LTC occurred.

In order to obtain Tier 2 qualification, LTC had to make dedicated investments in specialized production equipment, which would sit idle whenever it was not used to work on orders from Indian Railways (personal communication with the CEO of LTC, October 7, 2019). In addition, Indian Railways made no guarantees for future orders (personal communication with a Senior Manager at the Mechanical Department of an Indian Railways production plant, October 5, 2019), which LTC confirmed as well. Again, while this could plausibly be read as a manifestation of "the muscular approach" if not downright opportunism, the reason for refraining from offering any provisions is not a choice but an institutional constraint: Assurances about future orders would have been not only a violation of the government's purchasing policies but possibly also illegal.

Why did LTC decide to commit to specificity without explicit guarantees? LTC informed us that Indian Railways was considered "a prestigious customer [for LTC] to showcase [its] capabilities to other customers" (personal communication with the CEO of LTC, October 7, 2019) Conversely, Indian Railways management is fully aware of this prestige and knows that this attracts potential suppliers. Therefore, the commitment to specificity was possible even in the absence of formal safeguards (see Kang, Mahoney, & Tan, 2009).

Upon Tier 2 qualification, LTC started supplying subsystems. Currently, as a Tier 1 supplier, the subsystems LTC supplies to Indian Railways are "highly critical for production and maintenance of Diesel Locomotives and hence ensuring smooth running of [operations]...supplies of these components remain under constant watch as any failure can result into halt of production" (communication with a Senior Manager at the Mechanical Department at an Indian Railways production plant, October 5, 2019). Critically, the LTC supplies subsystems that are subject to performance ambiguity.

Our examination of the failure data revealed that over a period of approximately six years, Indian Railways made a total of 141 warranty claims to LTC. Roughly one fourth of these claims involved some disagreement, with several cases escalating to Indian Railways deciding to penalize LTC financially. Importantly, none of these conflicts occurred in the early stages of the relationship, before the fundamental transformation had occurred and bilateral dependency developed. Warranty records indicate that early in the relationship, LTC did not contest any of Indian Railways' decisions. But once the fundamental transformation had occurred and bilateral dependency had developed, LTC felt more comfortable voicing its concerns. Of course, it may be that over time, LTC management felt more confident in its own manufacturing capabilities, and consequently, more comfortable in voicing disagreements. In the early stages of the relationship, contesting Indian Railways or invoking arbitration could have jeopardized the relationship. Being sufficiently comfortable to voice one's concerns is a sign of voluntary, not coerced, cooperation. In sum, we read the emergence of conflict not as a problem but a sign of a healthy buyer–supplier relationship.

Interestingly, even in the case of conflicts that occurred after the fundamental transformation, LTC chose not to invoke arbitration, even when LTC felt it was being held responsible for a problem without sufficient cause. Implicitly following the tenets of TCE, LTC did escalate the problem, but did it without involving third parties. Specifically, LTC requested that Indian Railways' internal Railway Board seek a resolution. Importantly, Railway Board is not a third party, it is the board of directors of Indian Railways. As Williamson (1975: 30) pointed out, not involving third parties secures a cooperative (as opposed to adversarial) resolution of problems. Although once bilateral dependency had been established, and LTC could have adopted "the muscular approach," it decided to adopt "the benign approach." We read the abandonment of "the muscular approach," both by the buyer and the supplier, as an example of a mutual credible commitment, which is essential in safeguarding a long-term exchange relationship. Our data also show that joint problem-solving efforts between Indian Railways and LTC were initiated during later stages of the relationship, after the fundamental transformation had occurred. These efforts involve Indian Railways inviting LTC engineers to conduct the teardown inspections jointly.

The records of a high-level Indian Railways' internal meeting in February 2019 clearly echo "the benign approach:"

> "[Suppliers] whose performance is not up to the mark *must* be brought to the table...giving them opportunity to improve the delivery as per the *mutually* accepted terms and conditions of the contract. Should suppliers decide not to improve performance, they need to be systematically weeded out." (Indian Railways internal minutes of meeting, distributed on March 15, 2019; emphasis added).

The key point in this passage is the idea of engaging in joint problem-solving as the first resort, and termination of the exchange relationship as the last. In addition, giving suppliers a voice, which involves the buyer sacrificing power over the supplier to improve the supply chain relationship (Brown & Hendry, 1997), can also be regarded as a form of credible commitment.

3.4.3 | Rethinking the warranty process

A straightforward contractual solution to supplier performance ambiguity problem would be to purchase the entire engine from a single supplier. Indeed, in the 1990s, some railroads adopted what is best described as a *performance-based contracting* approach (Guajardo, Cohen, & Netessine, 2016). This mitigated the supplier performance ambiguity problem, but at the same time, required purchasing the entire locomotive engine from a single supplier. Ultimately, this turned out to be prohibitively expensive. The industry then shifted toward a component-centric fragmented approach akin to *time-and-material contracts* (Guajardo et al., 2016: 960). Although economically more efficient, this approach did little to alleviate the supplier performance ambiguity problem. The industry also experimented with preventive maintenance: swapping of subsystems every three years, which is half their rated life of six years. Although comparatively more efficient than earlier approaches, our analyses of the failure data suggest that this does not solve the problem: Majority of the failures occurs within one year of commissioning the engine. Furthermore, this approach is particularly problematic if only some subsystem suppliers participate: How much sense would it make to swap the pistons if the cylinder liners (with which pistons concomitantly fail) are not also replaced? However, despite the obvious problems associated with what was labeled the Unit Exchange Program, some variants of preventive maintenance have been widely adopted in North America in particular. This is because railroads believe that the revenues lost due to downtime exceed the costs of adopting these programs by several orders of magnitude.

