DATA MODELS AS ORGANIZATIONAL DESIGN: COORDINATING BEYOND BOUNDARIES USING ARTIFICIAL INTELLIGENCE

Tom Steinberger
tsteinbe@uci.edu

Margarethe Wiersema
mfwierse@uci.edu

University of California, Irvine

(forthcoming, *Strategic Management Review*)

DATA MODELS AS ORGANIZATIONAL DESIGN: COORDINATING BEYOND BOUNDARIES USING ARTIFICIAL INTELLIGENCE

Abstract

Organizational design scholars observe that advances in information technology are helping blur the boundaries between a firm's internal and external activities. Yet despite the central role of information processing in managing the coordination of activities within the firm, we know little about the firm's ability to process information beyond its boundaries. We provide a framework for understanding the coordination of activities beyond the firm's boundaries in terms of micro-structural solutions to information provision. Our core insight is that organizational design can be modeled at the level of data. The firm's 'data model' shapes processes of data integration using artificial intelligence, enabling agents to frame and find their problem contexts and self-organize activities. We contribute to the organizational design and strategy literatures by showing how coordination beyond boundaries has major, yet neglected, micro-structural effects on how firms organize. We discuss research implications for managerial capabilities, corporate strategy amid digitalization, and models of strategic representations.

DATA MODELS AS ORGANIZATIONAL DESIGN: COORDINATING BEYOND BOUNDARIES USING ARTIFICIAL INTELLIGENCE

INTRODUCTION

The scope choices underlying a firm's activities have traditionally concerned make-or-buy decisions within an industry value chain. Boundaries are clearly demarcated. The firm coordinates activities performed in-house by internal employees and divisions, while those performed by firms or agents[1] externally are handled through arms-length contracts such as supplier relationships, alliances, or partnerships. Organizational design scholars, among others, have observed that advances in information technology have helped blur the boundaries between internal and external activities (Benner and Tushman, 2015; Joseph, Baumann, Burton and Srikanth, 2018). Improved information processing capacity (e.g., 4g networks, cloud services, faster processors, smartphones) has enhanced coordination capabilities, reduced transaction costs, and made possible novel value propositions and business models (Helfat and Raubitschek, 2018; Teece, 2017; Fjeldstad and Snow, 2018). Scope choices increasingly concern a firm's ability to harness this improved information processing capacity to shape how agents self-organize their activities (Gulati, Puranam and Tushman, 2012; Fjeldstad, Snow, Miles and Lettl, 2012).

In comparison to the conventional industry value chain, firms able to harness improved information processing capacity to coordinate activities beyond boundaries can realize benefits of economies of scale and scope, leverage complementarities in resources, and develop more flexible capabilities for innovation and search (Thomas, Autio and Gann, 2014; Jacobides, Cennamo and Gawer, 2018). Yet while information processing capacity has been essential to the internal coordination of the firm (e.g., Tushman and Nadler, 1978), extant perspectives on coordinating beyond the firm's boundaries have tended to subordinate analysis of information processing. Instead the focus has been on relatively macro-level components of platforms or rules of collaboration in ecosystems, marketplaces or communities (e.g., Gawer, 2014; Afuah and Tucci, 2012). As a result, while we have rich insights into relatively macro-level components and rules, micro-structural mechanisms (Puranam, 2018) by which firms develop the capacity to process information beyond their boundaries remain little explicated.

[1] We adopt the broad sense of the term agent suggested by Puranam (2018: 6) as 'any entity capable of action'.

The importance of how firms develop information processing capacity beyond their boundaries is highlighted by the emergence of digital data as a central strategic resource. Vastly expanded ability to acquire and analyze a variety[2] of digital data has made integrating this data to provide information among the most complex and visible issues that firms face. The very existence of digital platforms (e.g., Google, Facebook, DropBox) is premised on scalable capabilities for harnessing diverse data to provide context-specific information to users. Ride sharing services such as Uber or Lyft depend on users' ability to interact with an evolving database of information about local drivers and passengers. Across typical enterprises, a virtually universal strategic goal has become the integration of silos of data into centralized data warehouses for more effective use of information in their activities. GE, for instance, estimates it saved $80 million per year in local managers' everyday negotiations with suppliers simply by integrating procurement systems across its divisions and subsidiaries (Davenport and Ronanki, 2018).

Harnessing digital data effectively can thus enhance the activities that the firm coordinates by augmenting agents' capacity to process information. The fact that this data is generated from multiple sources for use in diverse problem contexts[3], however, has led to data integration being viewed as the '800 pound gorilla in the corner' (Stonebraker, 2015: 2). To effectively integrate numerous silos of data, the firm must not only acquire and classify the data, but also determine how to make it available so that agents are provided with the requisite information they need in their activities. Given that the firm cannot sufficiently anticipate the problem contexts of its agents' activities, it needs to integrate data in such a way as to flexibly provide information for agents to frame and find problems on their own (Nickerson, Wuebker and Zenger, 2017). Explicating mechanisms by which firms integrate data effectively and flexibly is therefore important to understanding how firms coordinate activities beyond their boundaries.

Artificial intelligence (AI) has emerged as a key tool for enabling data integration beyond boundaries. Firms use AI (i.e., deep learning algorithms, logic programs, sensor networks) to filter, classify and make predictions from their data. To the extent that a firm is able to use AI to process data regarding agents' activities, the firm can determine which data can be used to provide information specific to agents' problem contexts. Determining which information to provide to

[2] We emphasize the increase in the *variety* of data that firms collect and analyze, rather than increases in the amount (*volume*) or speed (*velocity*) of data (Stonebraker and Ilyas, 2018). Data volume and velocity pose primarily technical challenges; data variety poses both technical and organizational challenges that arise even at low volume and velocity.

[3] By problem context, we refer to 'challenges, opportunities, situations [and]… alternative possible future states' (Nickerson and Argyres, 2018: 592) underlying activities. Nickerson and Argyres (2018) use this definition to characterize strategy formulation. We argue that analogous processes characterize agents who self-organize their activities.

agents is known within the organizational design literature as the problem of information provision. Solutions to information provision — or how agents get enough information to execute their activities and coordinate with others (Puranam, Alexy, and Reitzig, 2014) — are viewed as universal elements of a firm's organizational design for integrating the efforts of agents (Schelling, 1960; Lawrence and Lorsch, 1967; Puranam, 2018). Like firms' use of other information technologies (IT), then, how firms integrate data using AI has implications for organizational design strategy.

In this paper, we develop a framework to understand how firms integrate data using AI as a solution to information provision beyond boundaries. First, we describe how organizational design based on the division of labor constrains the ability of the firm to coordinate activities beyond its boundaries. Then, we lay out our framework by developing the idea of a firm's data model as organizational design based on information provision. By data model, we refer generally to any 'collection of high-level data description constructs' that can be easily accessed and manipulated by users (Ramakrishnan and Gehrke, 2000: 9). After laying out our framework, we give some empirical context through two short cases analyzing the role of data integration in coordinating activities beyond firm boundaries at Novartis and Airbnb. We then identify a firm's strategic decisions for developing its data model, which involve identifying archetypal problems of information provision, representing agents' problem contexts, and assigning credit for data integration efforts.

