Causal Machine Learning and Business Decision Making*

Paul Hünermund[†] Jermain Kaminski[‡] Carla Schmitt[§]

February 15, 2021

Abstract

Causal knowledge is critical for strategic and organizational decision making. By contrast, standard machine learning approaches remain purely correlational and prediction-based, rendering them unsuitable for addressing a wide variety of managerial decision problems. Taking a mixed-methods approach, which relies on multiple sources, including semi-structured interviews with data scientists and decision makers, as well as quantitative survey data, this study makes a first attempt at delineating causality as a critical boundary condition for the application of machine learning in a business analytical context. It highlights the crucial role of theory in causal inference and offers a new perspective on human-machine interaction for data-augmented decision making.

Keywords: Organizational decision making, strategic management, data science, causality, machine learning

*Names appear in alphabetical order. All authors contributed equally.

[†]Copenhagen Business School, Department of Strategy and Innovation, Kilevej 14A, DK-2000 Frederiksberg. Email: phu.si@cbs.dk

[‡]Maastricht University, School of Business and Economics, Department of Organisation, Strategy, and Entrepreneurship, Tomgersestraat 53, NL-6211 LM Maastricht. Email: j.kaminski@maastrichtuniversity.nl

[§]Maastricht University, School of Business and Economics. Email: carla.schmitt@maastrichtuniversity.nl

INTRODUCTION

The age of big data has given rise to data science and machine learning as a promising tool for promoting organizational and strategic decision making, whereby managers rely less on intuition and more on data (Brynjolfsson and McElheran, 2016). Machine learning technologies have thus become a workhorse in many organizations (Agrawal et al., 2018), significantly driving firm value and performance (Rahmati et al., 2020; Mithas et al., 2011). Examples for how data-driven strategies have become instrumental to achieving competitive advantage are abundant (LaValle et al., 2011; Brynjolfsson et al., 2011; Mithas et al., 2011; Bloom et al., 2012; Brynjolfsson and McElheran, 2019; Tidhar and Eisenhardt, 2020). It is for this reason that some authors have classified data science and machine learning as a newly emerging *general purpose technology* (Goldfarb et al., 2020).

Machine learning thereby commonly refers to a set of statistical algorithms that are designed to efficiently detect patterns in high-dimensional data and fit functional relationships between variables at a great degree of accuracy (Hastie et al., 2009; Shrestha et al., 2021). They form the architectural underpinning to modern approaches in business analytics and artificial intelligence (AI). This is exemplified by the spectacular success of deep learning algorithms in recent years that have been applied to a variety of different decision problems (Blei and Smyth, 2017; Athey and Imbens, 2019; Choudhury et al., 2020; Tidhar and Eisenhardt, 2020). Because of their superior forecasting abilities, compared to traditional statistical and econometric techniques, machine learning algorithms have been labeled as *"prediction machines"* (Agrawal, 2018); a term which captures well their main purpose of predicting the state of an output variable based on complex correlational patterns in the input data (the so-called *feature space*).

At the same time, however, this term also illustrates a potential boundary condition for using machine learning in a business analytical context. Organizational and strategic decision making involves deliberate actions and interventions in the environment (both internal and external) with the aim of achieving a desired result in line with organizational goals (Cyert et al., 1956; Simon, 1964; Christensen et al., 2016). Assessing the likely impact of these interventions ex-ante – a capacity that is crucial for optimized decision making – requires causal knowledge (Bertsimas and Kallus, 2016; Athey, 2017; Bareinboim et al., 2020). I.e., to generate and evaluate alternative strategic actions in terms of their effect on central business metrics, managers need to understand the causal mechanisms underlying a decision situation (Mintzberg et al., 1976). By contrast, most commonly used machine learning algorithms, including decision tress, supportvector machines, and deep learning, remain purely correlational and are thus only able to make accurate predictions in a static domain (Pearl, 2019). Once perturbations to the environment are introduced as a result of a deliberate managerial action their superior forecasting ability breaks down.

Therefore the question arises to what extent machine learning and the data-scientific approaches that build on it are really useful for improving business decision making? In this paper, we study whether there is a mismatch between the managerial problems that organizations try to tackle with data analytics and the methods they are using when it comes to the challenge of causal inference. If this hypothesis is confirmed, we would further like to know whether practitioners are aware of this gap and what actions they take to overcome it. To this end, we employ a mixed methods research design in which we combine qualitative interviews with a quantitative survey among practitioners, as well as a multitude of other data sources, such as online resources, educational material, blog posts and software packages originating from the data science community. The interviews we conducted thereby provide us with rich, contextual insights about the mechanisms underlying modern business analytics in contemporary organizations (Bettis et al., 2014), while the survey study allows us to solicit information from a much broader sample in a more systematic way (De Leeuw et al., 2008).

The results indicate an ongoing shift in the practitioners community towards the growing application of causal data science methods for business decision making. Traditional correlation-based machine learning approaches are increasingly perceived as not being suitable for informing a large variety of practical decision problems. Moving to causal methods, including experimental and observational approaches, by contrast, offers the prospects of increasing the reliability and robustness of obtained data analytic insights. Moreover, we find that the organizations under study plan to invest more into their causal inference capabilities in the coming years. Several examples of key players, particularly in the technology sector, that have started to significantly increase their efforts in this direction demonstrate that the topic of causality will grow in importance for the industry as a whole in the future. Yet, moving towards this new paradigm poses a number of practical as well as theoretical challenges that will be identified in the course of this paper.

Our study contributes to the literature by clarifying the epistemological foundation for causal learning in an organizational decision making context and delineating theoretical impediments for the use of standard machine learning approaches in business analytics (Pearl, 2019). We discuss the crucial role of ex-ante domain knowledge that cannot be obtained from pure observation alone for inferring causality (Bareinboim et al., 2020). In doing so, we connect to the newly emerging theory-based view of the firm (Camuffo et al., 2020; Felin and Zenger, 2009, 2017; Felin et al., 2020a,b) and demonstrate that theory is an essential input to *data-augmented decision making*. Conversely, we show how the causal inference literature in machine learning and AI can significantly contribute to the inferential power of managerial theorizing and map out avenues for the strategy literature to more effectively integrate data science into the strategy formulation and decision making process. Finally, we discuss the practical implications of our study with respect to developing causal learning as an important organizational capability.

THEORY

Causal Knowledge in Strategic Management

Causal knowledge revolves around the awareness and utilization of cause and effect relationships in the world. The ability to understand and acquire causal knowledge is seen as one of the most important components of human cognition; inseparable of our thought and essential to our survival (Pearl and Mackenzie, 2018; Waldmann, 1996). Knowledge of cause and effect relationships in particular, enables an actor to predict and understand the outcome of an action and the mechanism it is transmitted by, allowing her to deliberately change the state of the environment with selective interventions (Pearl and Mackenzie, 2018). Indeed, according to Woodward (2003), causal knowledge can be defined as ”knowledge that is useful for a very specific kind of prediction problem: the problem an actor faces when she must predict what would happen if she or some other agent were to act in a certain way on the basis of observations of situations in which she or the other agent have not (yet) acted” (p. 32).[1]

Such kind of (causal) prediction problems are ubiquitous in the field of management. To name a few examples: Marketing executives might be interested in the question whether ads on a mobile or desktop version of a social network create higher click-through rates (Lu and Du, 2020). Human resource managers might ask themselves whether increased teleworking would exert a positive influence on employee productivity and well-being (Vega et al., 2015). Founders of a start-up might wonder whether distinct communicative signals in a crowdfunding campaign result in better funding outcomes (Kaminski and Hopp, 2019). “What if” questions of this kind thereby typically arise in the context of strategic business problems and thus constitute an important parameter in taking central management decisions (Felin and Zenger, 2009).

Among organizational decisions taken, strategic decisions are generally identified as those

[1] Woodward is a representative of an interventionist theory of causation within the philosophy of science (Menzies, 2006)

managerial choices that are important in terms of the resources committed, the actions taken, and the precedents set (Shrivastava and Grant, 1985; Mintzberg et al., 1976; Eisenhardt and Zbaracki, 1992; Mitchell et al., 2011). As such, they decisively define the direction of the organization at large (Eisenhardt and Zbaracki, 1992) and thereby exert a fundamental long-term effect on the firm's administration, structure and performance (Shrivastava and Grant, 1985; Shivakumar, 2014). At the same time, however, strategic decisions are particularly difficult to make, as the problems evoking them are often complex, novel and poorly understood, without clearly defined routines and rules to approach them (Shivakumar, 2014; Shrivastava and Grant, 1985) and thus involve considerable uncertainty for decision makers (Schwenk, 1984). Cyert et al. (1956) therefore prescribe that those "non-repetitive [...], [...] basic long-range questions about the whole strategy of the firm or some part of it, initially [arise] in a highly unstructured form" (p. 238) and demand a very particular decision-making process. Organizational success and a firm's performance and efficiency thus depend predominantly on organizational decision making structures (Simon, 1964) and the process underlying managers' strategic choices (Mitchell et al., 2011), which, as the following discussion shows, depends largely on causal knowledge and assumptions. Adopting Mintzberg et al. (1976) descriptive approach towards understanding and defining managers' strategic decision activities, this process can best be conceptualized along *three phases of decision making.* In the first phase, managers are concerned with recognizing a performance-objective gap in the data and thoroughly defining the strategic problem and important relationships. Subsequently, management identifies and designs alternative actions to the problem and finally determines all feasible alternatives and evaluates them along organizational goals and objectives to select a strategic action.

Therefore, to fully identify the strategic problem faced and define a starting point of the process, "management seeks to comprehend the evoking stimuli and determine cause-effect relationships for the decision situation" (Mintzberg et al., 1976, p. 253). Specifying the causal structures underlying a complex problem thereby facilitates problem formulation (Baer et al.,

2013) and aids managers to adequately identify and define important variables and objectives of the decision task (Maule et al., 2003). Similarly, the subsequent generation and evaluation of alternative actions requires the decision maker to process causal assumptions in order to imagine and compare different action scenarios for effective strategic planning (Pearl and Mackenzie, 2018). In their seminal work, Cyert et al. (1956) stress that the unstructured and complex nature of non-programmed, strategic decisions requires a very particular search process. Thereby, alternative actions and the consequences attached to them in the context of the business problem, are not given but must be sought, making this search for cause and effect relationships an integral part of the strategic decision-making process (Cyert et al., 1956). Indeed, Mintzberg et al. (1976) emphasize that ”the largest share of manhours in the decision process [...] [is] devoted to gathering information to determine the consequences of alternatives” (p. 262). Building upon this theory, Nickerson and Zenger (2004) pose that to identify valuable solutions to complex problems and generate knowledge, managers need an implicit theory of the problem space to cognitively evaluate probable effects of choices and determine solution performance.

Practically, such cognitive evaluation can be realized by agents forming and consulting mental images of their information worlds and the problem space (Walsh, 1995; Gavetti and Levinthal, 2000; Pearl and Mackenzie, 2018). The literature on managerial cognition therefore finds that managers do in fact build *causal* mental maps to support their decision-making efforts (Gary and Wood, 2011; Maule et al., 2003; Hodgkinson et al., 1999). Those mental models are generally defined as graphical representations of an individual's causal beliefs in a certain domain (Axelrod, 1976). Cognitive maps thereby act as simplified working models that aid decision makers in overcoming their limited processing capacity when facing complex strategic problems (Walsh, 1995; Hodgkinson et al., 1999; Gavetti and Levinthal, 2000). Not surprisingly, managers' cognition is thus found to be a key determinant of managerial choice and action along the entire decision-making process (Stubbart, 1989; Walsh, 1995). Emphasizing the causal nature of cognitive maps, the literature claims that an understanding of cause and effect in the relevant business context in particular allows decision makers to focus on strategic actions (Hodgkinson et al., 1999), speeds problem solving (Walsh, 1995) and increases the quality of choice (Gary and Wood, 2011) and decision performance (Waldmann, 1996). Beliefs about causal structures thereby assist decision makers in covariate detection and distinguishing real from spurious correlation (Vera-Muñoz et al., 2007; Waldmann, 1996). More specifically, Gary and Wood (2011) argue that causal models guide managers in deciding when and how to intervene in their business by providing them with a tool to infer the effect of alternative strategic actions. Indeed, the authors' analysis shows that "accurate mental models about causal relationships in the business environment result in superior performance outcomes" (ibid., p. 570) and that managerial cognition is a significant driver of firm performance heterogeneity. Similarly, Gavetti and Levinthal (2000) conclude that "even simple models of the world have a tremendous potential to guide search processes" (ibid. p. 135).

In practice, causal knowledge is not only relevant to instantaneous strategic decision making but is also evident in other aspects of the business, such as strategy formation more broadly. As an integral part of the corporate strategy and the essential rationale behind a businesses' undertakings, the business model is generally defined as a cognitive tool that specifies the firm's value creation and capture activities (Baden-Fuller and Mangematin, 2013; Chesbrough and Rosenbloom, 2002). As such, more recent studies (Baden-Fuller and Mangematin, 2013; Furnari, 2015; Vera-Muñoz et al., 2007) reveal an underlying causality in the specification of business models, conceptualizing it as a system with underlying cause and effect relationships that define how the firm can achieve its long-run objectives by realizing concrete strategies. More specifically, Baden-Fuller and Mangematin (2013) describe the business model as a "characterization that captures the essence of the cause–effect relationships between customers, the organization and money" (ibid. p. 419). Hence, encoded in the business model specification are the causal knowledge and assumptions strategic leaders hold about the most important business parameters, relationships and mechanisms related to the strategic objectives and overall business goals. The business model as a cognitive instrument specifying the core activities and goals of the organization thus provides a reference frame of causal relationships to address strategic management questions and structure organizational decision making more generally (Simon, 1964). Indeed, Baden-Fuller and Morgan (2010) stress the manipulable, experimentable component of business models which "provides the kinds of descriptions that can be reasoned with, the kind of resources that can be investigated to answer questions" (p. 163).

Data Science and Machine Learning in Management

Fueled by new machine learning technologies and opportunities of data collection, the emergence of data-augmented decision making has changed the way managers take decisions – relying more on data and less on intuition (Brynjolfsson and McElheran, 2016). Correspondingly, *business intelligence* and *business analytics* capture increasing interest from researchers, job-market candidates and practitioners alike (Chen et al., 2012; Lycett, 2013; Sharma et al., 2014; Athey and Luca, 2019).

Machine learning evolved primarily as a tool for prediction problems (Davenport and Harris, 2009; Varian, 2014, 2016; Agrawal et al., 2018; Iansiti and Lakhani, 2020) and evidence on the positive relationship between data-augmented decisions, enhanced productivity, and the increase of intangible firm value is abundant (Bharadwaj et al., 1999; Brynjolfsson et al., 2011; Mithas et al., 2011; Bloom et al., 2012; Brynjolfsson and McElheran, 2019; Ghasemaghaei and Calic, 2020; Rahmati et al., 2020). By enabling firms to expand the search space of existing knowledge to combine technologies, data analytics is especially valuable and complementary to incremental process innovation (Wu et al., 2020). As organizational processes are generally seen as core features of many organizational capabilities and central to corporate strategy (Bingham and Eisenhardt, 2011), the relevance of data analytics for strategic management is apparent. More generally, LaValle et al. (2011) therefore stress the importance of data analytics in high-performing organizations: "The correlation between performance and analytics-driven management has important implications to organizations, whether they are seeking growth, efficiency or competitive differentiation" (p. 22).

The "promise" of big data in organizations today, hence, essentially lies in significantly advanced predictions derived from gradually improving machine learning models and technologies (Bajari et al., 2018). Consequently, *prediction machines* (Agrawal et al., 2018) have evolved as a workhorse in many companies, providing continuously better and cheaper forecasts to decision makers. Much of this expansion in data-augmented decision making has in fact been fueled by the success of deep learning architectures, that is, models that map from observable to outputs via multiple layers of latent representations (high-dimensional data). These deep learning algorithms represent effective tools for unstructured predictions and can be employed to solve complex classification problems (Athey, 2018; Blei and Smyth, 2017; Mullainathan and Spiess, 2017; Athey and Imbens, 2019; Choudhury et al., 2020). Both supervised and unsupervised, it is employed in contexts (Brynjolfsson and McAfee, 2014; Agrawal, 2018; Mullainathan and Spiess, 2017) ranging from predicting customer churn (Agrawal et al., 2018; Ascarza, 2018), to economic predictions with satellite images satellite images (Henderson et al., 2012; Donaldson and Storeygard, 2016; Athey, 2018), and applied to support hiring decisions (Chalfin et al., 2016). With regard to strategic questions, machine learning systems have shown to be capable of finding the optimal revenue-model fit (Tidhar and Eisenhardt, 2020), or, as the reinforcement learning algorithm of Google DeepMind's AlphaStar exemplifies, are even able to map out and predict strategies in a complex gaming simulation (Vinyals et al., 2019).