Our analyses can offer guidance on how to rethink the warranty process. The central idea is to shift back toward a system-centric perspective in an analysis-driven, selective manner. Specifically, some subsystems are more problematic than others are not because they fail more often, but because in case of failures, they interact with other subsystems (Novak & Wernerfelt, 2012). Concomitant failures are particularly problematic (Agrawal et al., 2017).

A closer reflection of the results (Table 2; also see Appendix B) reveals that most near-zero correlations involve the top deck. Further, once we include the engine block, we have accounted for nearly all cases. The three subsystems that comprise the power assembly—cylinder head, piston, and liner—are associated with the most problematic interactive failure patterns. Bundling these three into one *power assembly subsystem*, to be provided by one supplier, would alleviate contractual problems associated with warranties. The interactions of the three subsystems would still present a challenge to the power assembly supplier, but the problem would transform from a contractual problem to an internal engineering and management problem for the supplier. This has two benefits. One is that problems are always tackled more effectively within organizations than between organizations, indeed, this is why firms exist in the first place (Kogut & Zander, 1992; Williamson, 1975). The second is that contractual enforcement costs would be lower: In the case that two or three subsystems within the power assembly failed, only one warranty action would be needed.

To examine the implications of adopting the *bundling approach*, we reanalyzed our data under the assumption that any failure that involved several subsystems of the power assembly (cylinder head, piston, or liner) would result in just one warranty action (warranties are claimed "at the level of purchase"). The key question is obviously the magnitude of the reduction in the total number of warranty actions. Our analysis reveals that bundling would reduce the overall number of warranty actions by 17%, which is a significant reduction. Further, the number of warranty claims associated with the subsystems within the bundle would shrink by 30%. In addition to the direct savings from having fewer warranty claims, bundling would better channel relevant information to the supplier who is incentivized to monitor quality more closely. Specifically, instead of receiving only the failed subsystem from Indian Railways, the supplier would receive the entire power assembly, which would enable a more effective failure analysis.

Is bundling feasible? The obvious downside is that the small-numbers bargaining situation with suppliers would intensify, ultimately leading to *bilateral monopoly* concerns: Having one powerful buyer and one powerful supplier can lead to a situation in which the transacting parties are "inclined to expend considerable resources bargaining over the price at which the exchange is to take place" (Williamson, 1975: 28). The benefits of aggregation must be weighed against this tradeoff. However, based on our discussions with Indian Railways already in May of 2016, not only did they indicate that having one supplier provide the entire power assembly is feasible, it was in fact something that they had done in the past. The bundling approach could therefore be adopted within the existing purchasing policy framework, no restructuring of the supply chain would be required.

To further evaluate the viability of the bundling approach, we contacted Mr. Belisle to inquire whether other railroads had anything similar in place, and whether it had been successful in improving the warranty claims process and reducing failures. Mr. Belisle (personal communication, November 17, 2019) confirmed that mixing and matching subsystems provided by different suppliers—"unbundling"—inevitably increases supplier performance ambiguity. He further confirmed that the failure rates of "unitized power assemblies" were indeed lower: In the mix-and-match approach, problems are caused specifically by subsystems having different origins. We cross-checked this with the assembly procedures for the power assembly, and concluded that the tolerances are indeed so tight, and the assembly procedure so complex, that having different suppliers first supply the components to Indian Railways for assembly may be comparatively inefficient compared to the arrangement where the entire power assembly is outsourced to a supplier. It is reasonable to expect suppliers to be both more competent and better incentivized (due to high-powered incentives) to ensure the high quality of power assemblies.

4 | REFLECTIONS AND CONCLUSIONS

Many buyer–supplier relationships are characterized by power imbalances and asymmetric dependency: The comparatively stronger buyer takes the weaker, smaller supplier "hostage" and unilaterally sets the terms of the contract (e.g., Liker & Choi, 2004; MacDuffie & Helper, 2006). However, even purely transactional exchange where the costs of switching suppliers are low can ultimately develop into a relationship where identities of both exchange partners start to matter, giving rise to relational contracting. This *fundamental transformation* is often overlooked in research on supply chain structures. However, specifically because of the fundamental transformation, it would ultimately be myopic for the comparatively stronger party to adopt "the muscular approach" (Williamson, 2008: 10): Even if the pool of potential suppliers were large at the outset, the fundamental transformation means that at renegotiation, the suppliers who were first selected would have an advantage over new bidders. Due to the fundamental transformation, suppliers are no longer "captive" in the sense that switching costs would be consequential for them but not for the buyer. Instead, the buyer and the supplier become bilaterally dependent, and the buyer is well advised not only to demand credible commitments from the supplier, but also to give them in turn: It takes mutual credible commitments—not a "taking" but an "exchange" of hostages (Williamson, 1983: 519)—to support the long-term exchange relationship.

In this study, we have examined credible commitments in the context of the first-tier supply chain relationships of a locomotive engine manufacturer and railroad operator. Through this research, we have been able to contextualize credible commitments empirically. In the context of locomotive engines, these credible commitments are both important and constantly tested, because engines fail both sufficiently often to warrant attention, and in ways that defy even the best engineering understanding. Because different engine subsystems are provided by different suppliers, designing a warranty process that establishes credibility in the eyes of the supplier base is central. Indian Railways has learned from past experience that unilateral action should be replaced by mutual problem-solving and cooperation. The central problem of interest to us is how this is incorporated into not only the letter of the contract, but also to the process of contracting as it unfolds over time.