Our core insight is that organizational design can be modeled at the level of data. Decisions regarding the firm's data and how it integrates its data could be viewed as more technical than strategic in nature, and thus more in the domain of database management or software engineering. We believe that such view would be critically shortsighted. Consider the core process in strategy formulation of theorizing which activities the firm should engage in to create value (Felin and Zenger, 2017). For the firm to provide information beyond boundaries, it needs to in some way represent its activities within some sort of a database. Keeping the firm's database logically consistent yet adaptable to self-organizing agents' diverse, evolving problem contexts depends on understanding its theory of value creation, the information that agents need about activities, how these agents wish to access and manipulate this information, and the broad technical implications for the firm's database design. The firm's data decisions are therefore strategic, in that they clearly call for a ‘capacity to imagine and model complex interactions with both internal and external actors’ (Leiblein, Reuer and Zenger, 2017: 559).

We draw on our core insight that a firm’s organizational design can be modeled at the level of data to contribute to the strategy and organizational design literatures. Our framework shows how the coordination of activities beyond the firm’s boundaries has major, yet neglected, micro-structural effects on how the firm organizes that relate to its solutions to information provision. Extant research on firm boundaries does not fully account for these effects due to its tendency to focus on problems of division of labor. We discuss how the need to integrate data to coordinate beyond boundaries reveals the need for distinct managerial capabilities for data abstraction, and which call for a more micro-structural understanding of a firm’s architectural knowledge. We further discuss implications for literature on corporate strategy amid digitalization, showing how a firm’s data model can offer important insight into the nature of a firm’s activities beyond boundaries that are not captured by the more macro-structural approaches that are currently predominant. Finally, we lay out future directions for how a firm’s data model can be used as a basic organizational design variable in the nascent literature on firms’ strategic representations.

INFORMATION PROVISION BEYOND BOUNDARIES

According to Puranam *et al*. (2014:165), the design of a functioning organization must solve ‘two fundamental and interlinked problems: the division of labor and the integration of effort’. The division of labor consists of task division and task allocation, while the integration of effort consists of rewards provision and information provision.

Much research finds that a firm can coordinate effectively with an organizational design that is based on solutions to division of labor. Examples include the use of modular task structures for innovation, and the adoption of organizational hierarchies, networks or polyarchies according to the firm’s technological and competitive environments. Using solutions to division of labor to guide solutions to information provision (e.g., as in conventional enterprise resource planning (ERP) software), however, can impose too much structure for information to be adaptable agents’ diverse, evolving problem contexts (Kallinikos, 2004). Effective, flexible solutions to information provision instead tend to be characterized by being largely ‘pure’[4] — or independent from — structuring

[4] We adopt the term ‘pure’ from the computing and artificial intelligence literatures, where the analogous term for information provision-related processes is ‘messaging’. ‘Pure’ messaging to refer to the ability of an agent (human or artificial) to provide information without pre-determined communication channels (Hewitt, 2014).

based on how the firm divides and allocate tasks. As a simple example to contrast with ERP, a firm's email system can be viewed as 'pure' information provision in that an agent can send or receive messages to any agent for whom they have the email address, regardless of the tasks or roles involved. We describe the basic distinction we make between organizational design based on solutions to division of labor versus information provision in Table 1 below.

TABLE 1:
Organizational Design based on Division of Labor or Information Provision

Organizational design based on division of labor: Align firm's division of labor with technological and competitive environments

Division of Labor →	Information Provision
Modular task structures, organizational hierarchies, networks, polyarchies, etc.	Representations are guided by division of labor
	Example: ERP system that channels information according to how tasks are divided and allocated

Organizational design based on information provision strategy: Align firm's representations with agents' problem contexts

Division of Labor	← Information Provision
Self-organizing agents	Representations are 'pure' of solutions to division of labor
	Example: Email system that allows sending messages to any agent with an email address regardless of how tasks are divided and allocated

'Pure' solutions to information provision are broadly consistent with Carnegie School-inspired theories in organizational design in which coordination is based on shaping agents' adaptive search processes, rather than specifying task division or allocation directly (e.g., Levinthal and Warglien, 1999). While organizational design research using these theories has largely taken a behavioral approach, recent work on strategy process from the same tradition identifies managers' *representations* as mechanisms for modeling their search environments (Csaszar and Levinthal, 2016; Puranam and Swamy, 2016; Csaszar, 2018). By creating, modifying and manipulating representations, managers can augment their use of judgment, theorizing, and analogizing to shape the very nature of the firm's activities (Leonardi and Bailey, 2008; Foss and Klein, 2012; Helfat and Peteraf, 2015; Nickerson et al., 2017; Gavetti, Helfat and Marengo, 2017). We argue that solutions to information provision beyond boundaries likewise can be viewed in terms of how the firms' agents as a whole use representations. To the extent that the firm can augment such use of representations across all of its agents, it can enhance these agents' ability to frame and find their problem contexts and self-organize activities.

Next, we develop our framework for understanding solutions to information provision Beyond firm boundaries. We first introduce firms' use of AI as a key tool in developing such solutions amid firms' vastly expanded ability to access and analyze digital data. After defining AI as used by firms, we show how it plays a role in information provision beyond boundaries by helping the firm integrate pervasively semi-structured data regarding agents' activities. We then develop a definition of a firm's *data model* as an organizational design variable by which the firm can guide data integration processes.

Integrating Semi-Structured Data Using AI

Practitioners and scholars nowadays tend to consider a technology as AI not by some objective measure of intelligence, but merely in terms of whether it can be plausibly described as rational in the sense of acting on the basis of some set of beliefs (Agrawal, Gans and Goldfarb, 2018). Given complex environments, rationality is assumed to be limited and thus based on an AI tool's ability to respond *adaptively* (Gershman, Horvitz, Tenenbaum, 2015). AI tools (e.g. deep learning algorithms, logic programs, sensor networks) that fall under these criteria have become widely viewed as strategic resources. According to a survey, 85% of executives from 3,000 firms across diverse industries believe AI will help them 'obtain or sustain a competitive advantage' (Ransbotham, Kiron, Gerbert and Reeves, 2017:1). Examples of the potential value of AI for firms can be easily found. Pattern recognition is used in healthcare to classify medical images from diverse patients. In the automotive industry, visual and lidar sensors and GPS navigation enable semi-autonomous vehicles to react to road environments. Facebook's predictive and filtering algorithms display social media content according to user behavior.

While firms make use of diverse AI tools, the basic functions of these tools concern the adaptive processing of diverse, evolving data. We thus consider AI in its use by firms as machines with the ability to adaptively process diverse, evolving data regarding the firm's complex organizational and competitive environments.