As Agrawal et al. (2019, 31) importantly remark however, "machine learning does not represent an increase in artificial general intelligence of the kind that could substitute machines for all aspects of human cognition, but rather one particular aspect of intelligence: prediction". While modern decision-aiding systems are well suited "an exploratory tool to discover robust patterns in quantitative data" (Choudhury et al., 2020, p. 1), they are not capable of deriving causal effects: “First, the goal [of machine learning] is predictive power, rather than estimation of a particular structural or causal parameter.” (Athey and Imbens, 2019, p. 7). In the context of the preceding theoretical discussion, this implies an important mismatch between machine learning capabilities and analytical requirements of problems addressed (Bertsimas and Kallus, 2016), as pattern discovery by itself has been shown to rarely be relevant to strategic management questions. Considering that machine intelligence is mostly preferred by executives in strategic contexts, such as improving or creating new products (Davenport and Ronanki, 2018), this is especially consequential. Indeed, Christensen et al. (2016, p. 4) explain that “though it’s no surprise that correlation isn’t causality, we suspect that most managers have grown comfortable basing decisions on correlations”. The questions addressed by machine learning, however, should in fact be confined to “essentially, any problems that: (1) require prediction rather than causal inference [meaning that you are interested in how, on average aspects of the data relate]; and (2) are sufficiently self-contained, or relatively insulated from outside influences” (Fedyk, 2016, p. 3).

This limitation in explaining causal effects consequently implies that machine learning is constrained for most managerial decision tasks. Outcomes of deep learning models in particular cannot be easily interpreted because of the many feature layers involved in a decision (Rai, 2020, p. 14). As Athey (2017) points out, a number of gaps yet persist between making a prediction and making a decision. To optimize data-augmented decision making in organizations, these gaps need to be addressed by comprehending underlying assumptions (Athey, 2017). A majority of existing machine learning research however focuses on the relationship between data and prediction, while the relationship between prediction and decision is still underdeveloped. As Kleinberg et al. (2017, p. 40) have shown, “being clear about how predictions translate to decisions can substantially influence how the prediction function is evaluated”. Bertsimas and Kallus (2016) and Lycett (2013) explain that prediction models are business problems such as pricing or inventory management, which actually require more robust causal assumptions to guide optimal decision outcomes. Prediction systems lack “the knowledge of causal relationships among the various variables, relying solely on past data to make decisions”. As such, “[machine learning] cannot foresee future consequences as humans can” (Balasubramanian et al., 2020). Hence, the authors derive that while predictive approaches often perform well in business contexts, optimal decisions cannot be identified without a sound assessment of causal effects: “When we have a good understanding of where our data comes from, what has influenced [that] data, the causal relation between [input and output data], we understand where, how and why something happened” (Asatiani et al., 2020, p. 270).

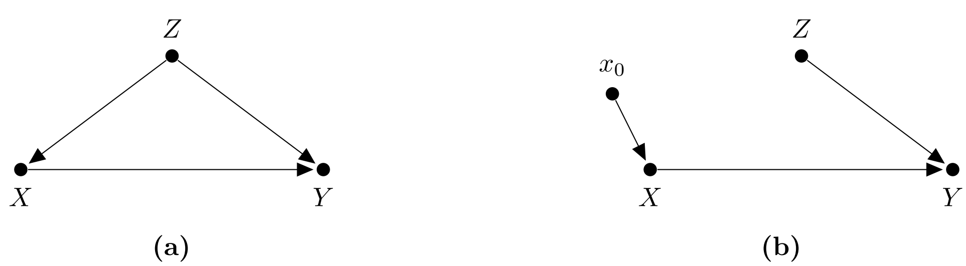
Shiffrin (2016, p. 7308) identifies, that while big data comes with its own challenges of collection and storage, it is itself not very helpful to organizational as the detection of patterns is only the first step towards causal inference. “Explaining those patterns (possibly with the help of experimental manipulations of some variables coupled with additional data collection), and then using the patterns and explanations for a variety of purposes“, such as strategic decision making, are essential steps to derive value from organizational data. As (Gillon et al., 2012, p. 290) put forward, causal approaches such as experimentation significantly advance businesses' understanding of “causal relationships between human behavior and economic value”, which can provide valuable new insights into decision making. Accordingly, Hartford et al. (2016, p. 20) argue more generally that “the next generation of problems in ML involve moving from raw prediction tasks into more complex decision-making domains” which require “knowledge of the true structure of the processes that we are modeling and, hence, causal inference.“. Foreshadowing the future development, Malone (2018, p. 258) projects that “cyber-human strategy machines”, or “strategy combinators” that merge human reasoning with machine learning, “could rapidly generate and evaluate various strategic possibilities”.

The field of causality thus presents a promising future path for business strategy, considerably growing along more recent contributions from statistics, social science and computer science, especially suited for observational data and experiments (Blei and Smyth, 2017). As such we conclude that “big data” needs to be complemented with “smart models” in order to unfold its full potential for data-augmented decisions in business practice.

The Fundamental Challenge of Causal Inference

The task of causal inference, to predict the likely outcome of an action (Woodward, 2003), is challenging. To carry it out, the analyst cannot simply rely on passive observations of the environment, as her action exerts a force on said environment and thereby changes it. The normal course of history, which would have prevailed without the action, gets interrupted and therefore carries only little information about the state of the world that will occur after the action has been taken. Formally, this idea can be best illustrated with the help of a causal diagram (Pearl, 1995; Durand and Vaara, 2009). Figure 1a shows a network of three variables, X, Y, and Z, depicted as nodes connected by edges. These edges are directed, with directions indicated by arrowheads that specify cause-and-effect relationships between the nodes.

FIGURE 1: (a) Directed acyclic graph corresponding to the SCM in equation (1). (b) Post-intervention graph of (a) for $do(X = x_0)$, corresponding to the SCM in equation (2).



The causal diagram in Figure 1a is a representation of the following underlying *structural causal model* (SCM; Pearl, 2009):

$$Z \leftarrow f_1(\varepsilon_1), \quad X \leftarrow f_2(Z, \varepsilon_2), \quad Y \leftarrow f_3(X, Z, \varepsilon 3) \tag{1}$$

Here, X is determined by a function f_2 that takes Z as an argument. The variable Z thus has a direct causal effect on X. Causal relationships are generally assumed to be asymmetric (Cartwright, 2007), captured by the assignment operator $\leftarrow$, which states that while Z is a cause of X, the reverse is not true.[2] The model furthermore contains a set of exogenous background factors, ε_i, that are considered to be determined outside of the model and are thus not further specified. For ease of notation, these background factors are not depicted in the causal diagram. Nonetheless, they exert an influence on the endogenous variables in the model. Because they are unobserved from the standpoint of the analyst, this renders the model stochastic with a probability distribution $P(\varepsilon)$ over the set of endogenous variables.[3]

As a graphical representation of the structural causal model, the causal diagram only relies on the qualitative causal dependencies between nodes. No assumptions about the exact form of the functional relationships, f_i, as well as the distribution of background factors, $P(\varepsilon)$, are needed (Bareinboim and Pearl, 2016). The only requirement is that causal relationships are acyclic (Pearl, 2009). That means that by tracing paths between nodes following the directed edges in the diagram (such as, e.g., $Z \leftarrow X \leftarrow Y$ in Figure 1a) it should not be possible to arrive at a node that has already been visited before on the same path. Hence, feedback loops such as $A \leftarrow B \leftarrow C \leftarrow A$ are ruled out, which captures the intuitive notion that a variable cannot be a cause of itself.[4] Due to this property of acyclicity, causal diagrams are also referred to as *directed acyclic graphs* (DAG) in the literature (Pearl, 1988).

Equipped with the notion of structural causal models, actions can now be defined as interventions on variables in the model (Haavelmo, 1943; Strotz and Wold, 1960). For example, intervening on X in SCM (1) amounts to deleting the function $f_2(\cdot)$, which normally

[2]Equations would not be able to capture this asymmetry, since $X = Z$ is equivalent to $Z = X$.

[3]This notion is analogous to error variables in standard statistical regression theory. It is important to note, however, that background factors have a causal interpretation and do not simply reflect a deviation from a conditional mean function.

[4]Acyclicity only rules out instantaneous feedback loops. Dynamic relationships such as $A_t \rightarrow A_{(t+1)} \rightarrow A_{(t+2)} \rightarrow \cdots$ are permissible though.

assigns values to X, and setting X to a constant value x_0:

$$Z \leftarrow f_1(\varepsilon_1), \quad X \leftarrow x_0, \quad Y \leftarrow f_3(X, Z, \varepsilon 3) \tag{2}$$

This operation is denoted by a special operator, called the do-operator: $do(X = x_0)$. Following this notation, the goal of causal inference is then to assess the quantitative effect of such an intervention on other variables of interest in the model. I.e., if Y is the outcome variable under study, the target quantity becomes $P(y|do(X = x_0))$; in words: the probability of Y, given that X has been set to x_0 (Pearl, 2009, def. 3.2.1). Once this probability distribution is known, other potential target quantities, such as average or quantile treatment effects (Heckman and Vytlacil, 2007), can easily be derived from it.

Interventions can also be illustrated graphically in a DAG. Figure 1b depicts the post-intervention situation corresponding to model (2), in which all the incoming arrows pointing into X are deleted and replaced by a single intervention node x_0. The graphical operation of removing arrows from the graph highlights the fact that an intervention eliminates all the causal relationships that usually exert an influence on X in the naturally occurring *data generating process* (DGP; Hünermund and Bareinboim, 2019). This change of the DGP as a result of the intervention implies, however, that the post-intervention distribution $P(y|do(x))$ is not readily observable from the pre-intervention state. This disparity is described as the difference between *seeing* and *doing* in the literature, which constitutes a formal epistemological hierarchy, also known as the *ladder of causation* (Pearl and Mackenzie, 2018; Bareinboim et al., 2020). [5][6]

[5]The hierarchy states that information at one layer (*seeing*) almost always (in a measure-theoretic sense) underdetermines information at higher layers (*doing*). Additionally, the hierarchy also contains a third layer (*imagining*), which relates to counterfactual reasoning that is enabled by an SCM. For the sake of brevity, we focus only on the step between the first and second layer of the hierarchy since the fundamental challenges of obtaining causal knowledge are already introduced there.

[6]Not every organizational decision requires causal knowledge in the form of $P(y|do(x))$. In many situations, decision making can be improved simply based on passive observations of the DGP, such as, e.g., accurate forecasts of demand Y given product characteristics X (Agrawal et al., 2018). However, decisions based on associational knowledge $P(y|x)$ need to rest outside the system of variables $\{X, Y, Z\}$ under investigation and cannot intervene on it. An example here would be the decision to optimally allocate storage capacity C based on seasonal demand patterns Y. However, if managers want to induce change in the sys-

Nevertheless, under certain circumstances, $P(y|do(x))$ might actually be transferable into an equivalent expression that can be computed from pre-intervention information. For the graph in Figure 1a it can be shown that, based on a powerful causal inference engine called the do-calculus (Pearl, 2009), the post–intervention distribution is expressible as:

$$P(y|do(x)) = \sum_{z} P(y|x,z)P(z), \tag{3}$$

where the right-hand side stands for the conditional probability of Y given X and Z, and integrating over all values of Z. Interestingly, while the left-hand side expression contains a do-operator, and thus relies on post-intervention information, this is not the case for the right-hand side. The expression on the right is comprised only of standard probability objects that can be estimated from the pre-intervention distribution of the variables in the model, $P(Y,X,Z)$. The equivalence in (3) therefore solves the identification problem of causal inference (Koopmans, 1949; Pearl, 2009), since it allows the analyst to estimate post-intervention distributions purely based on passive pre-intervention observations without manipulating the treatment variable X directly (so-called *observational* causal inference).

It is important to note that the theoretical justification for the mapping in (3) comes from the structural causal model and is only valid under certain conditions. In Figure 1a, for example, there are no other influence factors than Z that are jointly affecting X and Y, and thus assessing the conditional distribution Y given X for each value of Z separately, i.e., $P(y|x)$, eliminates all spurious influence factors from the relationship. A corollary of the fact that the equivalence in (3) can only be established based on the SCM, however, is that causal effects are generally not estimable without a causal model. In fact, model-free causal inference is a theoretical impossibility. Solving the identification problem always requires ex-ante causal assumptions and can thus not be done in a purely data-driven fashion (Bareinboim et al., 2020).

tem, e.g., increase demand by adjusting the characteristics of the product portfolio, optimal decision making requires predicting the effect of an action, $do(x)$, and therefore causal knowledge.

Experimentally manipulating a variable and measuring the resulting effect on an outcome, e.g., in a randomized control trial (RCT) or A/B test (Thomke, 1998, 2020), in principle renders $P(y|do(x))$ directly observable. However, also the capacity to carry out experimental studies does not alleviate the need for a causal model (Deaton and Cartwright, 2018). Experiments are necessarily run at a specific point in time and within a particular population (e.g., in a laboratory, for a selected group of customers, or within a confined geographical area). That means that the analyst will need to adapt experimental results to different empirical settings in order to use them productively. Establishing whether, and under which conditions, causal knowledge is applicable in varying contexts is a problem known under the name of *transportability* in the causal inference literature; while the social sciences commonly refer to it as *external validity* (Bareinboim and Pearl, 2016). Solving this problem requires ex-ante causal assumptions about the DGP, even if they are as trivial as assuming – as it is commonly done – that experimental results can be extrapolated without explicitly taking domain heterogeneity into account (Pearl and Bareinboim, 2014).

Moreover, in many practical settings, directly intervening on a variable of interest is not feasible, either because it would be too costly, unethical or simply impractical to do so. In such cases, the analyst might need to rely on surrogate experiments, which manipulate a target variable only indirectly (Bareinboim and Pearl, 2012a).[7] In a social media context, for example, online advertisers who want to estimate the impact of a campaign cannot directly control clients' exposure to an ad (Gordon et al., 2019). Instead, however, consumers can be randomly assigned to both a treatment group, who will be shown the ad once they log on to the platform, and a control group, who will only see a neutral message. That way, the advertiser is able to effectively manipulate ad exposure, but this control will only be imperfect due to, e.g., consumers who never visit the platform during the field phase of the experiment. Thus, these customers never get exposed to the ad, even if they have been assigned to the treatment group – a problem called "one-sided noncompliance" in the literature (Imbens

[7]In economics and management research, surrogate experiments are commonly referred to as instrumental variables designs (Imbens and Angrist, 1994; Bascle, 2008).

and Rubin, 2015). Such kind of surrogate experiments can be tremendously helpful in learning about causal effects, but they require very specific assumptions in order to be informative (Bareinboim and Pearl, 2012a; Semadeni et al., 2014), which highlights once again the necessity of a model for obtaining causal knowledge, even in situations where experiments are in principal possible.

To summarize the preceding theoretical discussion, we argued that managers require causal knowledge to generate and evaluate alternative courses of strategic actions. Inferring causal effects itself thereby requires causal models that encode theoretical assumptions about the data generating process. Standard machine learning approaches, however, refrain from causal modeling, which makes them unsuitable for the task of causal inference. As this raises an important mismatch between the questions that are being pursued and the capacity of methods that are employed to answer them, the purpose of this study, as presented in the following, is to assess the practical implications of this mismatch and to explore the use of causal inference methods in contemporaneous organizations.

METHOD & ANALYSIS

Given the novelty of this research focus, a mixed methods research design, combining interviews with a survey instrument in an exploratory sequential design (Creswell and Plano Clark, 2018), was employed to derive a comprehensive understanding and corroboration of the topic (Johnson et al., 2007; Creswell, 2014; Bettis et al., 2014). Additionally, acknowledging the association of the topic with ongoing discussions within the data science and machine learning community, emergent blog posts, discussions and other relevant online resources were followed up on and integrated throughout the data collection and analysis phase.

Interviews were conducted with 15 data science practitioners to obtain a descriptive account and learn facts, experiences and understandings from individuals in key positions to comprehend the topic (Rowley, 2012; Vaughan, 2013; Aguinis and Solarino, 2019). The research setting and sample were thus selected for their suitability to reveal existing relationships and underlying phenomena and are thereby not representative of some general population but rather chosen such that they facilitate the generation of new theoretical insights (Eisenhardt and Graebner, 2007). To that regard, practitioners from the field of data science and machine learning were deemed as particularly suitable to provide practical insights to the research questions for two reasons. First, as the topic of causal machine learning is based in the computer science and economics literature, it is reasonable to assume that it diffuses to the industry primarily via data scientists and machine learning engineers. Second, as this study is interested in the role of causal inference for organizational decision making, the topic can best be investigated by drawing on the experience of data scientists working on data-augmented strategies in today's organizations. Interviewees were recruited via email, professional social networking and development platforms (e.g. Twitter, LinkedIn, Kaggle) and referrals within the community.