4.1 | Relational contracting and organizational basis of trust

An analysis of mutual credible commitments can shed important light on the elusive and multifaceted notion of *trust* in supply chain relationships: How, exactly, does one build a long-term trust-based relationship (Liker & Choi, 2004), and in what sense is trust relevant? The answer is complex, which largely stems from two sources. One is that the very definition of trust is context dependent. Highly intertwined with the question of definitions is variability with regard to how researchers view the basis of trust: Where, exactly, is the locus of trust in interorganizational relationships? To be sure, it is contrived to think of interorganizational trust as being reducible to trust between their respective individuals: "[A]ggregating interpersonal trust as a proxy for interorganizational trust ignores the influence of social context… that [constrains] and [orients] its members" (Zaheer, McEvily, & Perrone, 1998: 154).

In terms of the definition of trust, we find it useful to start at Williamson's (1996: 216) words of caution: "Not only does the use of familiar terms (like trust) invite us to draw mistaken parallels between personal and commercial experience, but user-friendly terms do not encourage us to examine the deep structure of organization. Rather, we need to understand when credible commitments add value and how to create them, when reputation effects work well, when poorly, and why. Trust glosses over, rather than helps unpack, the relevant microanalytic features and mechanisms." Williamson's observation is important in the context of supply chain management in particular, because the majority of empirical examinations that involve trust are about supply chain relationships, that is, relationships between legal entities, not individuals (e.g., Hill, Eckerd, Wilson, & Greer, 2009; Ireland & Webb, 2007; Johnston, McCutcheon, Stuart, & Kerwood, 2004; Zhang, Viswanathan, & Henke Jr., 2011). In the spirit of Williamson's prescription, we must unpack the mechanisms, and in the spirit of Barden and Mitchell's (2007) finding that the organizational factors trump the personal in boosting credibility, should feel compelled to explore the organizational basis in particular.

Definitions are always context specific, and trust is no exception. Operations management scholars such as McCarter and Northcraft (2007) cite Rousseau, Sitkin, Burt, and Camerer's (1998: 395) definition of trust as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another." But this is not a general definition, as it is easy to think of many contexts where trust is relevant but accepting vulnerability is unfathomable. In our exposition (much like in the context of insurance, for example), attention turns to safeguards. We want to set aside labels and focus on the principles and processes that serve as the foundation of long-term buyer–supplier relationships. A useful way to understand how mutual credibility might be induced is to examine how the warranty resolution process developed as a result of the fundamental transformation: "The muscular approach" gave way to "the benign approach" (Williamson, 2008: 10). As Indian Railways realized, an effective warranty resolution system would have to focus on joint problem-solving instead of organizational expediency. Indeed, Liker and Choi (2004: 108) noted that partnering with suppliers should involve structured, selective information sharing and joint improvement activities (see also Zaheer et al., 1998)—whether we use the label "trust" to describe such cooperation seems redundant. In fact, our interpretation is that the term "trust" is used effectively as shorthand for whatever the alternative to arm's-length transactional contracting is—in TCE, this alternative is called *relational contracting*. However, such use of the term can be misleading, because cooperation may occur with or without trust: A supplier may voluntarily collaborate with a buyer simply because it has made sunk investments and has excess capacity (MacDuffie & Helper, 2006: 465–466). Similarly, both expectations of future gains and the ability to impose penalties may induce cooperation with no reference to psychological states involving the acceptance of vulnerability. Rather, our findings suggest that traceability may be more relevant (Skilton & Robinson, 2009). Traceability in supply chains can be enhanced not by "building trust" but by reducing complexity and by sharing information.

In the context of supplier performance ambiguity, improving traceability is key in improving quality (Skilton & Robinson, 2009), and information sharing is crucial in building credibility: Indian Railways understands the value of cooperative long-term relationships, and how and why subsystem performance is ambiguous, which leads to supplier performance ambiguity. In the past, this knowledge was internal in that a supplier of cylinder liners would receive warranty claims only for failed liners that had been involved in a system failure. For the purposes of contractual expediency, failed liners were treated as if they had been defective. By engaging in joint problem-solving and helping suppliers understand how the engines—that house several subsystems in addition to the ones they supply—fail in operation is central in nurturing the long-term relationship. Even though root causes of failures may elude even the most experienced experts, the best chances of advancing an understanding of failures is buyer–supplier collaboration (Johnston et al., 2004). However, we want to underscore the organizational basis of this cooperation: While cooperation ultimately takes place between individuals, it must be embedded within organizational structures, procedures, credible commitments, and incentives.

In summary, this study highlights the gains available to contracting parties that eschew "the muscular approach" in favor of "the benign approach," and focus instead on establishing an organizational foundation for long-term cooperation. Adopting "the benign approach" is challenging, but we argue that it leads to an improved contracting relationship if the parties are bilaterally dependent. In addition, it may also foster higher technical performance. The approach is likely to be useful in industries that require relation-specific investments—the automobile and aviation industries are good examples. More generally, "the benign approach" provides the requisite caution in approaching ambiguous performance.

4.2 | Limitations and future research

Both our exposition here and other supply chain management investigations that focus on undesirable events such as failures (e.g., Skilton & Robinson, 2009) tend to build on the premise that while there are complicating factors, we can ultimately understand failures better. Indeed, the empirical analyses as well as the solutions proposed here (e.g., the bundling approach) work toward that objective. However, as the frequent failures of root cause analysis suggest, perhaps emergent behaviors will remain significantly inaccessible for two reasons.