An AI tool's ability to respond adaptively develops through trial-and-error, and thus depends on the data on which it is trained. AI can be trained on 'unstructured data' that has little prior formatting or given purpose by the firm (e.g., images, audio clips or text scraped randomly from the web), as well as 'structured data' already organized into a firm's database (e.g., transaction records, customer data, or user behavior on websites). Firms' use of AI in solutions to information provision, however, is driven by the need to deal with *semi-structured* data. By semi-structured, we refer to data that is stored, but not well-integrated into a firm's database. Either certain attributes of data have not been sufficiently defined or related to other data to be of value for analysis by the firm, or relations between attributes of the data are inconsistent across different data sources. The data thus sits somewhere between silos of data, and integrated 'data warehouses'. Semi-structured data are important as they characterize the pervasively inconsistent, messy reality of the vast majority of data used by the firm for performing its ordinary activities (Hewitt, 2014).

Data regarding a firm's activities may be semi-structured simply since the firm does not have the knowledge or resources to interpret it in much detail. Lack of structuring more basically results from the fact that a firm's databases are created at different times, by different people, and are often managed and accessed independently by particular users. Most of the firm's data does not even reside in database software — the firm may represent budget numbers on a spreadsheet, reports on word processors and slides, and emails on an internet app (Stonebraker and Ilyas, 2018). These are merely administrative obstacles; at the level of practice, data is inherently semi-structured in that relevance is specific to an agent's problem context (Carlile, 2006; Leonard and Bailey, 2008). Such specificity poses interpretive challenges in integrating data, given insufficient understanding of these problem contexts as well as idiosyncrasies in how agents represent data. Variation in agents' roles and preferences further leads to inconsistency in how data is evaluated (March and Simon, 1958; Hewitt, 2014). Integrating data thus depends on ongoing processes of elaboration and evaluation. As a quintessential example, major initiatives to integrate electronic medical records to improve administrative efficiency, quality control and medical outcomes, have faced challenges from the idiosyncratic nature by which data for these records is generated and used across diverse facilities, hospital departments, and individual healthcare professionals.

The firm of course has diverse options for guiding information provision within this messy reality of semi-structured data. Most basically, it can decide that it need structure only a small part of this data, with the rest of little consequence to value creation and capture. Alternatively, the firm may judge that semi-structured data is reflective more of how its agents' problem contexts are bound up in individuals' tacit knowledge. In such case, data may be left in isolated silos, with information provision instead shaped by organizational mechanisms to structure or enrich communication channels, such as standard operating procedures, manuals, social media tools, or online knowledge sharing communities.

Given the growing importance of data as a resource for value creation and capture, however, firms increasingly seek to develop dedicated capabilities for data integration to provide information according to agents' problem contexts. We next introduce the idea of a data model as the core solution to information provision that the firm develops to guide data integration processes.

Data Models: Representations of the Firm's Theories of its Activities

We have laid out how information provision beyond a firm's boundaries depends on ongoing processes of integrating semi-structured data using AI. We next bring in our main theoretical assumption, which is that a firm's *model* for simplifying its understanding of the messy reality of its data drives the effectiveness and flexibility of its solutions to information provision.

Under the Carnegie School tradition from which the micro-structural approach to organizational design derives, the firm's strategy is based on its limitedly rational *model* of its search environment (Lave and March, 1993). Models correspond to 'conceptual structures… that encapsulate a simplified understanding of… reality' (Puranam, 2018: 38). The firm's model of its search environment has often been viewed in terms of goals and constraints that guide the solving of given problems within given (if uncertain and ambiguous) competitive and technological environments (March, 2006). Increasingly, however, scholars emphasize how a firm's use and generation of *representations* of its search environment enable the firm not just to solve given problems, but to shape the very search environment itself in which problems are identified (Csaszar, 2018; Nickerson and Argyres, 2018). The firm's representations of its search environment can, in one sense, be viewed as its theories of value creation (Felin and Zenger, 2017). That is, such representations correspond to 'theories and hypotheses about which activities [firms] should engage in, which assets they should buy and how they create value' (Felin and Zenger, 2017: 258). In regards to information provision, we likewise situate a firm's model within such 'theory for the firm' perspective. Given that the firm's solutions to information provision for coordinating activities relate to ongoing processes of data integration, we refer to such solutions as a firm's 'data model'. More specifically, we define a firm's data model as representations of theories regarding agents' activities that are implementable in a database for use by its agents and its AI.

How firms represent their activities in databases has traditionally concerned the study of management information systems (MIS) more than organizational design. This makes sense under a view of the firm as based on internal coordination, where the design of an MIS presupposes detailed representations of a firm's division of labor (e.g., organizational charts, standard operating procedures) that already embed theories and hypotheses of a firm's activities. To the extent that agents' activities are self-organized, however, the firm by definition must represent these activities independent of solutions to division of labor. It follows that the firm's representations of its theories of agents' activities correspond to a conceptual structure implementable in a database that do not require the database designer to necessarily account for the firm's division of labor. We next draw on the field of database design to identify basic criteria regarding the effectiveness and flexibility of a firm's data model.

Effectiveness and flexibility of data models

In the field of database design (e.g., Ramakrishnan and Gehrke, 2000), data models as conceptual structures correspond to representations of databases that abstract away technical detail in order to be accessibly defined or manipulated by users. It is possible to design such conceptual data models as *hierarchies* or *networks* of dimensions, just as in classic solutions to division of labor based on structuring organizations and tasks into hierarchies or networks. Database designers, however, overwhelmingly use much simpler *logical relations* that tend to allow far more effective and flexible information provision than hierarchies or networks (Codd, 1970; Stonebraker and Hellerstein, 2015).

On this point, we draw an important link between database design and the theory-based views of the firm mentioned above. Theory-based views of the firm analogize how a firm logically links its activities and assets to value creation to the logical relations of a scientific theory (Felin and Zenger, 2017). Considered as a conceptual structure composed of logical relations about a firm's activities, a firm's data model can likewise be analogized to a scientific theory. That is, the firm's data model can be analyzed in terms of the logical consistency, simplicity, generalizability and generativity[5] by which it relates representations of agents' activities to value creation based on information provision to these agents. The usefulness of viewing a firm's data model as theories of activities is implicit in the firm's need to coordinate ongoing processes of integrating semi-structured data. That is, given pervasive inconsistencies in its data, the firm would like to develop simple solutions to information provision that assume as little about this data in advance, while still being generalizable and generative across agents' problem contexts. We propose, then, that the effectiveness and flexibility of information provision depend on the extent to which a firm's data model is logically consistent, simple, generalizable and generative.

In Table 2 below, we situate these criteria for and our definition of data models within theory-based views of the firm, and the broader use of the term models in the strategy literature. We then flesh out our definition to sketch broad variation in data models, before providing some empirical context based on two short cases of firm's data integration processes beyond boundaries.