Table 1 provides profiles of all interview partners. Interviews were held in one consecutive round from September 2019 to May 2020 in the form of semi-structured interviews (Appendix A.1) to maintain flexibility towards the interview flow and encourage the interviewees to share experiences around the theme. Interviews took 30 to 45 minutes each and were conducted via video conferencing tools. Prior to the interview, participants were informed about the research project, procedures and the confidentiality of their responses (Rea and Parker, 2014). For means of analysis, the interviews were recorded and transcribed. To extract a holistic and descriptive account of the meaning of the textual material with respect to the research questions, the transcripts were analyzed using qualitative content analysis (Weber, 1990; Morris, 1994; Mayring, 2000). Primary content categories were therefore initially formulated based on the research focus and interview questions to determine the levels of abstraction for the subsequent inductive category development. The material was then open coded, extracting subcodes and additional categories emerging from the data, until theoretical saturation was reached. The final coding frame (Appendix A.2) consists of eleven

TABLE 1: Overview of interviews

ID	Company Alias	Role	Industry	Country
CONSI	IT & service integration consultancy	Data scientist	Consulting	GER
RETA1	Online fashion retail company	Senior applied (data) scientist	Retail & consumer goods	GER
TECH1	Software company	Research data scientist	Technology, media, telecommunications	USA
CONS2	Independent IT consultant	Consultant & software engineer	Consulting	GER
CONS3	IT Consulting Company	Data science consultant	Consulting	USA
MANU1	Automotive company	Data scientist	Industrial manufacturing	GER
TOUR1	Tech company in the travel industry	Chief technology officer	Hospitality & tourism	GER
TECH2	Tech company in consumer products marketing	Senior machine learning engineer	Technology, media, telecommunications	USA
MANU2	Engineering technology company	Senior vice president	Industrial manufacturing	GER
TECH3	Ridesharing company	Research data scientist	Technology, media, telecommunications	USA
CONS4	IT consulting company	Data science consultant	Consulting	GER
HEALTH1	Health care provider	Research data scientist	Health services	ISR
TOUR2	Online booking company	Data scientist	Hospitality & tourism	GER
TECH4	BDA Software Company	Senior applied (data) scientist	Technology, media, telecommunications	USA
TECH5	Communication platform	Data scientist & machine learning engineer	Technology, media, telecommunications	CAN

main categories, each with their own subcategories formulated out of the material.

Revealing important variables, clarifying relevant concepts and establishing a common terminology, the first eight interviews then provided the basis to inductively develop the survey instrument (Bryman, 2012; Creswell, 2014), which was administered as a web-based survey (Appendix A.3) in parallel with the second half of the interviews. Deriving from the preceding interviews, the target population of the survey was determined as all data scientists in organizations that emphasize big data and machine learning in their business. The respondents were taken as representatives of their field and their organization in particular. Potential respondents were recruited and contacted in the same way as interview partners. In total 342 responses were recorded, from which 108 were discarded as respondents did not go beyond the first page of general questions.[8] The majority of respondents (68.1%) who completed the survey are from the private sector. With 32.6%, the technology, media and telecommunications sector is most represented in the sample, followed by the education, research and public sector (16.7%) and the financial service sector (14.2%). Overall data and research scientists make up 62.3% of the respondents. The size of respondents' organizations is relatively equally distributed with 33.4% having 250 or less employees and 30.4% having 5000 or more employees. About one third (34.6%) of the organizations are younger than 10 years. Finally, most organizations (44.4%) are from Europe, closely followed by North American firms (40.6%). Detailed descriptive statistics can be found in Appendix A.4. Analysis of the survey responses in general focuses on descriptive figures to extend and validate findings from the semi-structured interviews.

RESULTS

The following presents interview and survey findings in parallel to allow for triangulation of results across cases as well as methods. The derivation of findings from different interviews

[8] This relatively large number of participants not responding to the main body of the survey is expected to primarily consist of academics who were curious about the content of the study but had no real interest in participating.

is presented transparently (codes in brackets) and supplemented with quantitative evidence from the primary survey data to provide validity and generalizability[9]. Where appropriate and conducive to the generation of richer theoretical insights, relevant online resources reviewed during the research process are incorporated. It should be noted that the last seven (of 15) interviews were performed in Spring 2020, during the Covid–19 pandemic. While this certainly needs to be considered an unexpected circumstance regarding the replicability of this study (Aguinis and Solarino, 2019), we expect that the situation did not significantly affect the insights generated. As responses were made retrospectively, based on past experiences, the Covid-19 pandemic was likely too recent at the time of the interviews to have had a significant impact on the interview and survey responses.

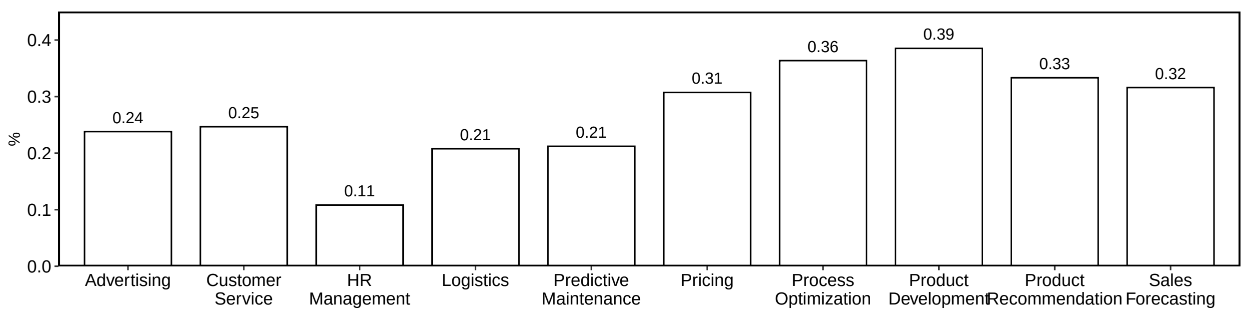
The types of questions firms address with their data science efforts

To obtain a more general understanding of the business problems practitioners are typically presented with and investigate whether practical use cases indeed involve causal inference components, we first explore the types of questions firms address with data science and machine learning. For 60% of the companies interviewed, in particular those offering web-based or software products, data science and machine learning is an integral part of their product (1.a.i), employed to ensure and optimize product functionality, including recommendation systems, automated pricing algorithms and price predictions in the case of, for instance, online marketplaces and meta search services. Accordingly, survey responses in Figure 2 also indicate that they employ data science for product recommendations (33%) and pricing (31%) on their platforms. Similarly, 33.3% of interviewees (1.a.iv) and 39% of survey respondents specifically mention the application of data science and machine learning to product (feature) development. Moreover, 53.3% of interviewees (1.a.ii) and 36% of survey participants indicate the applicability of data science to optimize business processes, such as the response time to a customer request or the scheduling of an airline operator. On a

[9]Additional figures not provided in the text can be found in Appendix A.7

related note, predictive maintenance is identified as a relevant area in the interviews (1.a.iii) and confirmed by 21% of survey respondents. Finally, a frequent application of data science and machine learning according to 53.3% of interviewees (1.a.v) and 32% of survey respondents, is forecasting of, primarily, sales but also demand and financial figures. Additionally, survey results identify customer service and advertising as important areas of application for 25% and 24% of the sample, respectively. Overall, data science and machine learning are mentioned as important inputs to managerial decision making by providing information on business parameters relevant to the decision situation (1.a.vi).

FIGURE 2: Use cases of data science applications in contemporaneous organizations



Note: multiple selection possible

When asked about the relevance of data science for strategic decisions in particular, all interviewees confirmed its importance and provided practical business cases (1.b). As TECH3 noted: *"One [of two ways in which data scientists contribute to the organization] is to help enable people make better decisions on a day-to-day basis. As leadership decides for instance what options to invest in or what products to launch, data scientists help inform those decisions."* From the practical examples offered by practitioners interviewed, five types of strategic applications can be synthesized. Most importantly, 86.6% of interviewees mention applications of data science that intend to understand the marketplace, such as customer segmentation, analyze revenue streams and customer churn or monitor and evaluate business variables to optimize for (1.b.ii). Other applications are in strategic planning, which includes market entry and exit decision and business model innovation (1.b.iii), pricing and revenue scheme decisions (1.b.iv), as well as product development (1.b.v) and investment decisions (1.b.vi). In support of these findings, 44% of survey respondents classify data science as highly important for strategic decision making in their organization.

An emergent theme during the interviews was that applications of data science and machine learning, and the business questions addressed, are to a large extent driven by the particular methods data scientists are familiar with and have at their disposal. As TECH5 noted: *"It's often rather, that they [executives] are faced with a business problem and some data scientist will come to them and present a toolkit to solve it."* The research directions specified by top management are thereby generally too broadly defined to recommend concrete actions to data scientists. *"It's interesting, with data and machine learning in general what I have observed is that there is very little top down actually."* (TECH5) Instead, analysts often resort to technical solutions and methodological approaches they are interested in or experienced with and focus on effectively applying them within the broader parameters of the business model. As CONS1 explained: *"we now often first see a particular technology that we want to use and later look for the right area of application."* The data analytical approach taken towards a particular business problem or (strategic) management question, is thus rarely decided by management but more so selected by data scientists commissioned with it. More generally, findings highlight the central role of data scientists in determining what the organization's analytical capacities are directed at, which implies that the approach taken towards strategic problems, and answers offered to that regard, in fact depend on, and are thus also limited to, the methods and capabilities available to data scientists in the organization.

Awareness of the difference between correlational and causal knowledge

To assess how far knowledge of causality is actually diffused among practitioners in the industry, we were interested in respondents' awareness of the topic of causal inference in the context of their work. All interviewees say that they know the conceptual difference between correlation and causation, which is also reflected by survey respondents who indicate familiarity with the distinction in 96.6% of the cases.[10] When asked what they associate with the phrase "correlation doesn't imply causation", 60% of practitioners recognize the limitations of their predictive models in determining causal mechanisms and the potential risk of, particularly actors in the broader organization, interpreting results of correlation-based analyses as causal relationships (2.a).

Indeed, despite this conceptual understanding of data scientists, the degree of recognition of causal inference in their professional work varies greatly across practitioners. At one end, 3 of 15 interviewees mention that causal inference is not at all considered in their projects (8.a.i). Two of the cases are firms offering consulting services (TECH4 and CONS3), whose data science efforts are found to be often restricted by the client's demands and decisions. The other case reflects the standpoint of an executive (MANU2), who appears to be inexperienced with the topic, a finding that seems to indicate that especially practitioners in management positions – typically the decision makers in organizations – do not recognize the applicability of causal inference in their daily work. The majority (60%) of interviewees however says that causal inference is actually beginning to slowly diffuse in their organization (8.a.ii) and 40% mention that they are new to the topic but highly interested in learning more (8.a.iii). Hence, overall, awareness of and interest in the topic of causal inference is growing in the industry. Diffusion thereby appears very much bottom-up (8.a.iv), driven by experts among data scientists: *"I think we rely on that small set of causal inference experts to inject their expertise wherever they can but it's very unevenly distributed.* TECH3 described. Similarly, TRAV1 noted: *"Together with one of our data scientists, I am the one who is currently pushing this topic. We are missing that view."* and CONS3 explained: *"In particular the data scientists are really aware of it [causal inference] and are following discussions and developments."*

[10]Results at this point are based purely on participants' own responses. We do not explicitly test whether the self-assessment of their understanding is correct i.e. whether respondents are in fact able to distinguish between correlation and causation

Importance of causal inference in business today

To explore more closely the practical use of causal methods in decision making, we further examine the importance and value that respondents attach to causal inference for practical business applications – particularly strategic decision situations. Highlighting the relevance of causal knowledge for their work (3), interviewees overall stress the importance of causal inference for addressing a variety of questions in the business context (4). Noting the relevance of causal inference for their data science projects, this observation is further confirmed by 47% of survey respondents. More generally, 87% of interviewees say that by identifying confounding variables and causal effects, causal inference allows firms to obtain a more thorough and robust model of their business environment (3.b). Causal knowledge thereby allows for more complete insights and understandings of the environment and wider applicability, generalizability and interpretability. CONS1 explained: *"The questions we deal with are generally larger than a specific context or a concrete data set. If I want my models to work in different scenarios, across data sets, I quickly arrive at such [causal inference] problems."*

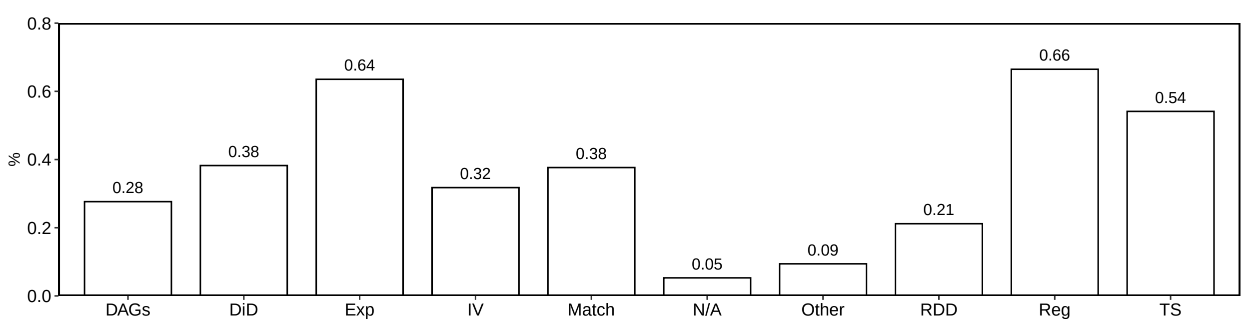
More practically, interviewees provide several exemplary business problems where they see potential for applying causal inference to better understand the relevant causal forces at play in the business environment (4.a) (see also Appendix A.6). As TECH2 described: *"One area where we are still interested in doing observational data science is in just fundamentally understanding the causes of a redemption in the app and having a true causal model of that phenomenon. That way we can apply interventions and ask questions like: What if we did X?"* Interviewees thus say that in practice causal inference is valuable to derive more robust predictions of business metrics (4.b), increase operational efficiency (4.c), solve particularly complex problems that require a more fundamental, generally applicable model of reality (4.d), and evaluate the performance of specific interventions such as product changes (4.e). *"Most of our experiments are about some feature change that we think will improve the product. [...] We just want to verify that it is an improvement and how much of an improvement it is."* TECH6 explained. The majority of respondents thereby illustrate the applicability of causal inference to managerial decision making in general (3.c) and specific situations of strategic choice about the long-term direction and scope of the firm such as product (roadmap) decisions, investment decisions, pricing, and prioritizing business objectives to optimize for (4.f). Stressing the importance of causal inference for strategic decision situations in particular, TOUR1 states that: *"Especially as a start-up, we need to manage our resources wisely, which actually links back to corporate strategy. [...] That's why understanding causes and bringing facts to the table when making these prioritization decisions, is really a key success factor that we believe in."*

Nonetheless, despite this evidence, survey results reveal that on average, practitioners (yet) find pure prediction to be more dominant in their data science projects and interviewees, too, state that most machine learning algorithms in practice are still mainly correlation based: *"At the moment, we are very often looking at correlations only, without investigating very much."* (TOUR1). *"All the classical machine learning algorithms rely on statistical correlation. They learn statistical correlations of input patterns. However, we know that this is not the whole truth."* (CONS3) Linking back to the observation that methods available to data scientists dictate the questions answered, the strict dominance of correlational methods indicated by interviewees (2.b), implies that causal questions are insufficiently answered or not addressed at all. Indeed, 73.3% of practitioners interviewed realize that the primarily correlational approaches in their work miss causal relationships (2.c). In practice this observation is critical, as CONS4 noted: *"what we see quite often is that when you are asked to do a data science project, the questions that the client asks actually can't be answered with the machine learning model you just trained with them. [...] My educated guess would be that the majority of questions in the end are causal questions, but the way we as data scientists have been trained to answer these questions is always in terms of classical machine learning."* From a retailing point of view, RET1 illustrated, that this means that they *"might be optimizing for something that doesn't really cause a change in [customer] behavior."* Emphasizing the importance of raising and resolving this mismatch in firms' data scientific approaches to strategic decisions, CONS4 in fact concluded: *"that while we can train machine learning to predict your outcome, it would be key to establish an understanding that there is still a gap between machine learning models and decision making."*

Diffusion of causal inference methods and techniques

To understand how contemporaneous organizations incorporate causal inference into their business practice, we explore the different techniques employed by firms and advantages and disadvantages practitioners identify to that regard. Survey results in Figure 3 show that practical causal inference techniques and approaches are unevenly distributed across organizations. The by far most prominent causal inference technique employed by firms today is experiments (5.h). Specifically, 73.3% of interviewees and 63.5% of survey respondents indicate that they apply A/B tests, bandits or reinforcement learning for causal inference at their organization. In contrast, besides regression (66%), observational causal inference techniques (i.e. based on ex-post data analysis without active manipulation or randomization) are less diffused. Those mentioned most often by interviewees and survey respondents respectively are difference-in-differences (33% (5.a), 38%), matching (26.6% (5.c), 38%) and instrumental variable estimation (20% (5.b), 32%). Distinguishing between experimental and observational causal inference techniques then, we can derive several findings about the use of causal inference methods and techniques in contemporary organizations' data science efforts.