First, systems behave in ways that are not only unpredictable, but even more fundamentally, unfathomable: Many things that could happen have in fact not happened to date, and it is impossible to anticipate something that cannot be fathomed. We submit that in light of the track record on failures in various settings, this is a plausible position. For example, in 2018, there were 11,476 rail incidents and accidents in the United States that led to a total of 7,999 injuries and 878 fatalities (Statista.com, 2019). Further, in his empirical analysis of risks associated with transporting oil by rail, Mason (2018: 1) observed that statistics pointed "to a steady stream of train derailments in the U.S. between 2009 and 2014." This suggests that failure is not an anomaly but rather, endemic to the artifact.

Why are we unable to do anything about this "steady stream" of failures? To be sure, intentionality cannot be the explanation: The proposition that we simply, against our better judgment, let accidents happen and people die is not plausible. Another example is Weick's (2004) discussion of how Centers for Disease Control and Prevention misdiagnosed West Nile Virus. We have dealt with virus outbreaks for decades, why do we still fail? There is at least some evidence that suggests that we are not always able to conduct successful root cause analyses and use the insights gained to prevent similar failures from reoccurring (Starbuck & Farjoun, 2005).

There is something strangely ambiguous about how failures in complex systems occur and why, even to the point that scholars such as Perrow (1984) have argued that accidents not only occur, we should expect them to happen, they are *normal*—hence the title of Perrow's book *Normal Accidents*. Indeed, the steady stream of train derailments, failures to diagnose viruses, and the general challenges associated with root cause analysis suggest that human beings are not limited merely in their orientation toward future events, but also subject to what could be labeled *bounded hindsight*. Bounded hindsight arises from the fact that understanding of past events may be not so much a process of analysis of evidence as it is one of *reconstruction of events*, where we invoke assumptions and schemas whenever empirical evidence is lacking (cf. March, 2010; Tetlock, 2005). Further, our understanding is obviously fundamentally conditioned by the theories and the tools we use. A case in point, the application of root cause analysis is founded—as the label implies—on the premise that there *is* a root cause. This is precisely the kind of a reductionist premise that underpins our thinking and that safety scientists have criticized as misguided (Dekker et al., 2011).

One of the limitations—and a potentially fruitful avenue for future research—would be to tackle the question of bounded hindsight in more detail, both theoretically and empirically; in this study, we have barely scratched the surface. In particular, it would be illuminating to investigate the detailed processes by which transacting parties ascribe opportunism versus bounded hindsight to failures. It would also be useful to examine the way evidence, expertise, and judgment are invoked in these processes. What kind of data are used? What are the key characteristics of root cause analyses that have been successful in that they have actually prevented future failures? Are these analyses successful due to the expertise and experience of those involved, or because in that particular failure, a singular root cause did indeed exist? What bases of expertise are involved in the processes? How is ambiguity resolved?

These questions will likely become even more relevant as information technology offers better ways of collecting data on failures as they occur. For example, contemporary sensor technology enables the collection of detailed real-time data that has the potential to significantly enhance the effectiveness of preventive maintenance, and may shed more light into failures. The most interesting question is whether these data can ultimately help us understand the systems so much better that we see a significant drop in unavoidable failures.

4.3 | Conclusion

Long-term buyer–supplier relationships tend to involve bilateral dependency, which means that they must be understood from the point of view of cooperation and mutual credible commitments instead of power dynamics, where the more powerful party makes unilateral, take-it-or-leave-it decisions. Understanding power in supply chains is of course critical, but it is equally crucial to understand that cooperative forms may be more effective. This "benign approach" lies at the foundation of transaction cost economics, which we have applied in this study to examine the complex buyer–supplier relationships between Indian Railways and some of its first-tier suppliers. The dynamics of these relationships reveal a gradual shift from a power-based approach to voluntary cooperation that emphasizes joint problem-solving and contractual arrangements that establish mutual credible commitments. This understanding helps us shed empirical light on the organizational basis of trust in buyer–supplier relationships, which is embedded more in the contractual nuances, interorganizational processes, and cooperation than interpersonal relationships. Understanding and building the organizational basis of trust is crucial, because it is simply too risky for long-term buyer–supplier relationships to be based on interpersonal trust. In the spirit of transaction cost economics, these considerations must be farsighted and seek mutual gains through cooperation.

ACKNOWLEDGMENT

We thank the review team, led by Dr. Sriram Narayanan, for thorough and constructive reviews and guidance throughout the review process. We also gratefully acknowledge the invaluable insights that both current and former executives and engineers at Indian Railways shared with us over the past five years. We are equally indebted to Dan Belisle, whose technical feedback and broader industry knowledge strongly complemented our analyses throughout this study. We further thank the CEO of Locomotive Technology Company (pseudoname) for his willingness to share information. More generally, we applaud all the work of executives and engineers on all sides of the transactions. We extend our gratitude in particular to those who are working to make the complex transactions described in this study possible. All the contractual challenges described in this study notwithstanding, we want to convey an expressly positive and constructive message of how Indian Railways' approach can potentially serve as a model for other organizations.