TABLE 2:
Data Models as Organizational Design

Variable	Definition	Purpose	Key criteria
Models of firm strategy	Conceptual structures of a firm's search environment (Lave and March, 1993)	Developing mechanisms for adapting firm behavior	Simplified understanding of reality (Puranam, 2018)
Models as theories for the firm	Logical relations of activities and assets to value creation (Felin and Zenger, 2017)	Developing a novel, valuable strategy	Novelty, simplicity, elegance, falsifiability, generalizability, generativity
Data models within a 'theory for the firm'	Logical relations among activities that can be represented in a database	Developing a novel, valuable strategy for information provision	Logical consistency (Codd, 1970), simplicity, generalizability, generativity

Variation in data models

Unfortunately, research and practice indicate that the messy reality of a firm's data makes

[5] Felin and Zenger (2017: 265) focus on 'valuable economic theories', which they identify as theories with 'novelty, simplicity and elegance, falsifiability, and generalizability and generativity'. We narrow these criteria to simplicity, generalizability and generativity as directly relevant to information provision. Further, given pervasive inconsistencies in semi-structured data creates inherent challenges to making a 'good theory' regarding information provision, we add 'logical consistency' as a criterion.

it impossible to have a single data model serve as a logically consistent ‘global schema’ of the activities of a firm, even if the model is ‘pure’ from assumptions about the division of labor (Stonebraker and Ilyas, 2018). A firm’s data model comprises not a unified theory for representing its agents’ activities, but rather a kludge of internally consistent data models that are mutually inconsistent (Hewitt, 2014). To develop our framework on data integration (and as a basis for future work), we thus limit our analysis to only a sketch of broad variations in data models.

At one extreme, a firm’s data model may meet all the criteria of a good theory, but offer little formal support for data integration processes. To draw on an earlier example, a firm’s email system may structure data only by email address (i.e., an agent can email any other agent, regardless of how tasks or roles are structured). While agents’ use of email may be critical to how a firm provides information, it is unlikely to have firm specificity or explicitly link to an organizational design for coordinating a firm’s data integration processes. On the other side of things, a firm’s MIS may be central to information provision, but only by mediating a firm’s decisions regarding its division of labor. We assume that few firms compete based on providing information using superior enterprise software. To give an example closer to what we mean, consider how Google initially developed capabilities for web search by integrating data regarding its users’ search activities. Many of the incumbents’ web search capabilities were in contrast tied to internal activities of manually classifying the content of web pages. What we might call Google’s distinct theories regarding its activities in web search (user-driven versus internally driven) were simply represented in a database (a network model based on clicks on web links) that enabled effective, flexible ongoing data integration processes using AI (its predictive algorithms for analyzing its data). By identifying a simple theory of value creation based on users’ web search activities, Google’s data model remained logically consistent regardless of changes in content and scale, and thus was both generalizable and generative.

To serve as an organizational design variable that involves strategic decisions, we thus assume that a firm’s data model should correspond to representations of valuable theories regarding agents’ activities that support data integration processes for the purposes of information provision. We next give some empirical context for how a firm’s data model can enable understanding firms’ processes of data integration for providing information beyond boundaries. We develop two short cases regarding Novartis and Airbnb, highlighting their common solutions to information provision despite quite different value propositions.

DATA INTEGRATION AT NOVARTIS AND IRBNB

Drug discovery at Novartis

Value creation at the Swiss pharmaceuticals conglomerate Novartis is driven by its innovation strategy, which is based on developing capabilities for coordinating drug discovery processes. In 2017, it spent nearly $10 billion on R&D to develop such capabilities. Novartis has opted to organize its drug discovery processes with highly decentralized research operations, little bureaucracy, and an external scientific advisory board as a primary formal coordination mechanism. It has around 10,000 lab scientists distributed globally, sources knowledge through hundreds of collaborations with research foundations, universities and ventures, and has made numerous multi-billion dollar acquisitions in recent years.

Challenges of data integration arise from the diversity of lab scientists and external sources, which leads to pervasive inconsistencies in how data regarding drug trials is stored and interpreted. For example, its R&D efforts and decentralized approach are reflected in a highly heterogeneous distribution of patents (Garg and Zhao, 2018), and databases that span millions of patient years based on trials run in parallel across over 60 countries. While its masses of data have prompted a major initiative to develop centralized analytics capabilities using machine learning, its global head of drug development stresses, "We want to focus on what we do well, [which is] running trials… We are not a tech shop."[6] Novartis views the core data relevant to drug discovery as generated, represented and used according to the needs of individual labs, where researchers have wide latitude to carry out experiments. While the lab scientists who are the main users of the data are typically employees of the firm, they effectively self-organize their experiments. Thus, Novartis' innovation strategy is premised on effective, flexible information provision to coordinate activities beyond traditional organizational boundaries.

A primary initiative in Novartis's solutions to information provision regarding agents' decentralized drug trials is to integrate all of its scientists' data in a unified database. The scalability of such a database is made possible through use of machine learning, which partially mechanizes the processes of classifying data generated by diverse lab scientists or other sources (Stonebraker and Ilyas, 2018). The idea is to allow its scientists to query a firm-level database in any way that might add useful insights to the design or interpretation of their own experiments. It desires to

[6] https://www.informationweek.com/big-data/big-data-analytics/novartis-seeks-hidden-cures-in-machine-learning-ai/d/d-id/1332269

understand how its scientists use and generate representations within these drug trial activities to enable them to 'ask the questions they weren't able to ask before'. Given that its agents are largely free to choose which drug trials to run and how to run them, it has developed what is essentially a scientific database. The content of and tools for querying its database can be related to a data model for 'pure' information provision, in that they require little or no specifications of how tasks in drug trials are divided and allocated. Novartis prioritizes merely a small number of broad disease areas and lab methods, which constitute simple, generalizable representations of the firm's theories regarding scientists' drug trial activities.

Data science at Airbnb

The strategy of accommodation sharing platform Airbnb is driven by its capabilities for coordinating a marketplace of hosts and guests. On the one hand, Airbnb's strategy depends on heavily centralized mechanisms for secure, seamless payment and scheduling, and for matching preferences by using analytics and inference techniques on massive amounts of structured data. Yet it also depends on enabling millions of users to engage with the semantically rich content of listings for deciding where to stay or who to host.

Airbnb has over five million listings that are continually changing and being updated by its users. It summarizes its approach to supporting user engagement as 'think[ing] of data as the voice of our users at scale'[7]. To generate insights from users' data to support engagement, Airbnb has adopted an organizational design based on viewing all employees as autonomous data scientists making 'data-informed decisions'. The driving assumption is that effective insights are necessarily developed within particular employees' problem contexts. As one executive put it, 'the person asking the question always has the best context on the question they are trying to answer… we believe that people have the capability to think critically and understand the data on their own, and we want to give them the tools to do it.'[8] Employees conduct data science activities using a wide range of both AI and statistical tools, such as machine learning, natural language processing, and computational social science models.

Given users' diverse data, employees' particular problem contexts, and the variety of data science tools, integrating insights at the firm-level regarding user engagement depends on strategic decisions regarding data integration as a solution to information provision. Airbnb has carefully

[7] https://medium.com/airbnb-engineering/at-airbnb-data-science-belongs-everywhere-917250c6beba
[8] https://medium.com/airbnb-engineering/how-airbnb-democratizes-data-science-with-data-university-3eccc71e073a

built an initial core set of data regarding user engagement, with rigid standards for data quality. It has developed a 'Data University' for employees regardless of their role, along with a suite of data querying tools for which usage is a core metric in evaluating the quality of the firm's decision-making processes. Further, it manages a knowledge repository driven by its own standardized formats for making 'posts' of the results of data science activities (called 'Knowledge RePo's') to enable sharing and querying any employees' insights. These choices of data standards, the content of its Data University curriculum, and its standardized querying tools and formats for knowledge posts all can be viewed representations regarding employees' data science activities. As they are all in some way identified at the level of data and are 'pure' of any solutions to division of labor, we argue that they can be viewed as constituting Airbnb's data model regarding user engagement.