FIGURE 3: Usage of causal inference methods in data science applications



Note: multiple selection possible

As insinuated above, experiments are the default causal inference technique for most data science practitioners (5.h.i). The greater majority of participants says that they regularly run a large number of A/B tests to find answers to their business questions. As TECH5 stressed: *”There’s a lot of causal inference techniques that we aren’t using, that we really could be here, but the massive hammer that tech companies swing around when it comes to causal inference is running experiments.”* In particular firms with web–based products thereby say they are continuously conducting a large number of experiments (5.h.ii). *”For actual causal inference in terms of impact, no matter what algorithm we develop, even if these algorithms are developed off of non-experimental data, we always run an A/B test. Every algorithm that we ever develop, at its final stage will go through A/B testing.”* TECH2 noted. This finding, is well in line with recent discussions in the literature, reflecting that awareness and relevance of experimental methods in the business domain is growing (Thomke, 2020; Bojinov et al., 2020; King, 2020). The popularity of experiments in the business context thereby appears to primarily stem from the fact that they are relatively easy to use, as TECH1 explained: *”A/B tests have been around for much longer and they have a much better understanding and support within the broader organization. [...] In that sense, I think almost everyone believes in the power of randomized experiments and they are increasingly becoming a part of decision making whenever they are possible.”* Survey respondents confirm this observation, indicating that ease of application and straightforward understanding, as they seemingly do not require specific causal modeling, are the most important advantages of experiments (Figure 4).

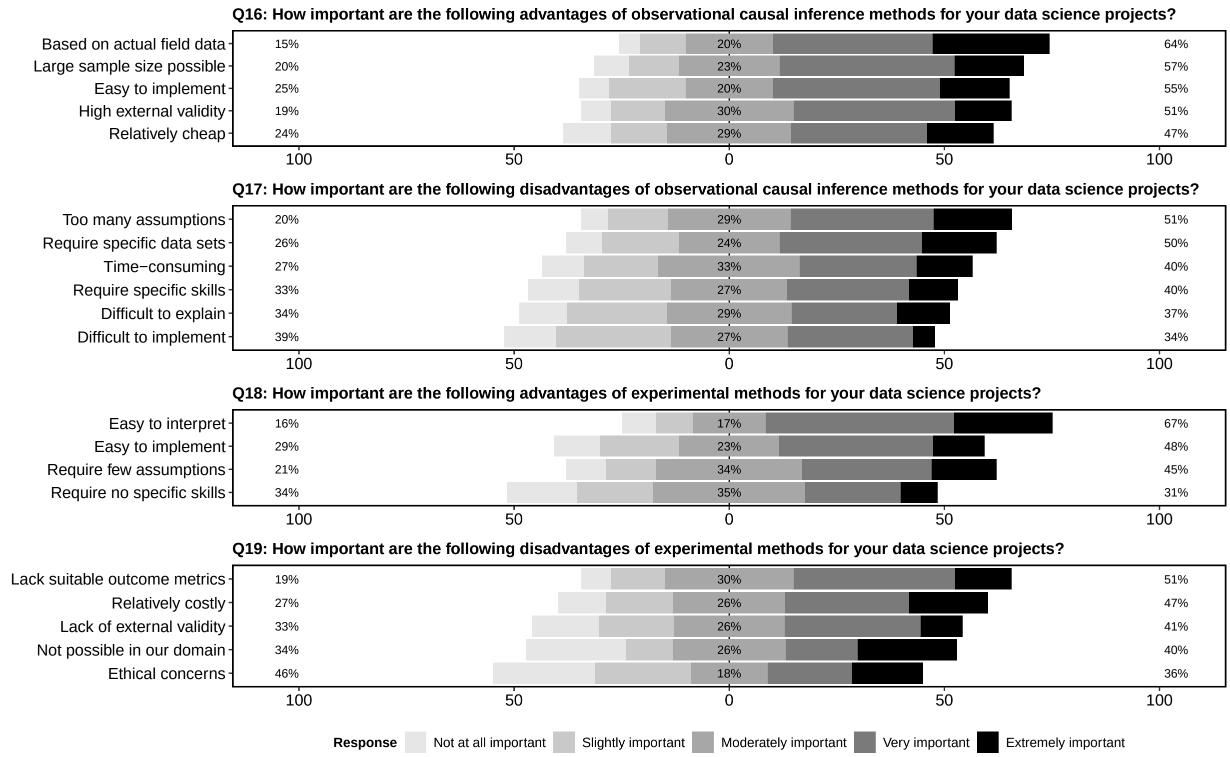
However, while experiments appear as the preferred choice to answer causal queries in the business context, several drawbacks (Figure 4), amongst others reflecting the A/B testing pitfalls identified by Bojinov et al. (2020), render them impractical in many problem spaces. To that regard, 73% of interviewees mention difficulties concerning the practical applicability (6.a) of A/B tests, relating to situations in which the business environment, data availability or parameters of interest are unsuitable for an experimental approach. Concretely, practitioners say that experiments are often time consuming, rendering them impractical for pressing business decisions, control and treatment cannot be sufficiently administered in certain situations (as it might not be possible to exclude people from the treatment) or the data collectible with the experiment are not suitable to the business question. Practitioners interviewed further explain that experiments cannot easily and safely be set up in, for instance, manufacturing or medical environments and 40% of survey respondents say that experiments are not possible at all in their domain. Additional ethical and legal concerns about providing divergent products, services or prices to different customers, present an important shortcoming to 27% of interviewees (6.a.i) and 36% of survey participants. In 40% of the cases, interviewees thereby associate high costs with providing inferior user experience and forgoing profits (6.b) under experimental designs, which is confirmed by 47% of survey respondents who see experiments as relatively costly. Moreover, interviewees identify particular technical shortcomings (6.c) that inhibit the reliability of the measured effects, whereby the lack of suitable outcome metrics appears to be the most important drawback of experimental methods, as indicated by 51% of survey respondents. Specifically, the outcomes of interest decision makers would like to affect are often difficult to observe, requiring analysts to rely on proxy metrics. In many settings, experiments are only able to measure a short-term metric, while the metric of interest is in fact long-term: *"A lot of those tests don't quite answer the questions that we have. Some of our tests might only run for two weeks, while we actually care about a long-term effect like for instance six months because that's the business relevant estimand."* TECH3 stated.

Finally, 40% of interviewees and 41% of survey respondents mention external validity, i.e., transportability of results to different circumstances, as a big concern (6.d). In practice, this becomes relevant when experimental results obtained in one (e.g. geographic) market are supposed to be used in another. The generalizability and applicability of experimental results as well as the value of data collected is thus limited under experimental approaches to causal inference. While the analysis shows that experiments are valued for their ease of implementation and interpretation, external validity presents an important obstacle. Considering, for instance, significant changes to a product, TECH2 explained that *"there, external validity becomes a concern. When we run previous experimental data and build models off of these experiments, we get concerned when the app drastically changes."* Interestingly, the global Covid-19 pandemic during some of the research phase provides another particularly vivid example of the external validity problem, as TECH3 described: *"External validity has indeed recently become a very important topic for us. With the Covid-19 pandemic, we are actually worried that experimental results from today won't extrapolate to the future as the marketplaces are quite different."* This finding is also reflected in broader discussions among data scientists and machine learning engineers. Particularly, the global pandemic in this case presents a drastic change in the world, altering peoples' behavior and consumption patterns. As a consequence, firms cannot know whether experimental results collected before the pandemic are still valid during or after it (Microsoft, 2020). In the realm of experimental causal inference methods, practitioners would have to rerun the experiment, as TECH2 confirmed, which potentially entails substantial resource investments.

Given those drawbacks of experimental methods (Figure 4) and the advantages practitioners perceive with respect to observational causal inference methods, such as difference-in-differences, DAGs or matching (Figure 3), some practitioners identify observational methods as relevant alternatives or complements to experiments, when those are not (fully) feasible (3.d). Indeed, survey results reveal that respondents especially value observational causal inference methods for being based on actual field data (64%), their ease of implementation (55%) and the high external validity of results (51%). TECH3 said: *"I think A/B tests tend to estimate the policy relevant estimand quite well for us when they work. So, they are the most desirable method. However, we often don't get the right answer from them, so we have to use something else. I do think having a system-wide causal understanding is something that we try to achieve."*

Nonetheless, data scientists in the industry yet also identify particular obstacles to using observational approaches (Figure 4). The biggest challenge thereby appears to be the under-

FIGURE 4: Observational versus experimental causal inference methods in data science applications



standability and applicability of methods to practical business cases (7.b) as practitioners perceive them as more complex than standard machine learning techniques and experiments. Specifically, interviewees remark that the analysis requires numerous untestable assumptions which need to be based on a good causal model of the business environment. Likewise, 51% of survey respondents view available methods as being based on too many assumptions. As such models are derived from expert domain knowledge, observational causal inference methods lack an objective standard of model evaluation, increasing the complexity of applying such methods to practical business problems. *"Since we are not randomizing, we can never be sure that we have not missed confounding variables."* TECH1 noted. To that regard, survey results highlight that observational causal inference methods are indeed seen as difficult to implement (34%) and explain (37%). More practically, interviewees further posit that analyses with observational methods entail lengthy and expensive deployment efforts, making them impractical in fast–moving business environments (7.a), which is confirmed by survey respondents who find such methods to be time–consuming (40%) and to require very particular skills (40%). Observational causal inference methods are consequently not as readily employed for analysis as *"it is often not easy to describe the direct benefit of causal inference [to customers]."* (CONS1) and non-data-scientists in management. TECH1 explained: *"Even if it's valid it's much harder to convince a businessperson on the basis of such a complicated analysis."*. From the interviews it can therefore be derived that industry examples could decisively help with this lack of understanding and applicability of causal inference in diverse business contexts. *"We are really missing experience and especially practical examples illustrating how causal inference can be applied to different areas, not only drug testing and the like."* TOUR1 stated. More generally, diffusion of practical causal inference tools and software libraries also appear to be in their infancy, as 34.5% of survey participants do not apply practical causal inference tools. As Table 2 suggests, a majority of libraries have only been published in mid-2019, with three industrial open source libraries leading in engagement on GitHub. Similarly, 46.6% of interviewees (7.e) are not aware of methods, practical applications of such methods or are not using any external tools but develop their own. To that regard, experts reveal important shortcomings (7.d) of existing software, which is actually found to be immature by some practitioners (7.d.iii). Specifically, interviewees say that the user experience is not very interactive and still very theoretical, tools lack graphical ability to make the models more understandable and offer some user interpretable what-if scenarios (7.d.ii), and more generally, the right, standard tools (especially in Python) are not yet available (7.d.i). Survey results confirm that many practitioners, although willing, perceive an important hurdle to adopting observational causal inference methods, indicating that only 27% find existing tools and software suitable for their purposes.

TABLE 2: Software libraries on causal inference (January 10, 2021, examples). *Contributors* and *Stars* refer to repository statistics at GitHub

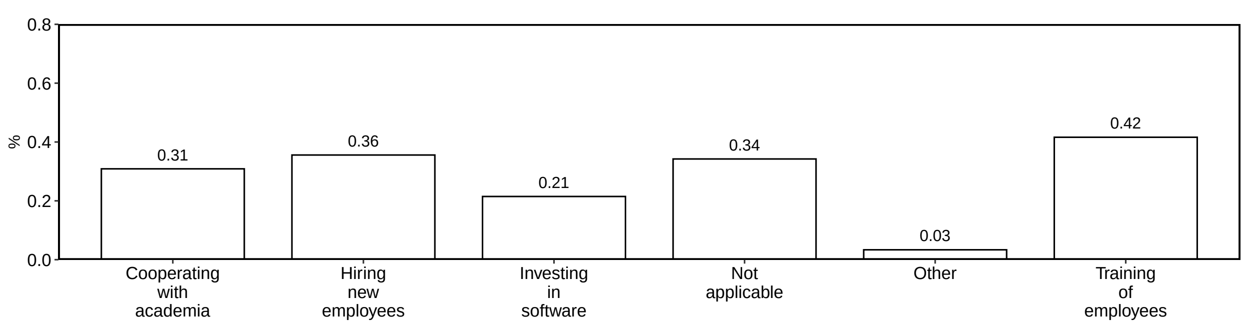
Entity	Library	Release	Source	Contrib.	Stars	Language
Microsoft	DoWhy	July 15, 2019	Public, GitHub	34	2,579	Python
Uber	CausalML	July 10, 2019	Public, GitHub	21	1,540	Python
Google	Causal Impact	August 2, 2014	Public, GitHub	7	1,209	R
Academic	ggdag	October 9, 2019	Public, CRAN	1	303	R
Academic	dagitty	August 26, 2016	Public, CRAN	5	129	R
IBM	Causal Inference	July 12, 2019	Public, GitHub	3	105	Python
Netflix XP	Causal Models	April 29, 2019	Private, Inhouse	NA	NA	Python/R

What is the future of causal inference in industry?

As the theory and results presented up to this point suggest causal inference to be an emerging topic among data scientists in the industry, the following focuses on deriving a future outlook of where the industry is moving towards with respect to causal inference. While the benefits of data science and machine learning in general are widely acknowledged across respondents (9.b), especially in terms of their informative value (business intelligence) and analytical capacity, a number of practitioners casts doubts about their benefits with respect to (strategic) decision making (9.a). Survey results show that while most practitioners (78%) are positive about the impact of data science on decision making in their company, 22% are not sure or even disagree completely. Yet the application of causal inference in particular is seen to have the potential to overcome some of the shortcomings of machine learning with respect to strategic decision making (9.c): *"I think causal inference does [improve human decision making]. It has the potential to improve decision making, at least more than machine learning does."* TECH1 said. Similarly, TECH5 noted: *"I worry that since it's all correlation rather than causation, it's unclear to which extent we are making great decisions based on that. [...] I would feel a lot better if we could narrow it down to some sort of causality instead of just correlation."* and CONS2 stated: *"From my experience, I think very few people are questioning the data generation process itself and the story behind the data and I think there* *is some value to be generated."*

Similarly, practitioners surveyed are slightly more positive, indicating that the majority of practitioners (83%) believes that moving to causal inference and making decisions on the basis of causal models could add considerable value to decision making in their organization. Accordingly, we document an overall willingness of 60% of interviewees and 45% (n=155) of survey respondents to invest into causal inference at their organization. Interviewees thereby intend to adopt additional causal inference techniques (10.b) and develop own, potentially open-source, solutions (10.c). Alternatively, 40% of interviewees (10.d) and 42% of survey respondents (Figure 5) want to train existing employees (in particular data scientists) more intensively in causal inference. *"We try to level people up by teaching and providing lots of best-practice and examples."* (TECH3). Additionally, 36% of practitioners surveyed plan to expand their team's capabilities by hiring suitable talent. To that regard, TECH3 specified that such causal inference experts come *"almost invariably from economics."* (TECH3). Likewise, survey results indicate that, together with computer science and statistics, economics is considered as one of the most important educational backgrounds of employees for improving causal inference capabilities in organizations today – a finding that reflects well an ongoing trend in the tech sector to increasingly hire economists for data science jobs due to their specific skill set (Athey and Luca, 2019).

FIGURE 5: Means of improving organizational causal inference skills and capabilities in the future



Note: multiple selection possible

Generally, as evidence points to an emerging trend of increased recognition and application of causal inference in organizations' data science efforts, the study also reveals four challenges that still need to be overcome for this trend to fully unfold (8.b). First, industry examples of practical causal inference applications are yet largely missing for many business sectors. Interviewees say that such industry leadership would help practitioners in adopting causal inference methods by showing where and how to apply causal techniques specifically to their business (8.b.iii), as TOUR1 explained: *"Uber, for instance, has a behavioral data science team, but it's not so common and that's why it's hard for us to make the decision of whether it is worth investing into it."* Second, practical causal inference methods need to become more accessible to practitioners, which relates to the lack of awareness and availability of applicable, standardized tools (8.b.iv). Only 27% of practitioners surveyed find existing causal inference software packages fitting for their purposes, which renders applying causal inference to business problems relatively expensive and time consuming.

Third, identifying a lack of respective skills within their organizations (8.b.ii), respondents reveal that a broader understanding of causal inference and its applicability is still necessary, but at the same time difficult to achieve because of the complexity of the topic. As TECH2 emphasized: *"There still is a big educational gap, even with professionals in higher-up positions"* that needs to be closed for causal inference to be applied more broadly to business decisions. To that regard, interviewees stress the need to overcome important structural challenges (8.b.i) in contemporary organizations that yet obstruct such broader diffusion. Specifically, this challenge refers to a lack of established processes, missing incentive structures, missing training and the pressure on data scientists to deliver fast results in practical business environments, which considerably limit data science approaches employed: *"When you build models, people always ask for the end product. It's often really only about getting stuff out the door, even if it's not right or even perfect, but if its good enough and making some impact, you go with it"* (TECH2). Given the finding that the broader organization yet often lacks training and involvement in causal inference, this top-down pressure on data scientists to deliver fast results implies that causal inference approaches are often not explored let alone exploited in approaching a business problem. *"We don't currently have processes to actually get to the root cause driving a particular phenomenon and that's why we are interested in how we can establish this kind of thinking"*, TOUR1 noted. Ultimately, results thus highlight the need for skill development, better integration of methodological and business knowledge, and establishing suitable processes in order for causal data science approaches to improve organizational decision making.