We dedicate this article to the memory Dr. Oliver Williamson, whose profound insights on the governance of contractual relationships have served as an intellectual foundation for our arguments both in this article and in our work more broadly.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

ORCID

Akhil Bhardwaj https://orcid.org/0000-0002-2231-0478
Mikko Ketokivi https://orcid.org/0000-0003-4510-4949

REFERENCES

Agrawal, A., Muthulingam, S., & Rajapakshe, T. (2017). How sourcing of interdependent components affects quality in automotive supply chains. *Production and Operations Management*, *26*(8), 1512–1533.

Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *American Economic Review*, *62*(5), 777–795.

Alchian, A. A., & Woodward, S. (1988). The firm is dead; long live the firm: A review of Oliver E. Williamson's the economic institutions of capitalism. *Journal of Economic Literature*, *26*(1), 65–79.

Anand, G., & Gray, J. V. (2017). Strategy and organization research in operations management. *Journal of Operations Management*, *53–56*, 1–8.

Bai, X., Sheng, S., & Li, J. J. (2016). Contract governance and buyer–supplier conflict: The moderating role of institutions. *Journal of Operations Management*, *41*, 12–24.

Baldwin, C. Y., & Clark, K. B. (2000). *Design rules: Volume 1. The power of modularity*. Cambridge, MA: MIT Press.

Barden, J. Q., & Mitchell, W. (2007). Disentangling the influences of leaders' relational embeddedness on interorganizational exchange. *Academy of Management Journal*, *50*(6), 1440–1461.

Benton, W. C., & Maloni, M. (2005). The influence of power driven buyer/seller relationships on supply chain satisfaction. *Journal of Operations Management*, *23*(1), 1–22.

Brinkhoff, A., Özer, Ö., & Sargut, G. (2015). All you need is trust? An examination of inter-organizational supply chain projects. *Production and Operations Management*, *24*(2), 181–200.

Brown, J. E., & Hendry, C. (1997). Industrial districts and supply chains as vehicles for managerial and organizational learning. *International Studies of Management & Organization*, *27*(4), 127–157.

Cai, S., Jun, M., & Yang, Z. (2010). Implementing supply chain information integration in China: The role of institutional forces and trust. *Journal of Operations Management*, *28*(3), 257–268.

Carroll, J. S., Hatakenaka, S., & Rudolph, J. W. (2006). Naturalistic decision making and organizational learning in nuclear power plants: Negotiating meaning between managers and problem investigation teams. *Organization Studies*, *27*(7), 1037–1057.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, *34*(2), 187–220.

Crook, T. R., & Combs, J. G. (2007). Sources and consequences of bargaining power in supply chains. *Journal of Operations Management*, *25*(2), 546–555.

Cui, Z., Loch, C., Grossmann, B., & He, R. (2012). How provider selection and management contribute to successful innovation outsourcing: An empirical study at Siemens. *Production and Operations Management*, *21*(1), 29–48.

Dekker, S., Cilliers, P., & Hofmeyr, J.-H. (2011). The complexity of failure: Implications of complexity theory for safety investigations. *Safety Science*, *49*(6), 939–945.

Dutta, J., & Agarwal, A. K. (2017). Diesel locomotives of Indian Railways: A technical history. In A. K. Agarwal, A. Dhar, A. Gautam, & A. Pandey (Eds.), *Locomotives and rail road transportation* (pp. 9–23). Singapore: Springer.

Goldberg, V. P. (1976). Regulation and administered contracts. *Bell Journal of Economics*, *7*(2), 426–448.

Gray, J. V., & Handley, S. M. (2015). Managing contract manufacturer quality in the presence of performance ambiguity. *Journal of Operations Management*, *38*, 41–55.

Grover, V., & Malhotra, M. K. (2003). Transaction cost framework in operations and supply chain management research: Theory and measurement. *Journal of Operations Management*, *21*(4), 457–473.

Guajardo, J. A., Cohen, M. A., Kim, S.-H., & Netessine, S. (2012). Impact of performance-based contracting on product reliability: An empirical analysis. *Management Science*, *58*(5), 961–979.

Guajardo, J. A., Cohen, M. A., & Netessine, S. (2016). Service competition and product quality in the U.S. automobile industry. *Management Science*, *62*(7), 1860–1877.

Handfield, R. B. (1993). A resource dependence perspective of just-in-time purchasing. *Journal of Operations Management*, *11*(3), 289–311.

Handfield, R. B., & Bechtel, C. (2002). The role of trust and relationship structure in improving supply chain responsiveness. *Industrial Marketing Management*, *31*(4), 367–283.

Handley, S. M., & Benton, W. C. (2012). The influence of exchange hazards and power on opportunism in outsourcing relationships. *Journal of Operations Management*, *30*(1–2), 55–68.

Hill, J. A., Eckerd, S., Wilson, D., & Greer, B. (2009). The effect of unethical behavior on trust in a buyer–supplier relationship: The mediating role of psychological contract violation. *Journal of Operations Management*, *27*(4), 281–293.

Hoetker, G., Swaminathan, A., & Mitchell, W. (2007). Modularity and the impact of buyer–supplier relationships on the survival of suppliers. *Management Science*, *53*(2), 178–191.

Hora, M., Bapuji, H., & Roth, A. V. (2011). Safety hazard and time to recall: The role of recall strategy, product defect type, and supply chain player in the US toy industry. *Journal of Operations Management*, *29*(7–8), 766–777.

Ireland, R. D., & Webb, J. W. (2007). A multi-theoretic perspective on trust and power in strategic supply chains. *Journal of Operations Management*, *25*(2), 482–497.