STRATEGIC DECISIONS: DEVELOPING DATA MODELS AS SOLUTIONS TO INFORMATION PROVISION

The above short cases illustrate how firms with diverse value propositions commonly seek solutions to information provision for coordinating beyond boundaries. By representing theories of activities — i.e., conducting drug trials at Novartis or engaging users at Airbnb — in terms of a data model that is largely 'pure' from solutions to division of labor, the firm can coordinate ongoing processes of integrating semi-structured data. In so doing, the firm enables agents to create, modify and manipulate representations to frame and find their problem contexts and self-organize their activities. We next consider the strategic decisions that a firm makes to develop its data model. First, we propose that the firm identifies 'archetypal problems' of information provision. Second it chooses how to represent these archetypal problems according to agents' contexts. Finally, the firm identifies procedures for assigning credit to induce collective data integration efforts.

Identifying 'archetypal' problems of information provision

Under our framework, the effectiveness and flexibility of a firm's data model depends on having good theories of agents' activities in regards to information provision. It follows that a baseline requirement is to carefully identify the broader phenomena to which these activities belong. For example, Novartis relies on the domain-specific expertise of a scientific advisory board to determine which health problems are important to represent regarding its lab scientists' drug trials. The ability to develop good theories further depends on characterizing the archetypal ways in which agents create, modify and manipulate representations of their problem contexts regarding their activities. The firm therefore must translate its understanding of the broad phenomena into logically consistent, simple, generalizable and generative representations that are implementable in a database and valuable for its agents to use. For example, Airbnb developed rigid standards for the structure and quality of those aspects of its database that need to be accessed across virtually all data science activities.

Whether supporting drug trials at Novartis or data science activities at Airbnb, information provision beyond boundaries thus depends on identifying, at the level of data, the most characteristic problems regarding agents' activities that they need to frame and find by creating, modifying and manipulating representations. We thus describe the foundational strategic decision underlying a firm's solutions to information provision beyond boundaries as the identification of recurring, 'archetypal'[9] problems that agents frame and find.

Representing agents' problem contexts

Archetypal problems identified at the level of data need to be made available by the firm as representations that agents can create, modify and manipulate according to their problem contexts. The more richly that agents can represent their problem contexts consistent with the firm's archetypal problems, the more the firm can gather data for training its AI to process information that is specific to one agent's context to make it relevant and available for use in other agents' contexts. Such ability to link context-specific representations to the firm's database is at the heart of how the data integration processes we described earlier enable value creation in a firm's activities. As one Airbnb executive simply put it regarding its data science activities, 'learning from your colleagues' knowledge and experience can be much better than spending hours searching for half-baked solutions on the internet'[10].

Consider a lab scientist who queries a firm's database regarding its drug discovery activities to explore how to design a particular drug trial. The query results may enable the scientist to update internal representations (i.e., representations held in the mind (Csaszar, 2018)) regarding the drug trial, without expressing it externally in the firm's database. Such a scenario may be sufficient if the firm effectively integrates agents' representations through organizational mechanisms, such as shaping communication channels or using IT tools for communication.

---

[9] We thank Jackson Nickerson for suggestion of this term. Nickerson Wuebker and Zenger (2017: 275) tentatively advanced the term of a 'problem domain' to get at the idea that there are core problems that agents need to frame and find. 'Archetypal' problems capture how the firm cannot anticipate which problems comprise the domain in advance, and thus can talk about the domain only in a highly abstract way.

[10] https://blog.socialcops.com/technology/engineering/airbnb-knowledge-repository-scale-knowledge/

Alternatively, as with Airbnb's 'knowledge RePo's' and suite of querying tools, the firm can seek to enable the lab scientist to externalize the updated representations into standard formats and mechanisms for creating, modifying and manipulating representations of problems.

Let's take this latter scenario and assume that the firm has identified good theories of its activities that indeed capture archetypal problems, and do so at the level of data. Further, let's assume that the firm develops standard formats and querying tools regarding its database that allow agents to represent their problem contexts consistent with how the firm's archetypal problems are represented. It follows that the lab scientist's representations would also be logically consistent with the database. To the extent that the lab scientist's representations are also simple, generalizable, and generative, the representations further effectively become good 'micro-theories' of the firm's activities (Kornfeld and Hewitt, 1981). This is a critical point as, by enabling the agents to themselves represent theories of the firm's activities, their representations can plausibly be integrated into the firm's database. What's more, by representing 'micro-theories' in standard formats, the firm develops powerful, theoretically-grounded 'genres' for structuring effective, flexible organizational communication (Yates and Orlikowski, 2002). We thus describe a second strategic decision in developing a firm's data model for information provision beyond boundaries as its choice of mechanisms — e.g., standard formats and querying tools — for agents to represent their problem contexts.

'Due process' in credit assignment

The need for the collective efforts of agents in ongoing processes of data integration makes it difficult to assign credit for outcomes (Levinthal and Warglien, 1999). Given that data integration depends on the costly efforts of agents to externalize representations of their problem contexts into the firm's database, the firm faces potential agency problems that call for solutions to rewards provision (Puranam, 2018). In the micro-structural approach to organizational design, agency problems are divided into those of execution and search according to whether tasks are known. We assume the firm has insufficient understanding of tasks underlying its division of labor, such that supporting ongoing processes of data integration relates to agency problems of search. At Novartis, for example, allocating rewards for outcomes is hard because drug discovery depends on long-run efforts that may be distributed across numerous drug trials. In regards to information provision, it is particularly hard to assign credit ex ante or ex post for efforts towards goals if the firm specifies a data model based only on identifying archetypal problems. Indeterminacy in credit assignment places limits on firms' conventional use of accounting systems. The firm may instead adopt procedures for evaluation and negotiation of credit, analogous to the idea of 'due process' under the law — that is, the firm does not endeavor to guarantee fair evaluation per se, but only a fair evaluation process (King and Star, 1990). Airbnb and Novartis, for example, have established procedures for evaluating employees' insights based on processes resembling peer-review from academia or code-review from software engineering.

DISCUSSION

The issue of coordinating beyond boundaries is not new. Firms' growing use of digital data and AI as strategic resources, however, creates distinct opportunities and challenges for coordinating the firm's activities based on solutions to information provision. Harnessing digital data can enhance the firm's ability to augment agents' capacity to process information, and thus for agents to self-organize their activities. A central challenge, however, is that the data that a firm needs to process for the purposes of information provision tends to be used and generated within isolated silos. As a result of the specificity of data to an agent's problem context, the data is difficult to harness and use by agents in different contexts. Even if the firm collects rich data, then, harnessing it in a firm-level database to support the information needs of other agents depends on ongoing data integration processes.

