Causal Machine Learning and Business Decision Making*

Paul Hünermund† Jermain Kaminski‡ Carla Schmitt§

This draft: May 18, 2021

First draft: March 1, 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 senior decision-makers, as well as quantitative survey data, this study argues that causality is 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, Data Science, Causality, Machine Learning, Theory-based View

JEL Classification: C01, C45, D80, M21

*Names appear in alphabetical order. All authors contributed equally. We like to thank Maastricht University for research support and participants of the Strategy Science Conference 2021 for helpful comments on a previous version of this paper.

†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, Tongersestraat 53, NL-6211 LM Maastricht. Email: j.kaminski@maastrichtuniversity.nl

§Maastricht University, School of Business and Economics, Department of Organisation, Strategy, and Entrepreneurship, Tongersestraat 53, NL-6211 LM Maastricht. Email: carla.schmitt@maastrichtuniversity.nl

1 Introduction

The age of big data has given rise to data science and machine learning as promising tools for organizational and strategic decision-making in which managers rely less on intuition and more on data (Brynjolfsson & McElheran, 2016). Machine learning technologies are seen as workhorses in many organizations (Agrawal, Gans, & Goldfarb, 2018), significantly driving firm value and performance (Mithas, Ramasubbu, & Sambamurthy, 2011; Rahmati, Tafti, Westland, & Hidalgo, 2020). In many industries, data-driven strategies have become instrumental to achieving competitive advantage (Bloom, Sadun, & Van Reenen, 2012; Brynjolfsson, Hitt, & Kim, 2011; Brynjolfsson & McElheran, 2019; LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011; Mithas et al., 2011; Tidhar & Eisenhardt, 2020). For this reason, some authors have classified data science and machine learning as newly emerging *general purpose technologies* (Goldfarb, Taska, & Teodoridis, 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 with a great degree of accuracy (Hastie, Tibshirani, & Friedman, 2009; Shrestha, Krishna, & von Krogh, 2021). These algorithms underpin modern approaches to business analytics and artificial intelligence (AI). In recent years, deep learning algorithms have been applied to a variety of different decision-making problems (Athey & Imbens, 2019; Blei & Smyth, 2017; Choudhury, Starr, & Agarwal, 2020; Tidhar & Eisenhardt, 2020). Because of their superior forecasting abilities, compared to traditional statistical and econometric techniques, machine learning methods have been called *prediction machines* (Agrawal et al., 2018), a term that 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, the label *prediction machines* 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) to achieve a desired result in line with organizational goals (Christensen, Hall, Dillon, & Duncan, 2016; Cyert, Simon, & Trow, 1956; Simon, 1964). Assessing the likely impact of these interventions ex-ante—a capacity that is crucial for optimal decision-making—requires causal knowledge (Athey, 2017; Bareinboim, Correa, Ibeling, & Icard, 2020; Bertsimas & Kallus, 2016).

In other words, 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, Raisinghani, & Theoret, 1976). By contrast, most commonly used machine learning algorithms, including decision trees, support-vector machines, and deep learning, remain purely correlational and are thus only able to make accurate predictions in a static domain (Pearl, 2019). Once perturbations 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 ask whether there is a mismatch between the managerial problems that organizations try to tackle with data analytics and the methods they use in relation to the challenge of causal inference. If so, we would further like to know whether practitioners are aware of this gap and what actions they take to overcome it. Toward this end, we employ a mixed methods research design in which we combine qualitative interviews with a quantitative survey of practitioners, as well as a multitude of other data sources including online resources, educational material, blog posts and software packages originating from the data science community. The interviews we conducted and auxiliary data sources thereby provide us with rich, contextual insights about the mechanisms underlying modern business analytics in contemporary organizations (Bettis, Gambardella, Helfat, & Mitchell, 2014), while the survey study allows us to solicit information from a broader sample in a more systematic way (De Leeuw, Hox, & Dillmann, 2008).

The results indicate an ongoing shift in the community of practitioners towards the growing application of causal data science methods for business decision-making. Traditional correlation-based machine learning approaches are increasingly perceived as unsuitable for informing a wide 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 our respondents plan to invest more into their causal inference capabilities in the coming years. Several key players, particularly in the technology sector, who 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 practical as well as theoretical challenges that will be identified in the course of this paper.

Our study contributes to the strategic management literature by clarifying the epistemological foundation for causal learning in an organizational decision-making context and delineating theoretical impediments to the success 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, Cordova, Gambardella, & Spina, 2020; Felin, Gambardella, Stern, & Zenger, 2020; Felin, Gambardella, & Zenger, 2020; Felin & Zenger, 2009, 2017) and demonstrate that theory is an essential input to *data-augmented decision-making*. At the same time, we show how the literature on causal inference in machine learning and AI can significantly contribute to the inferential power of managerial theorizing and support users in more effectively integrating data science into the strategy formulation and decision-making process. Finally, we discuss the practical implications of our study concerning the development of causal learning as an important organizational capability.