Jap, S. D., & Anderson, E. (2003). Safeguarding interorganizational performance and continuity under ex post opportunism. *Management Science*, *49*(12), 1684–1701.

John, G. (1984). An empirical investigation of some antecedents of opportunism in a marketing channel. *Journal of Marketing Research*, *21*(3), 278–289.

Johnston, D. A., McCutcheon, D. M., Stuart, F. I., & Kerwood, H. (2004). Effects of supplier trust on performance of cooperative supplier relationships. *Journal of Operations Management*, *22*(1), 23–38.

Kang, M.-P., Mahoney, J. T., & Tan, D. (2009). Why firms make unilateral investments specific to other firms: The case of OEM suppliers. *Strategic Management Journal*, *30*, 117–135.

Ketokivi, M., & Mahoney, J. T. (2020). Transaction cost economics as a theory of supply chain efficiency. *Production and Operations Management*, *29*, 1011–1031.

Klein, B., Crawford, R. G., & Alchian, A. (1978). Vertical integration, appropriable rents, and the competitive contracting process. *Journal of Law and Economics*, *21*, 297–326.

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

Liker, J. K., & Choi, T. Y. (2004). Building deep supplier relationships. *Harvard Business Review*, *82*(12), 104–113.

Liu, Y., Luo, Y., & Liu, T. (2009). Governing buyer–supplier relationships through transactional and relational mechanisms: Evidence from China. *Journal of Operations Management*, *27*(4), 294–309.

Lumineau, F., & Oliveira, N. (2020). Reinvigorating the study of opportunism in supply chain management. *Journal of Supply Chain Management*, *56*, 73–87.

MacDuffie, J. P. (1997). The road to "root cause": Shop-floor problem-solving at three auto assembly plants. *Management Science*, *43*(4), 479–502.

MacDuffie, J. P., & Helper, S. (2006). Collaboration in supply chains with and without trust. In C. Heckscher & A. Paul (Eds.), *The firm as collaborative community, re-constructing trust in the knowledge economy*. Oxford, England: Oxford University Press.

Mahapatra, S. K., Narasimhan, R., & Barbieri, P. (2010). Strategic interdependence, governance effectiveness and supplier performance: A dyadic case study investigation and theory development. *Journal of Operations Management*, *28*(6), 537–552.

Mahoney, J. T., & Qian, L. (2013). Market frictions as building blocks of an organizational economics approach to strategic management. *Strategic Management Journal*, *34*(9), 1019–1041.

Maloni, M., & Benton, W. C. (2000). Power influences in the supply chain. *Journal of Business Logistics*, *21*(1), 49–74.

March, J. G. (2010). *The ambiguities of experience*. Ithaca, NY: Cornell University Press.

Mason, C. F. (2018). *Analyzing the risk of transporting crude oil by rail*, Cambridge, MA: Working paper 24299. National Bureau of Economic Research.

McCarter, M. W., & Northcraft, G. B. (2007). Happy together? Insights and implications of viewing managed supply chains as a social dilemma. *Journal of Operations Management*, *25*(2), 498–511.

Mellewigt, T., Hoetker, G., & Lütkewitte, M. (2018). Avoiding high opportunism is easy, achieving low opportunism is not: A QCA study on curbing opportunism in buyer–supplier relationships. *Organization Science*, *29*(6), 1208–1228.

Montgomery, D. C. (2005). *Design and analysis of experiments* (6th ed.). New York, NY: Wiley.

Novak, S., & Eppinger, S. (2001). Sourcing by design: Product complexity and the supply chain. *Management Science*, *47*(1), 189–204.

Novak, S., & Wernerfelt, B. (2012). On the grouping of tasks into firms: Make-or-buy with interdependent parts. *Journal of Economics & Management Strategy*, *21*(1), 53–77.

Nyaga, G. N., Whipple, J. M., & Lynch, D. F. (2010). Examining supply chain relationships: Do buyer and supplier perspectives on collaborative relationships differ? *Journal of Operations Management*, *28*(2), 101–114.

Peerally, M. F., Carr, S., Waring, J., & Dixon-Woods, M. (2017). The problem with root cause analysis. *BMJ Quality & Safety*, *26*, 417–422.

Perrow, C. (1984). *Normal accidents*. Princeton, NJ: Princeton University Press.

Plambeck, E. L., & Taylor, T. A. (2005). Sell the plant? The impact of contract manufacturing on innovation, capacity, and profitability. *Management Science*, *51*(1), 133–150.

Poppo, L., Zhou, K. Z., & Zenger, T. R. (2008). Examining the conditional limits of relational governance: Specialized assets, performance ambiguity, and long-standing ties. *Journal of Management Studies*, *45*(7), 1195–1216.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-disciplinary view of trust. *Academy of Management Review*, *23*(3), 393–404.

Santos, F. M., & Eisenhardt, K. M. (2005). Organizational boundaries and theories of organization. *Organization Science*, *16*(5), 491–508.

Simon, H. A. (1997). *Administrative behavior* (4th ed.). New York, NY: Macmillan.

Skilton, P. F., & Robinson, J. L. (2009). Traceability and normal accident theory: How does supply network complexity influence the traceability of adverse events? *Journal of Supply Chain Management*, *45*(3), 40–53.

Sosa, M. E., Eppinger, S. D., & Rowles, C. M. (2004). The misalignment of product architecture and organizational structure in complex product development. *Management Science*, *50*(12), 1674–1689.

Starbuck, W., & Farjoun, M. (Eds.). (2005). *Organization at the limit: Lessons from the Columbia disaster*. Oxford, England: Blackwell Publishing.