Extant research on coordination beyond boundaries has mostly been confined to examining relatively macro-structural solutions to division of labor. Identifying how tasks are divided and allocated, however, does not directly capture the challenge of data integration raised here, and which is intimately related to firms' growing use of digital data and artificial intelligence as strategic resources. In this paper, we have thus sought to unpack this challenge by providing a framework for understanding coordination by the firm in terms of micro-structural solutions for information provision. Our framework proposes that a firm's choices regarding its data model serve as solutions to information provision regarding ongoing processes of data integration using AI. Effective and flexible data integration processes enhance the firm's ability to self-organize activities by enabling agents to create, modify and manipulate representations for framing and finding their problem contexts. The firm can enable effective integration by developing a data model that is largely 'pure' from solutions to division of labor. By 'pure', we mean a data model that does not depend on how a firm divides or allocates tasks, and thus can be flexible enough to adapt to pervasive inconsistencies in data. We propose that a firm's data model is effective and flexible when it resembles a good theory of information provision, meaning that it is logically consistent, simple, generalizable and generative across its agents' problem contexts. A firm's strategic decisions for developing an effective, flexible data model in turn concern identifying 'archetypal problems' of information provision, representing agents' problem contexts, and assigning credit for data integration efforts. We describe strategic decisions for developing data models as solutions to information provision in Table 3 below.

TABLE 3:
Developing Data Models as Solutions to Information Provision

Strategic Decisions	**Definition**	**Strategic Challenge**	**Examples**
Archetypal Problems	Representations of characteristic, recurring problems of information provision in agents' activities	Translating representations into a logically consistent database	Airbnb: rigid standards for the structure and quality of aspects of its database that need to be accessed in virtually all data science activities
Representations of Problem Contexts	Standardized formats and querying tools for creating, modifying and manipulating representations	Effort and costs for agents to translate context-specific representations	Airbnb: developing 'knowledge posts' and a suite of query tools to integrate data science insights
'Due Process' in Credit Assignment	Procedures for evaluation and negotiation of credit	Collective data integration efforts hard to evaluate *ex*	Novartis: adopting rules inspired by peer review in academia to evaluate drug

		ante and *ex post*	trials

Our framework contributes to the strategy and organizational design literatures by showing how coordinating activities beyond firms' boundaries has major, yet neglected, micro-structural effects on how firms organize. Our core insight is that organizational design can be modeled at the level of data. We next discuss how the framework we develop from our core insight suggests novel implications for the literatures on managerial capabilities and corporate strategy amid digitalization. Finally, we lay out future directions for how a firm's data model can be used as a variable to capture information provision in models of strategic representations.

Data Abstraction as a Managerial Capability

Scholars have given increasing attention to the role of managers' representations in processes of strategy formulation and decision-making (e.g., Csaszar, 2018; Nickerson, 2018). While this work has tended to draw on cognitive science theories, our framework implies that formulating and managing a strategy for information provision is so specific to the complexities of data integration as to call for distinct capabilities for designing and reasoning about representations at the level of data. Similarly, computing scholars and practitioners have argued that designing and reasoning about digital code requires 'computational thinking', which they refer to as the ability to build simple but powerful *data abstractions* (Wing, 2006). In software design, processes of data abstraction concern thinking about digital code at a high level. To provide information beyond a firm's boundaries, we argue that managers need a more expansive but analogous ability to construct representations of a firm's activities. Such ability depends on an integrated understanding of the firm's theories that relate its activities to value creation, the information needs of its agents, and the core requirements for implementing the firm's representations of activities in a database. Next, we draw on our framework to suggest that considering data abstraction as a distinct managerial capability for providing information beyond boundaries points to novel implications for identifying a firm's architectural knowledge at the level of data.

Architectural knowledge at the level of data

Classic perspectives within organizational design assume that managers' 'architectural knowledge' — or knowledge about how elements of products, tasks and roles fit together — can be specific to a firm's product architecture such that the firm may be unable to adapt sufficiently to avoid 'disruption' from even seemingly small innovations in an industry or ecosystem (Henderson and Clark, 1990). These perspectives continue to be drawn on in recent work on organizational design and technological change (e.g., Baldwin, 2015) yet they do not explicitly capture the dynamics of information provision underlying coordination beyond boundaries. Rather, these theories heavily derive from industrial product engineering, where organizational design is based on solutions to division of labor.

Under our framework, the firm's ability to adapt depends on coordinating ongoing processes of data integration, which are constrained or enabled according to the 'pureness' of a firm's solutions to information provision. We argue that the ability to integrate data depends on whether a firm identifies its data model regarding archetypal problems, or whether its data model is also significantly guided by how a firm divides and allocates its tasks. For example, to go back to the earlier example of Novartis, its managers note a need to avoid a 'disinvestment trap'. This trap refers to how the ability to implement scope decisions that will drive long-run growth (i.e., by identifying basic disease areas) can be easily eroded by resource allocation towards shorter-run opportunities based on market growth projections (i.e. of particular pharmaceuticals markets). We argue that, if architectural knowledge instead is viewed at the more abstract level of data — that is, in terms how the concepts underlying a firm's data model fit together — the firm may be able to adapt to short-run change by developing and representing knowledge about archetypal problems underlying its value proposition likely to persist over the long-run.

Based on systems dynamics methods (Rahmandad, 2015), research could use data models as a variable that shapes the robustness of the firm's architectural knowledge by guiding the long-run dynamics of its data integration processes. The hypothesis would be as follows. Aligning a data model with solutions to division of labor may ossify the firm's data integration processes, and thus the robustness of its architectural knowledge to adapt to change in its technological and competitive environments. A data model that instead equates to a simpler, more generalizable, and more generative 'theory for the firm' (Felin and Zenger, 2017) would increase the robustness of architectural knowledge by subsuming innovations in product architectures. For example, Tesla seeks to disrupt the automotive industry not through a small but powerful change in the components of a car, but by developing the ability to integrate data from cars to enable semi-autonomous driving.

Corporate Strategy amid Digitalization: A Plea for Micro-Structural Approaches

Our use of the micro-structural approach to organizational design for analyzing coordination beyond boundaries contrasts with the approaches of much recent work on corporate strategy amid advances in IT (i.e., literatures on digitalization or digital transformation). As the boundaries between a firm's internal and external activities become blurred, scholars observe that scope decisions go beyond positioning internal resources and capabilities within a value chain to involve integrating the activities of an entire ecosystem of value (e.g., Jacobides *et al.*, 2018). An ecosystems view, however, treats scope decisions increasingly in terms of variables once confined to internal organizational design elements (i.e., configuring product components across platforms, or orchestrating agents in an open innovation community), or newly considered as design elements (i.e., strategy formulation as business model design). Theories derived from such relatively macro approaches tend to require assuming that the firm maintains relatively arms-length relationships with its agents. Our framework highlights how organizational design beyond boundaries is in fact deeply dependent on how a firm integrates the efforts of agents at the micro-structural level of information provision, exemplifying how 'understanding the micro is necessary, if not sufficient, to truly understand and re-design the macro' (Puranam, 2018: 14).