2 Theory

2.1 Causal Knowledge in Management

Causal knowledge involves the awareness and understanding of cause and effect relationships in the world. It is one of the most important components of human cognition, inseparable from our thought and essential to our survival (Pearl & Mackenzie, 2018; Waldmann, 1996). It enables an actor to predict 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 & Mackenzie, 2018). 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. Marketing

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

executives might try to predict whether ads on a mobile or desktop version of a social network will lead to higher click-through rates (Lu & Du, 2020). Human resource managers might want to know whether increased teleworking would exert a positive influence on employee productivity and well-being (Vega, Anderson, & Kaplan, 2015). Founders of a start-up might wonder whether certain communicative signals in a crowdfunding campaign will result in better funding outcomes (Kaminski & Hopp, 2019). "What if" questions of this kind typically arise in the context of strategic business problems. Causal knowledge thus constitutes an important parameter in taking central management decisions (Felin & Zenger, 2009).

Among organizational decisions, strategic decisions are generally identified as those managerial choices that are important in terms of the resources committed, the actions taken, and the precedents set (Eisenhardt & Zbaracki, 1992; Mintzberg et al., 1976; Mitchell, Shepherd, & Sharfman, 2011; Shrivastava & Grant, 1985). They define the direction of the organization (Eisenhardt & Zbaracki, 1992) and thereby have long-term effect on the firm's administration, structure, and performance (Shivakumar, 2014; Shrivastava & Grant, 1985). As the problems strategic decisions address are often novel, complex and unstructured (Cyert et al., 1956; Shivakumar, 2014), business success, performance and efficiency depend on organizational decision-making structures (Simon, 1964) and on the processes underlying managers' strategic choices (Mitchell et al., 2011), which, as the following discussion shows, are based largely on causal knowledge. Adopting Mintzberg et al. (1976)'s descriptive approach to understanding and defining managers' strategic decision activities, this process can best be conceptualized as containing *three phases of decision-making*. In the first phase, managers recognize a performance-objective gap in the data and thoroughly define the strategic problem. In the second phase, managers identify and design alternative actions to solve the problem. In the third phase, managers select the most feasible solutions and evaluate them in relation to organizational goals to arrive at a final choice.

To fully identify the strategic problem and define a starting point for the process of solving it, "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 facilitates problem formulation (Baer, Dirks, & Nickerson, 2013) and aids managers in identifying and defining important variables and objectives of the decision task (Maule, Hodgkinson, & Bown, 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 (Pearl & 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. Alternative actions and the consequences attached to them, are not given but must be determined; this search for cause and effect relationships is an integral part of all stages 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). Nickerson and Zenger (2004) stress 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.

Such cognitive evaluation can be realized by agents forming and consulting mental images of their information worlds and the problem space (Gavetti & Levinthal, 2000; Pearl & Mackenzie, 2018; Walsh, 1995). The literature on managerial cognition finds that managers build causal mental maps to support their decision-making efforts (Gary & Wood, 2011; Hodgkinson, Bown, Maule, Glaister, & Pearman, 1999; Maule et al., 2003). These 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 (Gavetti & Levinthal, 2000; Hodgkinson et al., 1999; Walsh, 1995). 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 asserts that an understanding of cause and effect in the relevant business context allows decision-makers to focus on strategic actions (Hodgkinson et al., 1999), speeds problem-solving (Walsh, 1995), and increases the quality of decision making (Gary & Wood, 2011; Waldmann, 1996). Beliefs about causal structures thereby assist decision-makers in detecting covariates and in distinguishing real from spurious correlations (Vera-Muñoz, Shackell, & Buehner, 2007; Waldmann, 1996). Gary and Wood (2011) assert 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. 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 heterogeneity in firm performance. Similarly, Gavetti and Levinthal (2000) conclude that "even simple models of the world have a tremendous potential to guide search processes" (ibid. p. 135). Moreover, the literature on business models provides further evidence on causal mapping in the business context. Several scholars conceptualizes the business model itself as a system with underlying cause and effect relationships that define how the firm can achieve its long-run objectives by realizing concrete strategies (Baden-Fuller & Mangematin, 2013; Furnari, 2015; Vera-Muñoz et al., 2007). "The business model [...] should be a stripped-down characterization that captures the essence of the cause–effect relationships between customers, the organization and money" (Baden-Fuller & Mangematin, 2013, p. 419). As a cognitive instrument it thus provides a reference frame of causal relationships to address strategic management questions and structure organizational decision-making (Simon, 1964).

2.2 Data Science and Machine Learning in Business Analytics

The emergence of data-augmented decision-making, fueled by new machine learning technologies and opportunities for data collection, has changed the way managers make decisions, relying more on data and less on intuition (Brynjolfsson & McElheran, 2016). Consequently, the topics of *business intelligence* and *business analytics* are receiving increasing interest from researchers, job-market candidates and practitioners alike (Athey & Luca, 2019; Chen, Chiang, & Storey, 2012; Lycett, 2013; Sharma, Mithas, & Kankanhalli, 2014) and data science has assumed a central role in managerial decision-making. There is abundant evidence for a positive relationship between data-augmented decisions, enhanced productivity, and the increase of intangible firm value (Bharadwaj, Bharadwaj, & Konsynski, 1999; Bloom et al., 2012; Brynjolfsson et al., 2011; Brynjolfsson & McElheran, 2019; Mithas et al., 2011; Rahmati et al., 2020). As LaValle et al. (2011, p. 22) find, "the correlation between performance and analytics-driven management has important implications to organizations, whether they are seeking growth, efficiency or competitive differentiation."