Tang, C. S. (2008). Making products safe: Process and challenges. *International Commerce Review*, *8*(1), 48–55.

Terpend, R., Tyler, B. B., Krause, D. R., & Handfield, R. B. (2008). Buyer–supplier relationships: Derived value over two decades. *Journal of Supply Chain Management*, *44*(2), 28–55.

Tetlock, P. E. (2005). *Expert political judgment*. Princeton, NJ: Princeton University Press.

Transportation Safety Board of Canada. (2013). *Railway investigation report R13D0054: Runaway and main-track derailment, freight train MMA-002*. Gatineau, Canada: The Transportation Safety Board of Canada.

Tsay, A. A., Gray, J. V., Noh, I. J., & Mahoney, J. T. (2018). A review of production and operations management research on outsourcing in supply chains: Implications for the theory of the firm. *Production and Operations Management*, *27*(7), 1177–1220.

Tucker, A. L., & Edmondson, A. C. (2003). Why hospitals don't learn from failures: Organizational and psychological dynamics that inhibit system change. *California Management Review*, *45*, 55–72.

Ulrich, K. T., & Ellison, D. J. (2005). Beyond make-buy: Internalization and integration of design and production. *Production and Operations Management*, *14*(3), 315–330.

Villena, V. H., & Craighead, C. W. (2017). On the same page? How asymmetric buyer–supplier relationships affect opportunism and performance. *Production and Operations Management*, *26*(3), 491–508.

Walker, G. (2007). The incompleteness of research on firm boundaries. *Strategic Organization*, *5*(3), 227–236.

Wang, Y., Niu, B., & Guo, P. (2013). On the advantage of quantity leadership when outsourcing production to a competitive contract manufacturer. *Production and Operations Management*, *22*(1), 104–119.

Weick, K. E. (2004). Normal accident theory as frame, link, and provocation. *Organization & Environment*, *17*(1), 27–31.

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

Williamson, O. E. (1979). Transaction-cost economics: The governance of contractual relations. *Journal of Law and Economics*, *22*, 233–261.

Williamson, O. E. (1983). Credible commitments: Using hostages to support exchange. *American Economic Review*, *73*(4), 519–540.
Williamson, O. E. (1985). *The economic institutions of capitalism*. New York, NY: Free Press.
Williamson, O. E. (1991). Strategizing, economizing, and economic organization. *Strategic Management Journal*, *12*(Winter Special Issue), 75–94.
Williamson, O. E. (1994). Transaction cost economics and organization theory. In N. J. Smelser & R. Swedberg (Eds.), *The handbook of economic sociology* (pp. 77–107). Princeton, NJ: Princeton University Press.
Williamson, O. E. (1996). *The mechanisms of governance*. New York, NY: Oxford University Press.
Williamson, O. E. (2008). Outsourcing: Transaction cost economics and supply chain management. *Journal of Supply Chain Management*, *44*(2), 5–16.
Wu, A. W., Lipshutz, A. K. M., & Pronovost, P. J. (2008). Effectiveness and efficiency of root cause analysis in medicine. *JAMA*, *299*(6), 685–687.
Zaheer, A., McEvily, B., & Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science*, *9*(2), 141–159.
Zhang, C., Viswanathan, S., & Henke, J. W., Jr. (2011). The boundary spanning capabilities of purchasing agents in buyer–supplier trust development. *Journal of Operations Management*, *29*(4), 318–328.
Zu, X., & Kaynak, H. (2012). An agency theory perspective on supply chain quality management. *International Journal of Operations & Production Management*, *32*(4), 423–446.

How to cite this article: Bhardwaj A, Ketokivi M. Bilateral dependency and supplier performance ambiguity in supply chain contracting: Evidence from the railroad industry. *J Oper Manag*. 2020; 1–22. https://doi.org/10.1002/joom.1103

APPENDIX A.

A summary of observed failure modes and frequencies is given in Table A1.

TABLE A1 Failure modes and frequencies

Top deck	Cylinder head	Piston	Cylinder liner	Engine block	Number of failures	Frequency	Percent	Cum. percent
1	1	0	0	0	Multiple (2)	41	23.8	23.8
0	0	0	0	1	Single	25	14.5	38.3
1	0	1	0	0	Multiple (2)	14	8.1	46.5
1	0	0	0	1	Multiple (2)	13	7.6	54.0
0	0	1	0	0	Single	12	7.0	61.0
0	0	1	1	0	Multiple (2)	11	6.4	67.4
0	1	0	0	0	Single	10	5.8	73.2
0	0	0	1	0	Single	9	5.2	78.5
1	0	1	1	0	Multiple (3)	9	5.2	83.7
1	1	1	1	0	Multiple (4)	8	4.7	88.3
1	1	1	0	0	Multiple (3)	5	2.9	91.2
1	1	1	1	1	Multiple (5)	3	1.7	93.0
0	1	1	0	0	Multiple (2)	2	1.2	94.1
1	0	0	0	0	Single	2	1.2	95.3
1	0	1	1	1	Multiple (4)	2	1.2	96.5
0	0	0	1	1	Multiple (2)	1	0.6	97.1
0	1	0	1	0	Multiple (2)	1	0.6	97.6
0	1	1	1	0	Multiple (3)	1	0.6	98.2
1	0	0	1	1	Multiple (3)	1	0.6	98.8
1	0	1	0	1	Multiple (3)	1	0.6	99.4
1	1	1	0	1	Multiple (4)	1	0.6	100.0
				Total	Single	58		33.7
					Multiple	114		66.3
				Total		172		100.0