DISCUSSION

The main research question addressed in this study is of epistemological nature. Strategic and organizational decision making entails the choice between different courses of action (Simon, 1964), which in turn depends on causal knowledge in order to predict the likely outcomes of the managerial initiatives under consideration. Due to path-breaking technological progress in the last decade, machine learning and artificial intelligence offer the prospects of becoming an increasingly important input for optimized decision making in modern organizations (Brynjolfsson and McElheran, 2019). However, as long as learning about individual cause-and-effect relationships is the goal, this requires the use of an adequate methodology. Recent advances in the causal inference literature have shaped our understanding of the kind of knowledge that can be obtained based on different types of data inputs (Pearl and Mackenzie, 2018; Bareinboim et al., 2020; Hünermund and Bareinboim, 2019). In particular, it has been understood that any data-scientific method necessarily adheres to an epistemological hierarchy – called the *ladder of causation* – which stipulates that lower-layer (correlational) information almost always underdertermines information at higher (causal) layers of the hierarchy. Under certain circumstances it becomes possible to bridge these layers of the hierarchy and infer causal relationships from passive observations alone. However, that requires the data analyst to invoke untestable theoretical assumptions about the data generating process in form of a causal model. A fact which was eloquently summarized by

Cartwright (1989) with the maxim: *"no causes in, no causes out"*.

These causal models need to be based on theory. They originate from an organization's accumulated knowledge and shared beliefs about its mode of value creation and the business environment it is operating in. Decision making, as long as it relies on accurate predictions of cause-and-effect, can therefore never be purely *data-driven*. Information based on passive observations of an unperturbed environment is rarely rich enough to inform about strategic courses of action at a sufficient level of granularity. The need for a causal model in order to interpret and contextualize empirical patterns also does not disappear when the decision-maker has the possibility to directly intervene in the environment she is observing, e.g., via A/B tests or reinforcement learning algorithms (Thomke, 2020; Forney et al., 2017). The problem of transportability (or external validity) remains if experimental results ought to be used in contexts (e.g., temporal or geographical) that differ even just slightly from the ones they have been obtained in. Solving this problem requires to bring in ex-ante theoretical knowledge that is not yet already in the data themselves.

The most commonly used machine learning tools today refrain from making explicit assumptions about the data generating process and are thus unsuitable for the task of causal inference (Pearl, 2019; Mullainathan and Spiess, 2017). Traditionally, their objective is to maximize out-of-sample fit in a hold-out sample, which seemingly provides an objective standard of evaluation. Causal inference methods, by contrast, with their requirement to incorporate expert domain knowledge, are perceived to be more elusive. Different causal assumptions might lead to substantially different conclusions, which adds a layer of subjectivity to the analysis. As one of our interview partners expressed: *"The biggest challenge [with causal inference] is that you don't believe the results. Unlike predictive projects it's very hard to validate your results. [...] That's the primary issue."* At the same time, there is a growing recognition among data science practitioners that there is no way around this challenge. In the words of another respondent: *"We often use predictive models for making decisions. However, that is increasingly not the right thing to do, which is something that* *not just our organization, but many organizations are still coming forward with."*

Our empirical analysis documents a shift in the data science and machine learning community, which starts to recognize the importance of causal inference for practical business decision making. Our interview partners indicate a rising frustration with standard methods, as their correlational approach is increasingly perceived to be not well aligned with the organizational goals practitioners try to achieve. This development is still at its beginning, however, and knowledge about causal inference methods needs to be diffused more widely outside of a relatively small group of specialists. Our empirical findings indicate that a majority of respondents plans to invest more into causal inference capabilities in the future. Main channels thereby constitute training measures as well as hiring of new employees with educational backgrounds from statistics, computer science, and economics, who can contribute the required methodological skills to the organization (Athey and Luca, 2019).

Several of our interview partners further expressed their opinion that this kind of retooling, away from a purely correlation-based framework, will be a major trend in the data science community for the years to come. Examples of causal inference initiatives in major tech firms illustrate where the industry is heading. For instance, already today the video streaming platform Netflix employs causal inference methods in recommendation systems (Raimond, 2018) and rigorously runs experiments for any product change considered before it becomes a default component in the user experience (Urban et al., 2016). Likewise, the online lodging marketplace Airbnb utilizes various experimentation techniques to test product changes and continuously learn from developments in the market place (de Luna, 2018). The American ride-hailing company Uber dedicates increasing resources to establishing causal inference approaches as a means to improve their user experience (Harinen and Li, 2019). And tech giant Google is concerned with assessing the effectiveness of online advertising campaigns in (causally) affecting search-related site visits (Brodersen et al., 2015; Varian, 2016).

Theoretically, our discussion of the fundamental challenge of causal inference corroborates and contributes to recent findings from the literature on the theory-based view of the firm (Felin and Zenger, 2009, 2017; Felin et al., 2020a,b). Accurate predictions of the outcomes of future actions, which is instrumental for effective organizational decision making and strategic foresight, requires managers to build theories. Simple data-driven approaches relying on readily observable evidence and performance feedback in business experimentation alone are not sufficient for inferring value-creating strategies (Felin and Zenger, 2009). Theories allow managers to put empirical findings into context, develop the necessary cross-sight for identifying undervalued strategic resources, and imagine new courses of action based on scattered evidence (Felin et al., 2020b). The importance of theories, which constitute *"abstract, causal representation[s] of the world"* (Felin and Zenger, 2017, 262), as an input for causal learning once again highlights the truth value of *"no causes in, no causes out"*

At the same time, the origin of viable theories within the theory-based view, and their relations to empirical evidence and experimentation, is still underresearched (Felin et al., 2020a; Gavetti and Menon, 2016). The literature on causal inference in the field of machine learning and AI clarifies the interplay between theory and data for causal learning and offers a powerful inferential machine that managers can use in order to gain strategic foresight.[11] In particular, it demonstrates the kind of theoretical assumptions that are necessary for bridging the layers of the ladder of causation and establishing a mapping between correlation and causation (Pearl, 2019; Bareinboim et al., 2020). As a guiding principle, this becomes especially valuable when the minimum level of assumptions required to obtain practically relevant causal knowledge can be determined (Peters et al., 2017). Furthermore, the causal AI literature specifies what kind of data needs to be collected and which business experiments have to be performed in order to inform theory (Hünermund and Bareinboim, 2019). It offers remedies if data is only imperfect, due to limited perception and selective observation (Bareinboim and Pearl, 2012b; Bareinboim and Tian, 2015; Correa et al., 2019). And it

[11] For an overview and introduction into the causal inferential framework, the following resources are well suited: Athey and Imbens (2017); Bareinboim and Pearl (2016); Hünermund and Bareinboim (2019); Pearl (2009); Peters et al. (2017); Pearl and Mackenzie (2018); Pearl et al. (2016).

proposes tools for managers transport insights between various different contexts (Pearl and Bareinboim, 2011; Bareinboim and Pearl, 2012c; Pearl and Bareinboim, 2014; Lee et al., 2020), which is a necessary ingredient for effective theorizing.

The necessity of theoretical causal modeling in order to establish an information transfer across the layers of the causal hierarchy, in effect, stresses the importance of domain experts in integrating causal inference into data science and constitutes a substantial opportunity for human-machine cooperation. Indeed, the role of managers as domain experts is seen as critical in leveraging existing decision making algorithms. As the bi-directionality of human and machine decisions poses challenges to organizations leveraging decision making algorithms (Shrestha et al., 2019), the question of how human sensemaking and machine learning can work together to improve the generation of insights from business analytics pertains (Sharma et al., 2014). Highlighting the process model of task input (data: sound, text, images, and numbers), task processes (algorithms), and task outputs (solutions and decisions), von Krogh (2018) argues that human problem solvers themselves need to engage in sensemaking and interpretation of the prediction output offered by algorithms, to connect needs, problems, and alternative solutions. Similarly, Athey (2018) concludes that automated prediction algorithms do not leave domain experts out of the loop. Concerns about the identifiability of causal effects, the confounders measured in a particular setting, selecting the right outcome variables or deriving accurate strategies from (causal) relationships do remain. In light of this theoretical evidence, causal modeling could in fact play a crucial role in the ”automation-augmentation paradox” (Raisch and Krakowski, 2020) artificial intelligence poses to the management domain, and thus address concerns about the future role of managers under machine intelligence. Because of their characteristic that crucial assumptions on which conclusions rest need to be made explicit ex ante, causal AI methods are also able to make an important contribution towards solving the problem of explainability and potential fairness concerns that plague existing approaches to automated decision making (Shrestha et al., 2021; Zhang and Bareinboim, 2018).

Hence, as data has become a strategic resource (Hartmann and Henkel, 2020), this study argues that causal inference might emerge as an important organizational capability. However, to fully develop it, top management, as the originator of value-creating theories, and data science teams, as the business unit with the relevant technical expertise, will need to work more closely together than they currently do. The results of our interviews indicate that to this day data analytics, although appreciated for its general utility in business intelligence (Shrestha et al., 2021), is still not integrated well in the organizational strategy formulation process. Too often, probably, top management sees machine learning competencies as *"nice to have"* but not essential for decision making, while data scientists focus on implementation and methodological aspects without leveraging the full potential of the contextual business knowledge that is embedded in the wider organization. Effective cooperation is hampered by communication barriers between the two groups of specialists who speak different languages and adopt different institutional logics (Dunn and Jones, 2010; Besharov and Smith, 2014). Thus, there is a need for interdisciplinarily trained individuals who can act as boundary spanners between both domains (Argote et al., 2003; Gittelman and Kogut, 2003). This growing demand for combining deep business knowledge and strong data analytic skills will thereby likewise affect business school education, which needs to include more training in machine learning and causal inference methodology, also outside of specialized programs, in order to develop the kind of holistic competencies that are necessary for effective data-augmented decision making.

References

Agrawal, A. 2018. The economics of artificial intelligence. *McKinsey Quarterly*.

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

Agrawal, A., Gans, J. S., and Goldfarb, A. 2019. Artificial intelligence: the ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, 33(2): 31–50.

Aguinis, H. and Solarino, A. M. 2019. Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal*, 40(8): 1291–1315.

Argote, L., McEvily, B., and Reagans, R. 2003. Introduction to the speical issue on managing knowledge in organizations: Creating, reatining, and transferring knowledge. *Management Science*, 49(4): 5–8.

Asatiani, A., Malo, P., Nagbøl, P. R., and Penttinen, E. 2020. Challenges of explaining the behavior of black-box ai systems. *MIS Quarterly Executive*, 19(4): 15.

Ascarza, E. 2018. Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 55(1): 80–98.

Athey, S. 2017. Beyond prediction: Using big data for policy problems. *Science*, 355(6324): 483–485.

Athey, S. 2018. The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*, 507–547. University of Chicago Press.

Athey, S. and Imbens, G. W. 2017. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2): 3–32.

Athey, S. and Imbens, G. W. 2019. Machine learning methods that economists should know about. *Annual Review of Economics*, 11: 685–725.

Athey, S. and Luca, M. 2019. Economists (and economics) in tech companies. *Journal of Economic Perspectives*, 33(1): 209–30.

Axelrod, R. 1976. *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton, New Jersey: Princeton University Press.

Baden-Fuller, C. and Mangematin, V. 2013. Business models: A challenging agenda. *Strategic Organization*, 11(4): 418–427.

Baden-Fuller, C. and Morgan, M. S. 2010. Business models as models. *Long Range Planning*, 43(2-3): 156–171.

Baer, M., Dirks, K. T., and Nickerson, J. A. 2013. Microfoundations of strategic problem formulation. *Strategic Management Journal*, 34(2): 197–214.

Bajari, P., Chernozhukov, V., Hortaçsu, A., and Suzuki, J. 2018. The impact of big data on firm performance: An empirical investigation. In *AEA Papers and Proceedings*, volume 109, 33–37.

Balasubramanian, N., Ye, Y., and Xu, M. 2020. Substituting human decision-making with machine learning: Implications for organizational learning. *Academy of Management Review*. forthcoming.

Bareinboim, E., Correa, J. D., Ibeling, D., and Icard, T. 2020. On Pearl's hierarchy and the foundations of causal inference. Technical Report R-60, Columbia University. Forthcoming in ACM special volume in honor of Judea Pearl.

Bareinboim, E. and Pearl, J. 2012a. Causal inference by surrogate experiments: Z-identifiability. In *28th Conference on Uncertainty in Articial Intelligence*, 113–120.

Bareinboim, E. and Pearl, J. 2012b. Controlling selection bias in causal inference. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, 100–108.

Bareinboim, E. and Pearl, J. 2012c. Transportability of causal effects: Completeness results. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Bareinboim, E. and Pearl, J. 2016. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27): 7345–7352.

Bareinboim, E. and Tian, J. 2015. Recovering causal effects from selection bias. In Koenig, S. and Bonet, B.(Eds.), *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, Palo Alto, CA. Association for the Advancement of Artificial Intelligence, AAAI Press.

Bascle, G. 2008. Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, 6(3): 285–327.

Bazeley, P. 2008. Mixed methods in management research. *Dictionary of Qualitative Management Research*, 133–136.

Bertsimas, D. and Kallus, N. 2016. The power and limits of predictive approaches to observational-data-driven optimization. *arXiv preprint arXiv:1605.02347*.

Besharov, M. L. and Smith, W. K. 2014. Multiple institutional logics in organizations: Explaining their varied nature and implications. *Academy of Management Review*, 39(3): 364–381.

Bettis, R. A., Gambardella, A., Helfat, C., and Mitchell, W. 2014. Editorial: Qualitative empirical research in strategic management. *Strategic Management Journal*, 36(4): 637–639.

Bharadwaj, A. S., Bharadwaj, S. G., and Konsynski, B. R. 1999. Information technology effects on firm performance as measured by tobin's q. *Management Science*, 45(7): 1008–1024.

Bingham, C. B. and Eisenhardt, K. M. 2011. Rational heuristics: The 'simple rules' that strategists learn from process experience. *Strategic Management Journal*, 32(13): 1437–1464.

Blei, D. M. and Smyth, P. 2017. Science and data science. *Proceedings of the National Academy of Sciences*, 114(33): 8689–8692.

Bloom, N., Sadun, R., and Van Reenen, J. 2012. Americans do it better: Us multinationals and the productivity miracle. *American Economic Review*, 102(1): 167–201.

Bojinov, I. I., Sait-Jacques, G., and Tingley, M. 2020. Avoid the pitfalls of a/b testing. *Harvard Business Review*, 2(March-April): 48–53.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. 2015. Inferring causal impact using bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1): 247–274.

Bryman, A. 2006. Integrating quantitative and qualitative research: How is it done? *Qualitative research*, 6(1): 97–113.

Bryman, A. 2012. *Social Research Methods*, volume 4. Oxford University Press: Oxford.

Brynjolfsson, E., Hitt, L. M., and Kim, H. H. 2011. Strength in numbers: How does data-driven decisionmaking affect firm performance? *Available at SSRN 1819486*.

Brynjolfsson, E. and McAfee, A. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. WW Norton & Company.

Brynjolfsson, E. and McElheran, K. 2016. The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5): 133–39.

Brynjolfsson, E. and McElheran, K. 2019. Data in action: Data-driven decision making and predictive analytics in us manufacturing. *Available at SSRN 3422397*.

Camuffo, A., Cordova, A., Gambardella, A., and Spina, C. 2020. A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Science*, 66(2): 564–586.

Cartwright, N. 1989. *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.

Cartwright, N. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge University Press.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., and Mullainathan, S. 2016. Productivity and selection of human capital with machine learning. *American Economic Review*, 106(5): 124–27.

Chen, H., Chiang, R. H., and Storey, V. C. 2012. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4).

Chesbrough, H. and Rosenbloom, R. S. 2002. The role of the business model in capturing value from innovation: Evidence from xerox corporation's technology spin-off companies. *Industrial and Corporate Change*, 11(3): 529–555.

Cho, J. Y. and Lee, E.-H. 2014. Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences. *The Qualitative Report*, 19(32): 1–20.

Choudhury, P., Starr, E., and Agarwal, R. 2020. Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*.

Christensen, C. M., Hall, T., Dillon, K., and Duncan, D. S. 2016. Know your customers'"jobs to be done". *Harvard Business Review*, 94(9): 54–62.

Correa, J. D., Tian, J., and Bareinboim, E. 2019. Identification of causal effects in the presence of selection bias. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*.

Creswell, J. W. 2014. *A Concise Introduction to Mixed Methods Research*. SAGE Publications, Inc.

Creswell, J. W. and Plano Clark, V. 2018. *Designing and Conducting Mixed Methods Research. Third*. SAGE Publications, Inc.

Cyert, R. M., Simon, H. A., and Trow, D. B. 1956. Observation of a business decision. *The Journal of Business*, 29(4): 237–248.

Davenport, T. H. and Harris, J. G. 2009. What people want (and how to predict it). *MIT Sloan Management Review*, 50(2): 22.

Davenport, T. H. and Ronanki, R. 2018. Artificial intelligence for the real world. *Harvard Business Review*, 96(1): 108–116.

De Leeuw, E. D., Hox, J., and Dillmann, D. 2008. *International Handbook of Survey Methodology*. Routledge, 1st edition.

de Luna, B. 2018. Experimentation and measurement for search engine optimization. *Retrieved from https://medium.com/airbnb-engineering/experimentationmeasurement-for-search-engine-optimization-b64136629760*.

Deaton, A. and Cartwright, N. 2018. Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210: 2–21.