Management scholars do broadly agree that a fundamental implication of blurring firm boundaries is that scope decisions become more endogenous. The firm does not just choose a set of activities within a given ecosystem of value, but shapes the boundaries of the ecosystem itself (Felin, Kauffman, Koppl and Longo, 2014; Gavetti *et al.*, 2017). Macro-structural approaches capture this growing endogeneity of scope decisions amid advances in IT in terms of how firms more flexibly configure or orchestrate broad components or rules underlying the activities of an ecosystem of value. Microfoundations perspectives imply, however, that such an approach requires assuming that the nature of the activities in the ecosystem is mostly given (Alvarez and Barney, 2007). We highlight how a firm's ability to coordinate its activities is driven by its very theories for characterizing its activities — that is, its data model identified in terms of archetypal problems of information provision. Our framework thus contributes a micro-structural basis for modeling scope decisions more explicitly in terms of the nature of activities themselves that the firm coordinates. By drawing on our definition of a firm's data model, future work could examine how firms identify archetypal problems of information provision that would give insight into the nature and origins of a firm's activities in ecosystems or other contexts in which the firm coordinates beyond boundaries.

Future Directions: Information Provision in Models of Strategic Representations

Despite being a fundamental problem of organizing, information provision has been relatively neglected in organizational design research, let alone in the broader strategy literature. In organizational design, information provision has typically been identified more in terms of the other three fundamental problems of organizing of task division, task allocation and rewards provision. Even in classic 'information processing perspectives' (Thompson, 1967), information processing is derivative of assumptions about task interdependencies that have to do with the division of labor. In more recent work on digitalization and strategy, IT is invoked primarily as an enabler of the disaggregation of components of products or tasks (thus based on task allocation in a firm's solutions to division of labor), or of communication beyond firm boundaries (thus based on role allocation in a firm's solutions to division of labor).

We argue that, as the boundaries between a firm's internal and external activities become blurred, it becomes critical to develop models of strategic representations that explicitly account for the firm's micro-structural solutions to information provision. To the extent that blurred boundaries limits the firm's control over tasks and roles, 'information about the relevant capabilities gets more dispersed among the individuals in the firm' (Knudsen, Levinthal and Winter, 2014: 1571). If information is dispersed across individual agents, the firm may not sufficiently understand the contexts of the problems that these agents need to solve to deliver the firm's value propositions (Knudsen et al., 2014). In such case, the firm may have insufficient ability to shape information flows through conventional means such as structuring tasks or communication channels according to roles. Firms may instead seek to coordinate by enabling agents to self-organize their activities by providing information for them to represent problem contexts (Simon, 1947; Schelling, 1960; Beer, 1962).

A primary contribution of this paper is thus to identify how firms coordinate activities in terms of an organizational design *based* on solutions to information provision — that is, based on a data model, rather than structures of tasks, roles, agents, or product components related to solutions to division of labor. We next outline how this contribution offers directions for future research by showing how a firm's data model can be used as an organizational design variable to capture information provision in models of strategic representations.

Information provision in models of strategic representations

A hallmark advantage of the micro-structural approach to organizational design is that identifying micro-structures enables formal theorizing to allow precise insight into the dynamics of coordination not possible with verbal theory (Puranam, 2018). Regarding our framework's focus on how self-organizing agents frame and find problem contexts, for example, recent work has explored the micro-structural dynamics of coupled learning according to a firm's initial choice of representations (Puranam and Swamy, 2016). A firm's data model can extend such work by serving as an organizational design variable for capturing how a firm's solutions to information provision structure how the firm's initial representations given to agents change as agents interact.

To use a firm's data model as a variable for such research, we propose building on the methodology and agenda laid out by Csaszar (2018), in which representations can be operationalized as *functions*. In Csaszar (2018), where the focus is on strategic decisions, in which the representations of managers are considered as profit functions that, from our perspective, can be related to a 'theory for the firm' (Felin and Zenger, 2015). Our framework can be further used to extend this methodology to the organizational-level by treating the firm's data model for information provision as a variable within the manager's profit function. Specifically, the firm's data model could be evaluated according to assumptions regarding its logical consistency, simplicity, generalizability and generativity. Given the lack of prior theorizing and inherent difficulty in identifying a firm's data model, such approaches could benefit from a range of approaches, including NK models, mixed-methods using systems dynamics (Rahmandad, 2015), or fieldwork regarding the nature of a firm's data model or its data integration processes. Drawing further on Csaszar (2018), operationalizing a firm's data model would enable tapping into diverse areas of the strategy literature, as the representation of a data model could be used to relate a firm's data to its resources (Makadok, 2012) or the strategic dimensions to which management pays attention (Ocasio, 1997).

CONCLUSION

We have developed a framework for understanding how firms coordinate beyond boundaries based on ongoing processes of integrating semi-structured data using AI. By integrating data, the firm enables agents to create, modify and manipulate representations to frame and find their problem contexts and self-organize activities. We have argued that the effectiveness and flexibility of data integration processes is driven by the extent to which a firm's data model corresponds to a 'pure' solution to information provision, independent from how a firm's tasks are divided and allocated. Developing such data model requires strategically identifying archetypal problems, structures for externalizing agents' representations and converting these into a database for the firm's AI, and procedures for 'due process' in assigning credit for data integration efforts. Our micro-structural approach contributes by theorizing organizational design based on information provision, a fundamental problem of organizing that has nonetheless remained relatively neglected even amid recent advances in IT.

REFERENCES

Agrawal A, Gans J, Goldfarb A. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Press: Cambridge, MA.

Afuah A, Tucci CL. 2012. Crowdsourcing as a solution to distant search. *Academy of Management Review*, **37**(3): 355-375.

Alvarez SA, Barney JB. 2007. Discovery and creation: alternative theories of entrepreneurial action. *Strategic Entrepreneurship Journal*, **1**(1-2): 11-26.

Baldwin CY. 2015. Bottlenecks, modules and dynamic architectural capabilities. *Harvard Business School Finance Working Paper* (15-028).

Beer S. 1962. *Brain of the Firm*. Allen Lane, Penguin: Harmondsworth.

Benner M, Tushman M. 2015. Reflections on the 2013 decade award—"Exploitation, exploration, and process management: The productivity dilemma revisited" Ten years later. *Academy of Management Review*, **40**(4): 497-514.

Carlile P. 2006. Artifacts and knowledge negotiation across domains. In, Rafaeli A, Pratt MG, eds. *Artifacts and Organizations: Beyond Mere Symbolism*: 101–118. Lawrence Erlbaum Associates: Mahwah, NJ.

Codd EF. 1970. A relational model of data for large shared data banks. *Communications of the ACM*, **13**(6): 377-387.

Csaszar FA, Levinthal DA. 2016. Mental representation and the discovery of new strategies. *Strategic Management Journal*, **37**(10): 2031-2049.

Csaszar FA. 2018. What makes a decision strategic? Strategic representations. *Strategy Science*, **3**(4): 606-619.

Davenport TH, Ronanki R. 2018. Artificial intelligence for the real world. *Harvard Business Review*, **96**(1): 108-116.

Felin T, Kauffman S, Koppl R, Longo G. 2014. Economic opportunity and evolution: Beyond landscapes and bounded rationality. *Strategic Entrepreneurship Journal*, **8**(4): 269-282.