APPENDIX B. CONFOUNDING OF EFFECTS AS BOUNDED RATIONALITY AND SUPPLIER PERFORMANCE AMBIGUITY

In our analyses, we discovered that the effects of the failure modes on outcome variables such as time to failure are confounded with one another. Here, the notion of *confounding* is used in the same technical meaning as in an experimental design (Montgomery, 2005, chap. 7). In an experiment, a central principle is to manipulate the factor levels such that the treatment effects of interest are not confounded with one another. For instance, in a 2^5 factorial experiment where five factors are varied at two levels, we would not want the main effects to be confounded with one another. We would be less concerned with main effects being confounded with four- or five-way interactions, because higher-order interactions are often negligible.

Although in our study nothing is manipulated, we can think of the failure modes analogously with treatment effects: We can think of the different failures modes as "treatments" of sorts, and the 21 observed failure modes as "treatment combinations." One question of interest would be to see whether we can predict time to failure as a function of these "treatment combinations." In an attempt to do this, we specified a Cox regression model with the time to failure as the dependent variable and the main effects and the interactions of the "treatment" variables as independents. Due to confounding, the estimation algorithm ran into serious difficulty in trying to estimate the model.

In order to demonstrate the confounding, we thought of the data as a 2^5 factorial design and created by hand the *contrasts* for each main effect and interaction. We then calculated all bivariate correlations among these contrasts in our data. Predictably enough, these correlations were not only nonzero, but also many of them were rather strong, indicating strong confounding. Figure B1 summarizes these correlations.

Let us take the high correlation between TD*P and TD*CH*EB as an example. This correlation is +.67. This correlation means that in the experimental design terminology, the two effects are confounded in that the two-way interaction of the top deck (TD) and the piston (P) is empirical indistinguishable from the three-way interaction of the top deck (TD), the cylinder head (CH), and the engine block (EB). As Figure B1 shows, there not only are dozens of strong correlations among the contrasts, but also that some of these are between the main effects and the two-way interactions. We interpret this as another manifestation of performance ambiguity: Based on these data, it is simply impossible to determine *ex post* which factors contributed to time to failure. The only way to get around the problem would be to try to observe

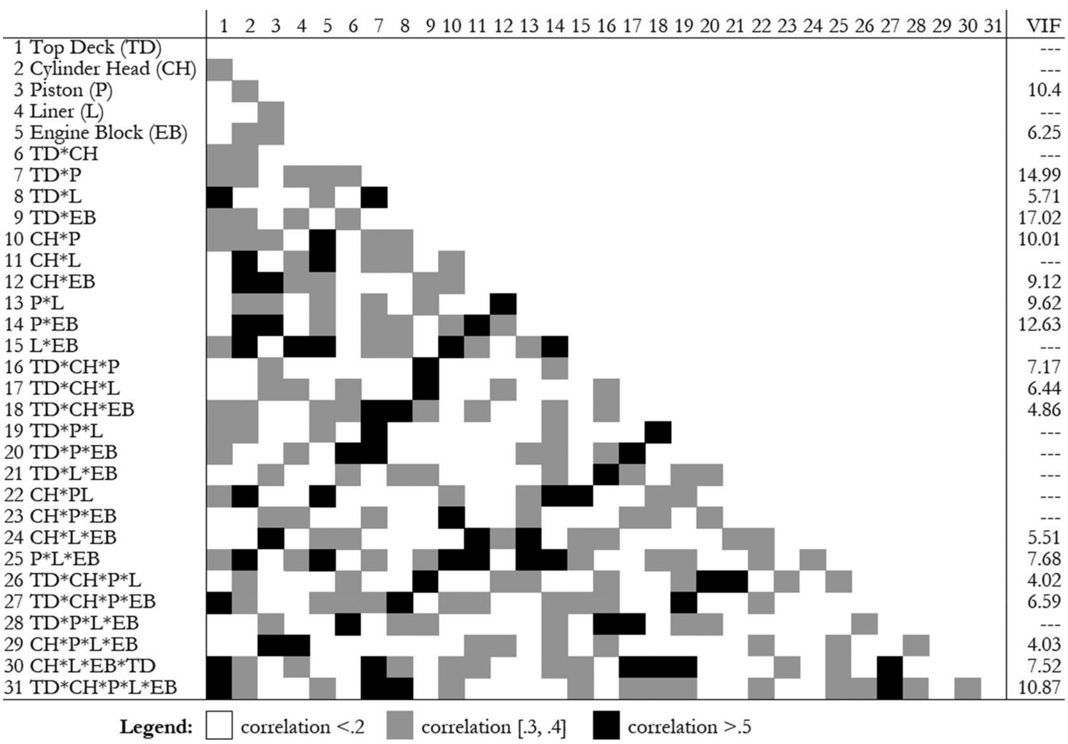

FIGURE B1 Confounding of the failure mode effects the failures as the occur. As we mention in the discussion section, advanced sensor technology can make that possible in the future.

The last column of Figure B1 gives the variance inflation factors (VIFs) from a regression analysis where time to failure is specified as the dependent variable in the 2^5 full-factorial ANOVA. The VIFs are generally very high, which is another symptom of estimation difficulty. Also, in 12 out of the 31 cases (the ones indicated by"---"), Stata excluded the variable in question from the analysis because collinearity was so high that estimation of the effect was impossible.