Donaldson, D. and Storeygard, A. 2016. The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4): 171–98.

Dunn, M. B. and Jones, C. 2010. Institutional logics and institutional pluralism: The contestation of care and science logics in medical education 1967–2005. *Administrative Science Quarterly*, 55.

Durand, R. and Vaara, E. 2009. Causation, counterfactuals, and competitive advantage. *Strategic Management Journal*, 30(12): 1245–1264.

Eisenhardt, K. M. and Graebner, M. E. 2007. Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, 50(1): 25–32.

Eisenhardt, K. M. and Zbaracki, M. J. 1992. Strategic decision making. *Strategic Management Journal*, 13(S2): 17–37.

Fedyk, A. 2016. How to tell if machine learning can solve your business problem. *Harvard Business Review*, 11: 2–4.

Felin, T., Gambardella, A., Stern, S., and Zenger, T. 2020a. Lean startup and the business model: Experimentation revisited. *Long Range Planning*, 53(4).

Felin, T., Gambardella, A., and Zenger, T. 2020b. Value lab: A tool for entrepreneurial strategy. *Management & Business Review*. forthcoming.

Felin, T. and Zenger, T. R. 2009. Entrepreneurs as theorists: On the origins of collective beliefs and novel strategies. *Strategic Entrepreneurship Journal*, 3(2): 127–146.

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

Forney, A., Pearl, J., and Bareinboim, E. 2017. Counterfactual data-fusion for online reinforcement learners. In *Proceedings of the 34th International Conference on Machine Learning*.

Furnari, S. 2015. A Cognitive Mapping Approach to Business Models: Representing Causal Structures and Mechanisms, 207–239. Emerald Group Publishing Limited.

Gary, M. S. and Wood, R. E. 2011. Mental models, decision rules, and performance heterogeneity. *Strategic Management Journal*, 32(6): 569–594.

Gavetti, G. and Levinthal, D. 2000. Looking forward and looking backward: Cognitive and experiential search. *Administrative Science Quarterly*, 45(1): 113–137.

Gavetti, G. and Menon, A. 2016. Evolution cum agency: Toward a model of strategic foresight. *Strategy Science*, 1(3): 207–233.

Ghasemaghaei, M. and Calic, G. 2020. Assessing the impact of big data on firm innovation performance: Big data is not always better data. *Journal of Business Research*, 108: 147–162.

Gillon, K., Brynjolfsson, E., Griffin, J., Gupta, M., and Mithas, S. 2012. Panel-business analytics: Radical shift or incremental change. In *Proceedings of the 32nd International Conference on Information Systems (16-19 December)*.

Gittelman, M. and Kogut, B. 2003. Does good science lead to valuable knowledge? biotechnology firms and the evolutionary logic of citation patterns. *Management Science*, 49(4): 366–382.

Goldfarb, A., Taska, B., and Teodoridis, F. 2020. Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings. http://dx.doi.org/10.2139/ssrn.3468822.

Gordon, B. R., Zettelmeyer, F., Bhargava, N., and Chapsky, D. 2019. A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science*, 38(2): 193–225.

Greene, J. C., Caracelli, V. J., and Graham, W. F. 1989. Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3): 255–274.

Guthrie, J., Petty, R., Yongvanich, K., and Ricceri, F. 2004. Using content analysis as a research method to inquire into intellectual capital reporting. *Journal of Intellectual Capital*, 282–293.

Haavelmo, T. 1943. The statistical implications of a system of simultaneous equations. *Econometrica, Journal of the Econometric Society*, 1–12.

Harinen, T. and Li, B. 2019. Using causal inference to improve the uber user experience. *Retrieved from https://eng.uber.com/causal-inference-at-uber/*.

Hartford, J., Lewis, G., Leyton-Brown, K., and Taddy, M. 2016. Counterfactual prediction with deep instrumental variables networks. *arXiv preprint arXiv:1612.09596*.

Hartmann, P. and Henkel, J. 2020. The rise of corporate science in ai: Data as a strategic resource. *Academy of Management Discoveries*, 6(3): 359–381.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

Heckman, J. J. and Vytlacil, E. J. 2007. Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. *Handbook of Econometrics*, 6: 4779–4874.

Henderson, J. V., Storeygard, A., and Weil, D. N. 2012. Measuring economic growth from outer space. *American Economic Review*, 102(2): 994–1028.

Hodgkinson, G. P., Bown, N. J., Maule, A. J., Glaister, K. W., and Pearman, A. D. 1999. Breaking the frame: An analysis of strategic cognition and decision making under uncertainty. *Strategic Management Journal*, 20(10): 977–985.

Hünermund, P. and Bareinboim, E. 2019. Causal inference and data-fusion in econometrics. *arXiv preprint arXiv:1912.09104*.

Iansiti, M. and Lakhani, K. R. 2020. Competing in the age of ai. *Harvard Business Review*, 61–67.

Imbens, G. W. and Angrist, J. 1994. Estimation and identification of local average treatment effects. *Econometrica*, 62: 467–475.

Imbens, G. W. and Rubin, D. B. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Johnson, R. B., Onwuegbuzie, A. J., and Turner, L. A. 2007. Toward a definition of mixed methods research. *Journal of mixed methods research*, 1(2): 112–133.

Kaminski, J. C. and Hopp, C. 2019. Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals. *Small Business Economics*, 1–23.

King, J. 2020. "the power of thesetechniques is only getting stronger"-a conversation with pinterest's jeremy king. *Harvard Business Review*, 16–17.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. 2017. Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1): 237–293.

Koopmans, T. C. 1949. Identification problems in economic model construction. *Econometrica, Journal of the Econometric Society*, 125–144.

Kriauciunas, A., Parmigiani, A., and Rivera-Santos, M. 2011. Leaving our comfort zone: Integrating established practices with unique adaptations to conduct survey-based strategy research in nontraditional contexts. *Strategic Management Journal*, 32(9): 994–1010.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., and Kruschwitz, N. 2011. Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2): 21–32.

Lee, S., Correa, J. D., and Bareinboim, E. 2020. General transportability – synthesizing observations and experiments from heterogeneous domains. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*.

Lu, C. and Du, R. Y. 2020. Click-through behavior across devices in paid search advertising: Why users favor top paid search ads and are sensitive to ad position change. *Journal of Advertising Research*.

Lycett, M. 2013. 'datafication': Making sense of (big) data in a complex world. *European Journal of Information Systems*, 22(4): 381–386.

Malone, T. W. 2018. *Superminds: The surprising power of people and computers thinking together*. Little, Brown Spark.

Maule, J. A., Hodgkinson, G. P., and Bown, N. J. 2003. Cognitive mapping of causal reasoning in strategic decision making. In Hardman, D. and Macchi, L.(Eds.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making*, book 13, 253–272. West Sussex, England: Wiley Online Library.

Mayring, P. 2000. Qualitative content analysis. *A Companion to Qualitative Research*, 1: 159–176.

Menzies, P. 2006. Making things happen: A theory of causal explanation by james woodward. *Mind*, 459: 821–826.

Mintzberg, H., Raisinghani, D., and Theoret, A. 1976. The structure of" unstructured" decision processes. *Administrative Science Quarterly*, 21: 246–275.

Mitchell, J. R., Shepherd, D. A., and Sharfman, M. P. 2011. Erratic strategic decisions: When and why managers are inconsistent in strategic decision making. *Strategic Management*

Journal, 32(7): 683–704.

Mithas, S., Ramasubbu, N., and Sambamurthy, V. 2011. How information management capability influences firm performance. *MIS Quarterly*, 35(1): 237.

Morris, R. 1994. Computerized content analysis in management research: A demonstration of advantages & limitations. *Journal of Management*, 20(4): 903–931.

Mullainathan, S. and Spiess, J. 2017. Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2): 87–106.

Nickerson, J. A. and Zenger, T. R. 2004. A knowledge-based theory of the firm — the problem-solving perspective. *Organization Science*, 15(6): 617–632.

Pearl, J. 1988. Probabilistic reasoning in intelligent systems. *San Mateo, CA: Kaufmann*, 23: 33–34.

Pearl, J. 1995. Causal diagrams for empirical research. *Biometrika*, 82(4): 669–688.

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, 2nd edition.

Pearl, J. 2019. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3): 54–60.

Pearl, J. and Bareinboim, E. 2011. Transportability of causal and statistical relations: A formal approach. In *25th AAAI Conference on Artificial Intelligence*, 247–254. Menlo Park, CA: AAAI Press.

Pearl, J. and Bareinboim, E. 2014. External validity: From do-calculus to transportability across populations. *Statistical Science*, 579–595.

Pearl, J., Glymour, M., and Jewell, N. P. 2016. *Causal Inference in Statistics: A Primer.* West Sussex, United Kingdom: John Wiley & Sons Ltd.

Pearl, J. and Mackenzie, D. 2018. *The Book of Why: The New Science of Cause and Effect.* New York, NY: Basic Books.

Peters, J., Janzing, D., and Schölkopf, B. 2017. *Elements of Causal Inference.* The MIT Press.

Rahmati, P., Tafti, A. R., Westland, J. C., and Hidalgo, C. 2020. When all products are digital: Complexity and intangible value in the ecosystem of digitizing firms. *Forthcoming, MIS Quarterly.*

Rai, A. 2020. Explainable ai: From black box to glass box. *Journal of the Academy of Marketing Science*, 48(1): 137–141.

Raimond, Y. 2018. Time, context and causality in recommender systems. Presented October 2018 at France is AI, Paris, France.

Raisch, S. and Krakowski, S. 2020. Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review.*

Rea, L. M. and Parker, R. A. 2014. *Designing and Conducting Survey Research: A Comprehensive Guide.* John Wiley & Sons.

Rowley, J. 2012. Conducting research interviews. *Management Research Review.*

Schwenk, C. R. 1984. Cognitive simplification processes in strategic decision-making. *Strategic Management Journal*, 5(2): 111–128.

Semadeni, M., Withers, M. C., and Trevis Certo, S. 2014. The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. *Strategic Management Journal*, 35(7): 1070–1079.

Sharma, R., Mithas, S., and Kankanhalli, A. 2014. Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, 23(4): 433–441.

Shiffrin, R. M. 2016. Drawing causal inference from big data. *Proceedings of the National Academy of Sciences*, 113(27): 7308–7309.

Shivakumar, R. 2014. How to tell which decisions are strategic. *California Management Review*, 56(3): 78–97.

Shrestha, Y. R., Ben-Menahem, S. M., and von Krogh, G. 2019. Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4): 66–83.

Shrestha, Y. R., Krishna, V., and von Krogh, G. 2021. Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research*, 123: 588–603.

Shrivastava, P. and Grant, J. H. 1985. Empirically derived models of strategic decision-making processes. *Strategic Management Journal*, 6(2): 97–113.

Simon, H. A. 1964. On the concept of organizational goal. *Administrative Science Quarterly*, 9: 1–22.

Strauss, A. and Corbin, J. 1990. *Basics of Qualitative Research*. London, UK: SAGE Publications.

Strotz, R. H. and Wold, H. O. 1960. Recursive vs. nonrecursive systems: An attempt at synthesis (part i of a triptych on causal chain systems). *Econometrica: Journal of the Econometric Society*, 417–427.

Stubbart, C. I. 1989. Managerial cognition: A missing link in strategic management research. *Journal of Management Studies*, 26(4): 325–347.

Thomke, S. H. 1998. Managing experimentation in the design of new products. *Management Science*, 44(6): 743–762.

Thomke, S. H. 2020. Building a culture of experimentation. *Harvard Business Review*.

Tidhar, R. and Eisenhardt, K. M. 2020. Get rich or die trying… finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal*, 41(7): 1245–1273.

Urban, S., Sreenivasan, R., and Kannan, V. 2016. It's all a/bout testing: The netflixexperimentation platform. *Retrieved from: https://medium.com/netflix-techblog/its-alla-bout-testing-the-netflix-experimentation-platform-4e1ca458c15*.

Varian, H. R. 2014. Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2): 3–28.

Varian, H. R. 2016. Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences*, 113(27): 7310–7315.

Vaughan, S. 2013. Elite and elite-lite interviewing: Managing our industrial legacy. *Researching Sustainability: A Guide to Social Science Methods, Practice and Engagement, Earthscan, Abingdon*, 105–119.

Vega, R. P., Anderson, A. J., and Kaplan, S. A. 2015. A within-person examination of the effects of telework. *Journal of Business and Psychology*, 30(2): 313–323.

Vera-Muñoz, S. C., Shackell, M., and Buehner, M. 2007. Accountants' usage of causal business models in the presence of benchmark data: A note. *Contemporary Accounting Research*, 24(3): 1015–1038.

Vinyals, O., Babuschkin, I., ..., ., and Silver, D. 2019. AlphaStar: Mastering the Real-Time Strategy Game StarCraft II. https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/.

von Krogh, G. 2018. Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*.

Waldmann, M. R. 1996. Knowledge-based causal induction. *Psychology of Learning and Motivation*, 34: 47–88.

Walsh, J. P. 1995. Managerial and organizational cognition: Notes from a trip down memory lane. *Organization Science*, 6(3): 280–321.

Weber, R. P. 1990. *Basic Content Analysis*, volume 2. SAGE Publications, Inc.

Woodward, J. F. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford, UK: Oxford University Press.

Wu, L., Hitt, L., and Lou, B. 2020. Data analytics, innovation, and firm productivity. *Management Science*, 66(5): 2017–2039.

Zhang, J. and Bareinboim, E. 2018. Fairness in decision-making – the causal explanation formula. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*.

A APPENDIX

A.1 Interview guide

1. What role do data science and machine learning play in your organization?
    - What are typical questions you are trying to answer?
    - Can you tell us a couple of examples (setting the stage for later is important here)?
    - Is data science also relevant for corporate strategy in your organization?
2. What do you associate with the phrase “correlation doesn’t imply causation”?
    - How would you define causal inference?
3. Does causal inference play a big role in your data science projects?
    - How do you make sure to model causal?
    - How do you make sure not to model correlational?
4. What are typical causal questions you are trying to answer in your organization?
    - Can you give examples of a typical project?
    - What tools do you use in order to answer them?
    - Is this relevant for corporate strategy too?
5. What are typical prediction problems you are dealing with?
    - What tools do you use in order to answer them?
    - What are the biggest challenges?
    - How do you deal with uncertainty?
6. Which causal inference methods are currently known to you?
    - What is your most used approach?
    - Do you know about other approaches?
    - What are the biggest shortcomings of current causal inference methods you see in practice?
7. Do you have the perception that there are different methodological camps when it comes to causal inference?
    - Rubin / Imbens / Athey versus Bareinboim / Pearl?
8. Which software tools and environments do you work with?
    - Software?
    - Which libraries are you using / planning to use for causal inference?
    - How did you take notice of these software solutions?
    - Are existing tools / libraries suitable for your purposes?
    - Do you plan to contribute own open-source solutions?
9. Do you run experiments (A/B testing, reinforcement learning, etc.)?
    - In which domains do you use experiments?
    - What are the shortcomings of experiments, in your opinion?
    - If you face the choice between experiments and observational data analysis, how do you decide which method to use?
    - How do you make sure that experimental results remain valid also in other contexts? External validity?
10. Does your organization currently hire data scientists? What skills are you looking for in particular?

    - Which majors (CS, econ, math, etc.) do you mostly hire for data science jobs?
    - How is your team composed?
    - Is everyone on the team aware of the difference between causality and correlation?
    - Do you plan to invest more into your causal inference capabilities in the future?
11. Do you have the feeling that machine learning improves human decision-making in your organization?
    - What about causal ML in particular?
12. Which question do you think we should have asked but haven't in this interview?