Felin T, Zenger TR. 2015. Crossroads—Strategy, problems, and a theory for the firm. *Organization Science*, **27**(1): 222-231.

Felin T, Zenger TR. 2017. The theory-based view: Economic actors as theorists. *Strategy Science*, **2**(4): 258-271.

Fjeldstad ØD, Snow CC, Miles RE, Lettl C. 2012. The architecture of collaboration. *Strategic*

Management Journal, **33**(6): 734-750.
Fjeldstad ØD, Snow CC. 2018. Business models and organization design. *Long Range Planning*, **51**(1): 32-39.
Foss NJ, Klein PG. 2012. *Organizing Entrepreneurial Judgment: A New Approach to the Firm*. Cambridge University Press: London.
Garg P, Zhao M. 2018. Knowledge sourcing by multidivisional firms. *Strategic Management Journal*, **39**(13): 3326-3354.
Gavetti G, Helfat CE, Marengo L. 2017. Searching, shaping, and the quest for superior performance. *Strategy Science*, **2**(3): 194-209.
Gawer A. 2014. Bridging differing perspectives on technological platforms: Toward an integrative framework, *Research Policy*, **43**(7): 1239-1249.
Gershman SJ, Horvitz EJ, Tenenbaum JB. 2015. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, **349**(6245): 273-278.
Gulati R, Puranam P, Tushman M. 2012. Meta-organization design: Rethinking design in interorganizational and community contexts. *Strategic Management Journal*, **33**(6): 571-586.
Helfat CE, Peteraf MA. 2015. Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal*, **36**(6): 831-850.
Helfat CE, Raubitschek RS. 2018. Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems. *Research Policy*, **47**(8): 1391-1399.
Henderson RM, Clark KB. 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly:* 9-30.
Hewitt C. 2014. Offices are open systems. In, *Readings in Distributed Artificial Intelligence*: 321-329.
Jacobides M, Cennamo, C, Gawer A. 2018. Towards a theory of ecosystems. *Strategic Management Journal*.
Joseph J, Baumann O, Burton R, Srikanth K. 2018. Reviewing, Revisiting, and Renewing the Foundations of Organization Design. In, *Organization Design* (1-23). Emerald Publishing Limited.
Kallinikos J. 2004. Farewell to constructivism: Technology and context-embedded action. In, *The*

Social Study of Information and Communication Technology: Innovation, Actors, and Contexts, 140-161.

King JL, Star SL. 1990. Conceptual foundations for the development of organizational decision support systems. In *Proceedings of the Twenty-Third Annual Hawaii International Conference on System Sciences*, **3**: 143-151.

Knudsen T, Levinthal DA, Winter SG. 2014. Hidden but in plain sight: The role of scale adjustment in industry dynamics. *Strategic Management Journal*, **35**(11): 1569-1584.

Kornfeld WA, Hewitt CE. 1981. The scientific community metaphor. *IEEE Transactions on Systems, Man, and Cybernetics*, **11**(1): 24-33.

Lave CA, March JG. 1993. *An Introduction to Models in the Social Sciences*. University Press of America: Lanham, MD.

Lawrence PR, Lorsch JW. 1967. Differentiation and integration in complex organizations. *Administrative Science Quarterly*: **12(**1-47).

Leonardi PM, Bailey DE. 2008. Transformational technologies and the creation of new work practices: Making implicit knowledge explicit in task-based offshoring. *MIS Quarterly*: 411-436.

Levinthal, D. 2000. Organizational capabilities in complex worlds. In, *The Nature and Dynamics of Organizational Capabilities*, Winter, S. (ed): 363-379. Oxford University Press: Oxford.

Levinthal DA, Warglien M. 1999. Landscape design: Designing for local action in complex worlds. *Organization Science*, **10**(3): 342-357.

Makadok R. 2001. Toward a synthesis of the resource-based and dynamic-capability views of rent creation. *Strategic Management Journal*. **22**(5): 387–401.

March J, Simon H. 1958. *Organizations*. Wiley: New York.

March J. 2006. Rationality, foolishness, and adaptive intelligence. *Strategic Management Journal*, **27**(3): 201-214.

Ng A. 2017. The state of artificial intelligence. www.youtube.com/watch?v=NKpuX_yzdYs

Nickerson J, Argyres N. 2018. Strategizing before strategic decision making. *Strategy Science*, **3**(4): 592-605.

Nickerson JA, Wuebker R, Zenger T. 2017. Problems, theories, and governing the crowd. *Strategic Organization*, **15**(2): 275-288.

Ocasio W. 1997. Towards an attention-based view of the firm. *Strategic Management Journal*.

18(S1):187–206.

Puranam P. 2018. *Microstructure of Organizations*. Oxford University Press: Oxford.

Puranam P, Swamy M, 2016. How initial representations shape coupled learning processes. *Organization Science*, **27**(2): 323-335.

Puranam P, Alexy O, Reitzig M. 2014. What's "new" about new forms of organizing? *Academy of Management Review*, **39**(2): 162-180.

Rahmandad H, 2015. Connecting strategy and system dynamics: An example and lessons learned. *System Dynamics Review*, **31**(3): 149-172.

Ramakrishnan R, Gehrke J. 2000. *Database Management Systems*. McGraw Hill: New York.

Ransbotham S, Kiron D, Gerbert P, Reeves M. 2017. Reshaping business with artificial intelligence, *MIT Sloan Management Review and The Boston Consulting Group*.

Leiblein MJ, Reuer JJ, Zenger T. 2018. What makes a decision strategic?. *Strategy Science*, **3**(4): 558-573.

Schelling TC. 1960. *The Strategy of Conflict*. Harvard University Press: Cambridge, MA.

Simon, H. 1947. *Administrative Behavior*. Free Press: New York.

Stonebraker, M. 2015. A biased take on a moving target: data integration. In, *Readings in Database Systems*, **4**: 1614-1644.

Stonebraker M, Hellerstein J. 2015. What goes around comes around. *Readings in Database Systems*, **4**: 1724-1735.

Stonebraker M, Ilyas IF. 2018. Data integration: The current status and the way forward. *IEEE Data Engineering Bulletin*, **41**(2): 3-9.

Teece DJ. 2017. Dynamic capabilities and (digital) platform lifecycles. In *Entrepreneurship, Innovation, and Platforms*: 211-225. Emerald Publishing Ltd.: New York.

Thomas LD, Autio E, Gann DM. 2014. Architectural leverage: Putting platforms in context. *Academy of Management Perspectives*, **28**(2): 198-219.

Thompson JD. 1967. *Organizations in Action: Social Science Bases of Administrative Theory*. Transaction Publishers: London.

Tushman ML, Nadler DA. 1978. Information processing as an integrating concept in organizational design. *Academy of Management Review*, **3**(3): 613-624.

Wing JM. 2006. Computational thinking. *Communications of the ACM*, **49**(3): 33-35.

Yates J, Orlikowski W. 2002. Genre systems: structuring interaction through communicative norms. *Journal of Business Communication*. **39**(1): 13-35.