A.2 Coding frame

ID	Subcode	Definition	Freq.
1	**Data science efforts**		**15**
1.a	Data science application	Areas of application and problems addressed with firms data science efforts in general.	15
1.a.i	Product functionality / improvements	Machine learning is part of the product and thus data science is employed to ensure functionality and improve the product. (e.g. recommendation engines; pricing algorithms)	9
1.a.ii	Process optimization		8
1.a.iii	Predictive maintenance		2
1.a.iv	Product development	Identifying and testing product and feature innovations (incl. ad systems).	5
1.a.v	Forecasting	Forecasting of business (decision) relevant parameters.	8
1.a.vi	Decision making	Provision of relevant data for decision making in general.	7
1.b	Data science & strategic questions		15
1.b.i	Important	Data and / or data driven decision making is mentioned to be important to corporate strategy / strategic decisions.	8
1.b.ii	Monitor/ understand marketplace	Data science is employed to monitor and understand the market place e.g. segment customers; identify high from low value customers; monitor & evaluate KPIs, customer churn or revenue streams.	13
1.b.iii	Strategic planning	Decisions concerned with strategic planning such as market entry and exit, market scoping, business model innovation.	6
1.b.iv	Pricing & revenue scheme	Inform and optimize (potentially automate) pricing.	5
1.b.v	Product decisions	Inform decisions on which products to launch; in which feature innovations to invest into; designing a product roadmap.	6
1.b.vi	Investment decisions	Inform decision on (a) financial investment and (b) time & (human) resource investment.	7
2	**Difference correlation & causation**	**Meaning of the phrase "correlation does not imply causation".**	**14**
2.a	Awareness	Awareness that correlational approaches used provide only limited insights as they do not reveal causal relationships and thus ought not be interpreted as such.	9
2.b	Dominance correlation	Correlation is dominant in data science efforts.	9
2.c	Miss causal effect	Practitioners say that correlational approaches they employ miss causal effects and so results do not represent the whole truth.	11

ID	Subcode	Definition	Freq.
3	**Relevance causal inference**	**Relevance of causal inference in practitioners' work environment.**	**14**
3.a	Social relevance	Relevance of causal inference for society at large (understanding organization as economic actor in society and considering effect of actions).	2
3.b	Model of environment	Causal inference (tools) allow firms to obtain a more complete, robust and generalizable model of the respective business environment by identifying important confounding variables and causal effects.	13
3.c	Decision making	Causal tools are relevant for making important (high investment, high value-creating, high risk, limited resources) business decisions by providing important decision making aid: identify spurious correlation; derive action alternatives; estimate effect of interventions and evaluate strategic action alternatives.	10
3.d	Experiment alternative	Firms recognize and employ causal tools as alternatives for experimental methods (when those are not feasible).	6
4	**Causal questions and problems**	**Causal questions and problems that arise in practitioner's work.**	**15**
4.a	Model business environment	Employ causal inference to model the business environment to understand the drivers of observed phenomena (identify variables of interest) in the business environment.	11
4.b	(Robust) Forecasting	Employ causal inference to make more robust (long run) predictions of diverse metrics (incl. predictive maintenance).	5
4.c	Process optimization	Employ causal inference to increase operational efficiency (reduce response time; make tools easier to use; develop / improve standard procedures for high value leads; error analysis).	4
4.d	Address complex problems	Employ causal inference to address particularly complex problems in the respective business context.	6
4.e	Performance evaluation	Employ causal inference to evaluate performance of specific interventions (often product changes, new features) with regards to relevant business metrics and check if the intervention has the desired outcome in the business environment.	8
4.f	Inform strategic choices	Employ causal inference to inform strategic decisions: product feature decision; pricing; investment decisions; inventory/ product choices; KPI selection.	11

ID	Subcode	Definition	Freq.
5	**Causal methods and tools**		**12**
5.a	Difference-in-differences		5
5.b	Instrumental variable		3
5.c	Matching		4
5.d	Regression discontinuity		4
5.e	Directed acyclic graphs (DAGs)		3
5.f	None		2
5.g	Internally built tools		5
5.h	Experiments		11
5.h.i	Default	Experiments are the default causal inference method.	6
5.h.ii	Multiple & continuous	Organization runs multiple and continuous experiments.	5
5.i	Generalizing / test validity	Run multiple tests on the same data set, randomize treatment and control group to validate results and check whether they generalize.	2
5.j	Causal segmentation methods		1
5.k	Inverse probability weighting		1
5.l	Covariate adjustment		1
5.m	Synthetic control methods		1
5.n	Time split design		1
5.o	Knowledge graphs		1
6	**Shortcomings experiments**	**Shortcomings of experimental approaches identified in practice.**	**12**
6.a	Practical application	Experiments are unpractical in the business environment; with data available or parameters of interest.	11
6.a.i	Social / legal reasons	Experiments cannot be run for social or legal reasons: discriminating customer groups in an A/B test; charging different prices for the same product; unethical experiments.	3
6.b	Costs	Experiments entail rel. high costs: profits forgone; inferior user experience (i.e. customer loss).	6
6.c	Technical shortcomings	Experiments have several technical shortcomings: non-stationarity; unsuitable proxy for outcome metric; the risk of self selecting into experiments that are feasible/easier; biased control group; novelty effect.	7
6.d	External validity	Experiments have low/no external validity i.e. limited transportability of results to different circumstances: seasonality, different markets (e.g. countries), different customer groups, drastic product changes.	6

ID	Subcode	Definition	Freq.
7	**Shortcomings observational causal methods**	**Shortcomings of observational causal inference methods identified.**	**12**
7.	Practicality	Observational causal inference methods are seen as impractical due to time and cost it takes to run them, develop the expertise for them or deploy packages in own infrastructure.	7
7.b	Understandability & applicability	Observational causal inference methods are relatively complex (compared to standard statistical techniques) as they require numerous untestable assumptions, thus their applicability is not clear and methods and results are difficult to understand.	11
7.c	Technical shortcomings	Technical shortcomings of current observational causal inference tools.	6
7.d	Software shortcomings	Shortcomings of observational causal inference methods in terms of the software available.	8
7.d.i	Availability	The right tools (in the right environments) are not available.	4
7.d.ii	Usability	Software and user experience lacks usability and features that allow more user friendly application.	3
7.d.iii	Maturity	Observational causal inference methods are underdeveloped.	3
7.e	Diffusion	Practitioners are not aware of methods; their practical applicability and means to use tools or are not using any external models or tools but develop their own.	7
8	**Diffusion of causal inference**	**Diffusion of causal inference (as a topic and corresponding techniques).**	**0**
8.a	In organization	Diffusion of causal inference within organizations.	15
8.a.i	Not relevant	Causal inference is not relevant in practitioners' organizations.	3
8.a.ii	Beginning	Discussion about and application of causal inference is only at the beginning and slowly diffusing into the wider organization.	9
8.a.iii	Interested in learning more	Participants are fairly new to the topic but interested in learning more.	8
8.a.iv	Bottom-up	The diffusion of causal inference is bottom up in organizations, meaning that mainly data scientists; machine learning experts & researchers are investigating and pushing the topic.	10
8.a.v	Methodological debate	Participants report on the methodological debate regarding causal inference in their organization. (Rubin versus Pearl)	5

ID	Subcode	Definition	Freq.
8.b	Challenges	Challenges to diffusion and more wide-scale adoption of the causal discussion and (observational) causal inference techniques.	12
8.b.i	Structural	Structural challenges to wider diffusion e.g. the lack of established processes; missing experts/ knowledge; no demand (from client side); missing incentive structures; missing education in universities	9
8.b.ii	Educational gap	Educational/ knowledge gap w.r.t. causal inference, within the industry, the organization and even data science community.	10
8.b.iii	Missing examples	Missing practical examples for business problems/industries to illustrate where and how to apply causal techniques.	4
8.b.iv	Accessibility	Lack of availability and awareness of applicable methods and tools.	6
8.c	In Industry	Diffusion of causal inference in the business world more generally.	9
8.c.i	Not diffused in industry		7
8.c.ii	Beginning to diffuse		5
9	**Strategic decisions**	**How data science and machine learning affect strategic decision making in organizations.**	**15**
9.a	Doubts	Doubts about whether machine learning helps to make better decisions.	9
9.a.i	Context	In specific contexts machine learning is not (perceived as) helpful to decision making or complicates the business decision and justification.	5
9.a.ii	Window-dressing	Machine learning is perceived as a means to justify predetermined managerial decisions.	2
9.b	Improve/ facilitate decisions	Data science and machine learning facilitate or even improve strategic decision making.	14
9.b.i	Informative value	Data science improves decision making by preparing, visualising and analysing data to enable humans to make decisions.	7
9.b.ii	Smarter products	Machine learning affects corporate strategy by making products smarter.	3
9.b.iii	Analytical performance	Machine learning improves decision making by providing advanced analytical capacities.	6
9.c	Causal inference & decision making	Causal inference in particular improves/has the power to improve strategic decision making.	8
10	**Causal inference investments**	**Plans to invest more into firms' causal inference capabilities in the future.**	**5**
10.a	Not decided	Not (yet) determined on investing into causal inference in the future.	3
10.b	Adopt causal methods/ tools	Adopting causal methods and tools currently not applied.	4
10.c	Develop (open-source) solutions		5
10.d	Training	Train incoming talent and existing employees in causal inference.	6
10.e	Hiring	Hire suitable talent (by educational background)	8
10.e.i	Social sciences		5
10.e.ii	Mathematics		1
10.e.iii	Computer sciences		3
10.e.iv	Statistics		2

ID	Subcode	Definition	Freq.
11	**Technology**	**Libraries and software environments used by practitioners.**	**11**
11.a	Software environments		11
	Apache Airflow		1
	Apache Hive		3
	Amazon SageMaker		1
	BigQuery		1
	DAGitty		1
	Google Cloud		1
	Google Docs/Sheets		2
	Hadoop		2
	Java		1
	Julia		1
	Jupyter Notebook		4
	Kafka		1
	Mode		1
	Presto		2
	Python		9
	R		7
	Scala		1
	Spark		3
	SQL		6
	Stan		1
	Tableau		1
11.b	Libraries & packages		8
	Causal Forests		1
	DoWhy		3
	EconML		2
	Matchit		1
	Pandas		2
	PySpark		2
	PyTorch		1
	scikit-learn		4
	Sparkml		1
	TensorFlow		2

A.3 Survey questionnaire

Dear participant,
Thank you for taking the time to respond to this research questionnaire! The survey will take about 10 minutes to complete.

The aim of this research: We are interested in the role causal inference plays in a business context, the types of questions practitioners attempt to answer with their data-science efforts, and what kind of tools they apply to inform important business decisions.

Data use: The information provided by you will be treated strictly confidential. As participants in this questionnaire you will have access to the final results (scientific paper, executive summary). The results will be presented in statistical form and will not contain references to individual cases.

Thank you for your support!

Consent
I understand the information given above and agree to participate in this study under these terms.

	Selection
Yes	
No	

General information

Q1: What is your current job affiliation?

	Selection
Academic institution (university, publicly funded research institute, etc.)	
Private sector	
Both	
Other (please specify)	

Q2: What industry does your organization (primarily) operate in?

	Selection
Energy, utilities and resources	
Financial services	
Health services	
Hospitality & tourism	
Industrial manufacturing	
Pharma and life sciences	
Public sector, education, and research	
Real estate	
Retail and consumer goods	
Technology, media, and telecommunications	
Other (please specify)	

Q3: What is your primary role in your organization?

	Selection
Top-executive (CEO, CFO, COO)	
Product manager	
Research scientist	
Data scientist	
Software engineer	
Machine learning engineer	
Consultant	

Q4: How large is your organization (in FTE)?

	Selection
1–250 employees	
251–500 employees	
01–1,000 employees	
1,001–5,000 employees	
5,001–10,000 employees	
10,000+ employees	

Q5: How old is your organization?

	Selection
$<$10 years	
$>$ 10 years	

Q6: Where is your organization based?

	Selection
North America	
South / Latin America	
Europe	
Asia / Pacific	
Middle East	
Africa	

Data science in your organization

Q7: How important is data science in your business?

	Selection
Not at all important	
Slightly important	
Moderately important	
Very important	
Extremely important	

Q8: What problems do you usually address with your data science efforts? (select all applicable categories)

	Selection
Pricing	
Sales forecasting	
Product development	
Advertising	
Customer service	
Process optimization	
Human resource management	
Logistics	
Predictive maintenance	
Product recommendations	
Not applicable	
Other (please specify)	

Q9: How important is data science for strategic decision-making in your organization?
In the context of this survey, strategic decisions refer to management decisions that entail a considerable resource commitment and significantly determine the long-term direction and goals of an organization. Amongst others, resource investment, market or pricing decisions would typically fall into this category.

	Selection
Not at all important	
Slightly important	
Moderately important	
Very important	
Extremely important	

<u>Correlation versus causation</u>

Q10: Do you know the difference between correlation and causation?

	Selection
Yes	
No	
Not sure	

<u>Causal inference in data science</u>

Q11: How important is causal inference in your data science projects?
In the context of this survey, causal inference methods refer to all statistical and data-science methods that are suitable for uncovering a causal relationships between two (or more) variables. They stand in contrast to pure prediction problems, which are solely based on the correlation between two (or more) variables.

	Selection
Not at all important	
Slightly important	
Moderately important	
Very important	
Extremely important	

Q12: Do you find pure prediction or causal inference more important for your data science projects?

-5	*slider item*	5

Q13: To what extent do you agree with the following statement: "In our organization we have the necessary skills and capabilities for causal inference"?

	Selection
Strongly disagree	
Somewhat disagree	
Neither agree nor disagree	
Somewhat agree	
Strongly agree	

Causal inference methods

Q14: Which of the following causal inference methods do you use in your organization? (select all applicable categories)

	Selection
Directed acyclic graphs (DAG)	
Experiments (A/B testing, reinforcement learning)	
Instrumental variable estimation	
Matching	
Regression	
Regression discontinuity designst	
Time series methods	
Not applicable	
Other (please specify)	

Q15: Do you find observational or experimental causal inference methods more important for your data science projects?
Observational methods = Based on ex-post observed data (quasi-experimental methods, DAGs, causal modeling, etc.); Experimental methods = A/B testing, reinforcement learning, etc.

-5	*slider item*	5

Q16: How important are the following advantages of observational causal inference methods for your data science projects?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Easy to implement					
Relatively cheap					
Large sample size possible					
High external validity					
Based on actual field data					

Q17: How important are the following disadvantages of observational causal inference methods for your data science projects?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Require specific skills					
Time-consuming					
Require specific data sets					
Difficult to explain					
Based on too many assumptions					
Difficult to implement					

Q18: How important are the following advantages of experimental methods for your data science projects?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Easy to implement					
Require few assumptions					
Easy to interpret					
Require no specific skills					

Q19: How important are the following disadvantages of experimental methods for your data science projects?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Relatively costly					
Lack of external validity					
Ethical concerns regarding experiments					
Lack of suitable outcome metrics					
Not possible in our domain					

Software tools and packages

Q20: Which software environments do you mainly use in your data science projects?

	Selection
Python	
R	
SPSS	
SAS	
Julia	
Stata	
Matlab	
Excel	
Other (please sepcify)	

Q21: Which causal inference tools / software libraries are you aware of?

	Selection
causaleffect (R)	
Causal Impact (R)	
CausalML (Python)	
DAGitty (R)	
DoWhy (Python)	
EconML (Python)	
ggdag (R)	
pcalg (R)	
Not applicable	
Other (please sepcify)	

Q22: How suitable do you find existing causal inference tools / software libraries for your purposes?

	Selection
Not at all suitable	
Slightly suitable	
Moderately suitable	
Very suitable	
Extremely suitable	

The future of causal inference in your organization

Q23: Does your organization plan to invest in its causal inference skills and capabilities in the future?

	Selection
Yes	
No	
Not sure	

Q24: If so, how does your organization plan to improve its causal inference skills and capabilities?

	Selection
Training of existing employees	
Hiring of new employee	
Investing in our software architecture	
Cooperating with academic experts	
Not applicable	
Other (please specify)	
Matlab	
Excel	
Other (please sepcify)	

Q25: How important are the following disciplines / educational backgrounds of employees for improving the causal inference capabilities of your organization?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Computer science					
Mathematics					
Economics sets					
Statistics					
Social sciences					
Natural sciences					
Engineering					
Psychology					

Data-driven decision-making

Q27: Do you think that data science is improving human decision making in your organization?

	Selection
Definitely not	
Probably not	
Might or might not	
Probably yes	
Definitely yes	

Q28: To what extent do you agree with the following statement: "Causal inference methods will become more important for data-driven decision making in the future"?

	Selection
Strongly disagree	
Somewhat disagree	
Neither agree nor disagree	
Somewhat agree	
Strongly agree	

A.4 Descriptives

TABLE 3: Characteristics of survey respondents ($n = 234$)

	Frequency	Proportion (excl. NA)
Job Affiliation		
Academic institution	50	21.5%
Private sector	159	68.2%
Both	13	5.6%
Other	11	4.7%
NA	1	
Industry		
Energy, utilities and resources	13	5.6%
Financial services	33	14.1%
Health services	11	4.7%
Hospitality & tourism	3	1.3%
Industrial manufacturing	8	3.4%
Pharma and life sciences	6	2.6%
Public sector, education, and research	39	16.7%
Real estate	0	0.0%
Retail and consumer goods	23	9.8%
Technology, media, and telecommunications	76	32.5%
Other	22	9.4%
Role		
Top executive	16	6.9%
Product manager	10	4.3%
Research scientist	57	24.5%
Data scientist	88	37.8%
Software engineer	11	4.7%
Machine learning engineer	13	5.6%
Consultant	13	5.6%
Other	25	10.7%
NA	1	
Size		
1–250 employees	78	33.5%
251–500 employees	20	8.6%
501–1,000 employees	18	7.7%
1,001–5,000 employees	46	19.7%
5,001–10,000 employees	19	8.2%
10,000+ employees	52	22.3%
NA	1	
Age		
<10 years	81	34.6%
>10 years	153	65.4%
Region		
North America	95	40.6%
South / Latin America	10	4.3%
Europe	104	44.4%
Asia / Pacific	22	9.4%
Middle East	2	0.9%
Africa	1	0.4%

A.5 Technical appendix

To address the research purpose of this study in sufficient depth and breadth, data collection, analysis and interpretation followed a mixed methods design (Johnson et al., 2007; Creswell, 2014). As the topic is characterized in particular by its novelty and timeliness, the primary purpose of this empirical approach is to obtain a comprehensive understanding and corroboration of the topic. Thus, data collection and analysis encompassed three elements. First, multiple interviews with elite informants were conducted to explore the topic qualitatively and triangulate findings across cases. Second, following the exploratory sequential design proposed by Creswell and Plano Clark (2018), a survey instrument was developed from the interviews, to investigate the topic at a larger scale and test whether the qualitative results generalize (Greene et al., 1989; Bryman, 2006). Additionally, acknowledging the association of the topic with ongoing discussions within the data science and machine learning community, emergent blog posts, discussions and other relevant online resources were followed up on and integrated throughout the data collection and analysis phase. As this study intends to serve both, academics and practitioners, this research design in particular is expected to make the study accessible and useful to diverse stakeholder (Bazeley, 2008; Bryman, 2012; Aguinis and Solarino, 2019).

In total 15 interviews with practitioners were conducted to obtain a descriptive account and learn facts, experiences and understandings from individuals in key positions to comprehend the topic of interest (Rowley, 2012; Vaughan, 2013; Aguinis and Solarino, 2019). The research setting was thus selected for its suitability to reveal existing relationships and underlying phenomena. The interview sample and context are thereby not representative of some general population, but rather chosen such that they facilitate the generation of new theoretical insights (Eisenhardt and Graebner, 2007). To that regard, practitioners in the field of data science and machine learning were deemed as particularly suitable to provide practical insights to the research questions for two reasons. First, as the topic of causal inference in machine learning is based in the computer science and economics literature, it is reasonable to assume that it diffuses to the industry primarily via data scientists and machine learning engineers. Second, as this study is interested in the role of causal inference for (strategic) decision making in organizations, the topic can best be investigated by drawing on the experience of data scientists working on data-augmented strategies in today's organizations. Interviewees were recruited via two channels. One, by means of public postings with a call for participation on professional social networking and development platforms (e.g. Twitter, LinkedIn, Kaggle). Two, via e-mail and referrals within the community. Potential interview partners were provided with a short description of the research project. A selection was then made such that the sample of respondents varied in terms of industries, countries and size of companies. Table 1 provides profiles of the practitioners interviewed. All participants were deemed as equally important to the research. Interviews were held in one consecutive round from September 2019 to May 2020. All interviews were conducted in English[12], in the form of semi-structured interviews to maintain flexibility towards the interview flow and encourage the interviewees to share experiences around the theme. To that regard, a guide of twelve open-ended questions with one to four sub-questions each was

[12]With the exception of the interview with CONS1, which was conducted in German and subsequently translated for the analysis.

prepared, starting with general questions such as "What role do data science and machine learning play in your organization?" and "Is data science also relevant for corporate strategy in your organization?", followed by questions related more specifically to causal inference. Throughout the interviews, questions were selected such that they on the one hand facilitate a detailed, flowing conversation, allowing the interviewees to speak relatively freely about their knowledge and on the other hand, provide insights relevant to the research purpose. The question guide was iteratively revised and updated during the first interviews (Bryman, 2012). The final version can be found in Appendix A.1. Interviews were scheduled to take 30 to 45 minutes each and were conducted via video conferencing tools such as Skype or Zoom, which allowed to speak to interviewees at distinct locations and at the same time removed potential interviewer bias (Bryman, 2012). Concerning research ethics, prior to the interview, participants were informed about the research project, procedures and the confidentiality of their responses (Rea and Parker, 2014). For means of analysis, the interviews were recorded, anonymized and transcribed.

To study the experiences and understandings held by informants within the exploratory research design adopted, qualitative content analysis was used to extract a holistic and descriptive account of the meaning of the textual material with respect to the research topic (Weber, 1990; Morris, 1994; Mayring, 2000). As an established method for qualitative analysis, it achieves a systematic description of the material by reducing it into identified content categories that describe the phenomenon of interest. The recording unit was identified as words (when applicable to the code), sentences and paragraphs. Main content categories were initially derived from the research and interview questions, determining the levels of abstraction for the inductive codes formulated in the second round of coding. Categories 3 to 10 (of the final code system in Appendix A.2) were thus initially defined, providing a criterion of selection. As a first step of reduction, the unit of analysis was established as those textual passages (paragraphs and sentences) of the transcripts that were relevant to these main categories (Guthrie et al., 2004; Cho and Lee, 2014). After retrieving the relevant textual material, the first 8 interviews were coded with the pre-determined main categories. Strauss and Corbin (1990) argue that when research questions are open and no hypotheses are formulated, grounded theory methods can effectively be utilized to extract a descriptive account from qualitative data. Therefore, the material was coded a second time using the open coding approach from the grounded theory framework to extract codes emerging from the data. The codes obtained from the first round of open coding were further grouped into inductive categories, formulated out of the material. Those categories and codes were then either subsumed to one of the main categories or formed a new category. On the basis of these subcategories, the material was coded again, taking into consideration the remaining 7 interview transcripts until no new codes were added to the code system, suggesting theoretical saturation. This point of saturation was reached after two rounds of open coding 12 interviews. Finally, the entire material was coded with the defined code system. Throughout the rounds of coding, the main categories were consistently revised on the basis of codes emerging from the material, providing an iterative development and formative check of the code system (Weber, 1990; Mayring, 2000). The final coding frame (Appendix A.2) consists of eleven main categories each with its own subcategories that were inductively formulated out of the material. Reliability of the code system was ensured through the involvement of two researches in the process. Disagreements were discussed and resolved by conceptual clarification.

On the basis of the first 8 interviews, the survey instrument was developed and administered in parallel with the conduction of the second half of the interviews. This inductive approach improves construct validity and the data obtained (Greene et al., 1989) and significantly increases the probability that the survey is relevant to the research and capable of providing meaningful insights (Creswell and Plano Clark, 2018). To that regard, the interviews in particular provided a clarification of relevant concepts and a common terminology and revealed important variables and questions to be investigated with the survey (Bazeley, 2008; Bryman, 2012; Creswell, 2014). Hence, closed question responses, categories and scales were derived from interview insights. A pre-test of the instrument was run with 3 of the former interviewees, instructed to pay special attention to the understandability and adequacy of the survey with respect to the topic of interest. Feedback was incorporated to generate the final instrument. The final questionnaire consisted of 7 parts with a total of 28 questions and can be found in Appendix A.3. Multiple choice closed question responses were displayed in random order to respondents. Deriving from the preceding interviews, the target population of the survey was determined as all data scientists in organizations that emphasize big data and machine learning in their business. The respondents were taken as representatives of their field and their organization in particular. The survey was conducted as a web-based survey. Potential respondents were identified and recruited via two means. One, from a list of suitable organizations a random sample of 112 firms was contacted by email. Two, professional contacts and referrals were contacted either directly by email or by a public posting with a call for participation on professional social networking and development platforms (e.g. Twitter, LinkedIn, Kaggle). At the beginning of the questionnaire respondents were offered a summary report of the research project (Kriauciunas et al., 2011), which was expected to be especially interesting to practitioners currently engaged with the topic in their organization. In total 342 responses were recorded, from which 108 were discarded due to non-response which includes respondents that only filled out the general information and dropped out of the survey when questions became more specific regarding data science and causal inference.

A.6 Practical causal inference applications

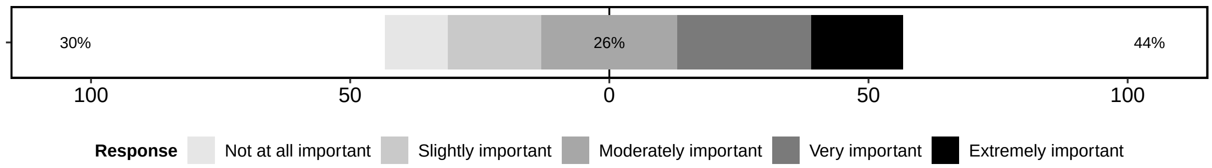
TABLE 4: Interview excerpts

Model the business environment	
TOUR1	”what we really want to know is: Is the processing time really the cause of it (a sale) or are there other variables that we don’t record?”
TECH1	”One is the problem of customers churning out and you want to know why they are churning out. The second one is more about understanding what causes revenue. You may have any kind of product, which you have high dimensional data on, and some of the data surely is important for understanding what causes an increase in revenue. Hence, you want to figure out, using data science, which parts of the actions you are taking, or characteristics of customers are actually leading to higher revenue or lower churn or any of the business metrics that you may care about.”
RETA1	”At the moment a big project we are working on is to consider all the variables we are optimizing for and try to work out whether they are actually good variables to optimize for. Good in terms of, if we can cause a customer to take these actions or go through this journey, they will have a different relationship with our business and will become qualitatively a better customer and thus spend more money. So, we are trying to identify behaviors that are indicative of people leaving and identify what is causing this behavior.”
CONS1	On the topic of fuel efficiency, we have questions such as: What are the actual effects of the individual components on fuel efficiency and how do I have to coordinate or exchange them so that my fuel efficiency is as high as possible?
(Robust) Forecasting	
TECH3	”So, there’s these causal inference problems that involve taking a more limited amount of randomization and trying to project what would happen in the case that everybody got some kind of treatment for a longer time.”
TOUR1	”When we get a lead, we are actually interested in how likely it is that this lead responds positively to us when we send out an offer.”
RETA1	”One project we are working on at the moment is linking our A/B testing infrastructures, so we can get a short run metric like, for instance, revenue per user for different versions of the website. (...) to predict what might happen over the next 1-2 years.”
Process optimization	
TOUR1	”It’s actually a process optimization step. The processing time in fact occurs when the human has to go into our back-office tools and adjust this trip. For technical reasons, these adjustments cannot be done on the website at the moment. In this case – and we invest a lot in this – we are interested to see: How can we reduce the response time? How can we make the tools easier to use? How can we invest in standard procedures for these 20 % of the cases which drive 80 % of the value?”
MANU1	”For example, I have a large typical 99% accuracy where my system works and I want to analyze the 1% which I think is not typical in the observational data and I would want to run a causal analysis how statistically important that factor is.”
Performance evaluation	
TOUR2	”Many things are difficult for us to test offline, because it will change what we show to the user, so we don’t know if the change that we are going to do is going to cause a positive impact.”

TECH3	”Then there’s also a lot of causal inference questions around what would happen if we changed the way we ran the business in various ways that would be quite disruptive for us to run tests with. (…) A good example of that would be our memberships program. We can’t actually A/B test a memberships program, because once we launch it, we can’t exclude people from participating. (…) However, it’s important for us to estimate what happens when we launch it more broadly.”
TECH5	”Most of our experiments are about some feature change that we think will improve the product. So, we are not terribly worried about exposing people to it. We just want to verify that it is an improvement and how much of an improvement it is.”
Inform strategic choices	
CONS1	”In pricing, questions surrounding the drivers behind certain variables are of interest. More specifically, that means: Do the improvements that we see come from the pricing strategy or are there forces outside your own market that create this effect, changing everything structurally without you exactly knowing how and why?”
ONS4	”“How should I address the individual user to maximize the click rate?”. I think most questions really are about, how I should change my business process to achieve some optimization goals. (…) Essentially you are asking, how you should change your status quo. That is usually the question and for that you need causality.”
TOUR1	”When we identify an actual correlation, we might decide to invest into this feature. Given that as a small company, our resources are limited, we need to distribute our efforts efficiently. If, for instance, we run two months of development with our team to improve that feature and it turns out that it is not a causal relationship, that would constitute a considerable loss of resources. That’s why understanding causes and bringing facts to the table when making these prioritization decisions, is really a key success factor that we believe in.”

A.7 Additional survey results

FIGURE 6: (Q9) How important is data science for strategic decision making in your organization?



Note:. For the percent numbers in likert scale graphs, responses are grouped into low, neutral and high

FIGURE 7: (Q13) To what extent do you agree with the following statement: "In our organization we have the necessary skills and capabilities for causal inference"?

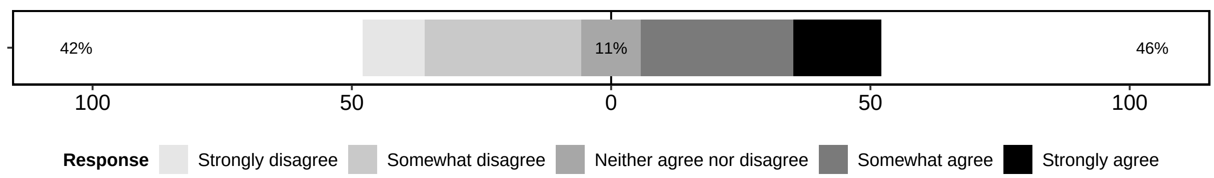


FIGURE 8: (Q21) Which causal inference tools / software are you aware of?

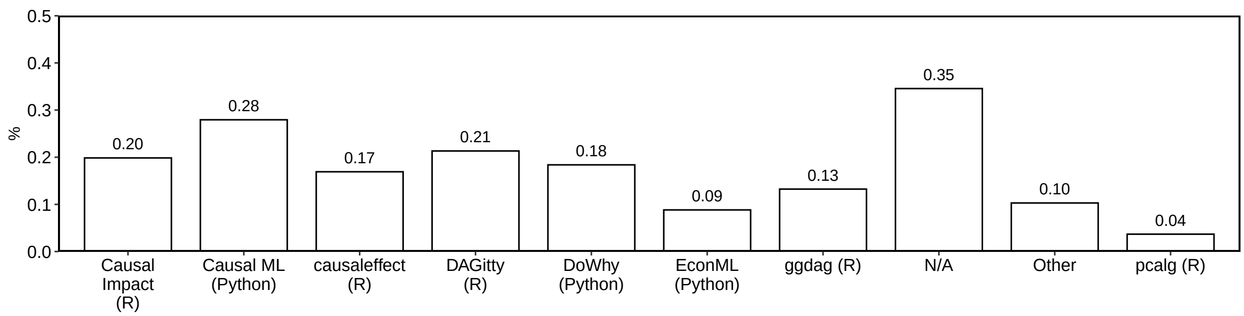


FIGURE 9: (Q22) How suitable do you find existing causal inference tools / software libraries for your purposes?

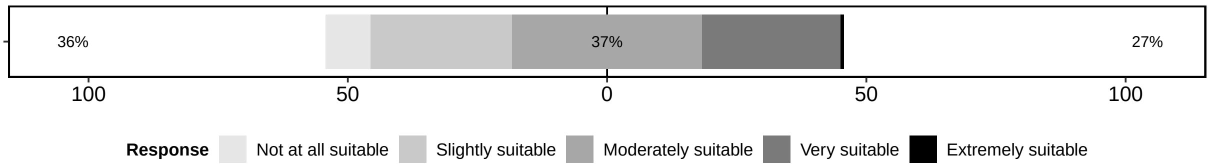


FIGURE 10: (Q25) How important are the following disciplines / educational backgrounds of employees for improving the causal inference capabilities of your organization?

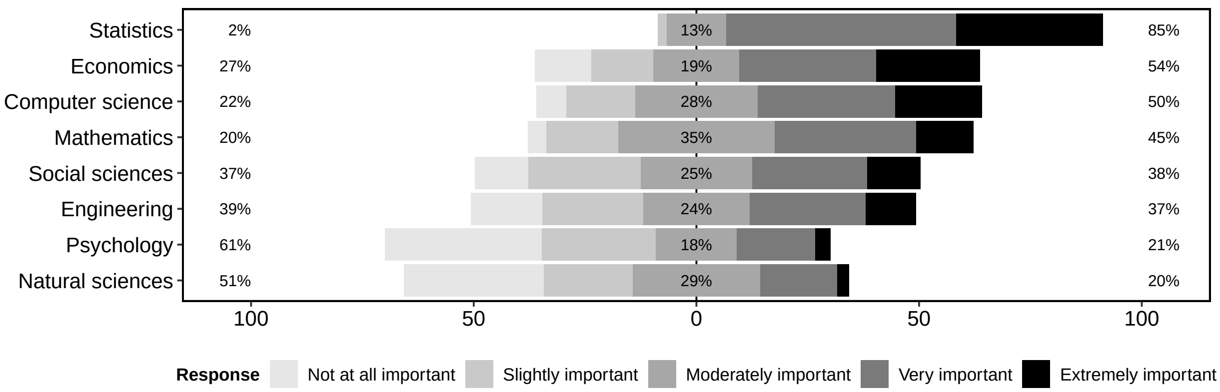


FIGURE 11: (Q27) Do you think that data science is improving human decision making in your organization?

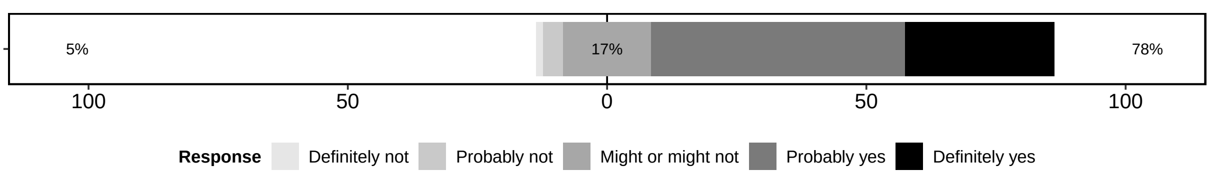


FIGURE 12: (Q28) To what extent do you agree or disagree with the following statement: "Causal inference methods will become more important for data-driven decision making in the future"?