Machine learning evolved primarily as a tool for prediction problems, that is, problems that use an input to predict the outcome through observed associations or relationships (Agrawal et al., 2018; Davenport & Harris, 2009; Iansiti & Lakhani, 2020; Varian, 2014, 2016). The promise of big data in organizations today, therefore, essentially lies in significantly advanced pattern detection abilities derived from gradually improving machine learning models and technologies (Bajari, Chernozhukov, Hortaçsu, & Suzuki, 2018). Consequently, so-called *prediction machines* (Agrawal et al., 2018)—systems using these models and technologies—are seen as workhorses in many companies, providing continuously better and cheaper forecasts to decision-makers.

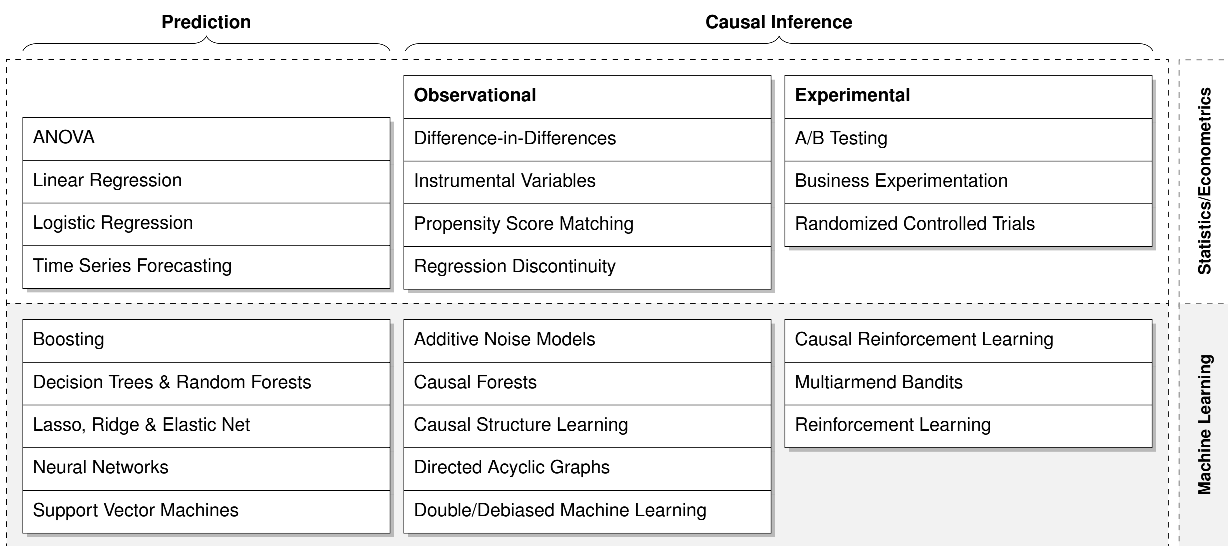
Much of this expansion in data-augmented decision-making derives from the success of deep learning architectures that map from observable inputs to outputs via multiple layers of high-dimensional data. These constitute effective tools for unstructured predictions and can be employed to solve complex classification problems (Athey, 2018; Athey & Imbens, 2019; Blei & Smyth, 2017; Choudhury et al., 2020; Mullainathan & Spiess, 2017) in contexts ranging from predicting customer churn (Agrawal et al., 2018; Ascarza, 2018) to making economic predictions with satellite images (Athey, 2018; Donaldson & Storeygard, 2016; Henderson, Storeygard, & Weil, 2012), or supporting hiring decisions (Chalfin et al., 2016). In a strategic context, machine learning systems have shown to be capable of finding the optimal revenue-model fit (Tidhar & Eisenhardt, 2020), and, as the reinforcement learning algorithm of Google DeepMind's AlphaStar exemplifies, are able to map out and predict strategies in a complex gaming simulation (Vinyals, Babuschkin, ..., & Silver, 2019).

As Agrawal, Gans, and Goldfarb (2019, p. 31) 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 amount to "an exploratory tool to discover robust patterns in quantitative data" (Choudhury et al., 2020, p. 1), they are not capable of deriving causal effects: "[T]he goal [of machine learning] is predictive power, rather than estimation of a particular structural or causal parameter" (Athey & Imbens, 2019, p. 7).

In the context of the preceding theoretical discussion, this point implies an important mismatch between machine learning capabilities and the analytical requirements of (some of) the problems they address (Bertsimas & Kallus, 2016). Pattern discovery by itself has rarely been shown to be relevant to strategic management questions. Classical machine learning models are thus incapable of providing the causal knowledge that is required by strategic decision-making processes. As Fedyk (2016, p. 3) points out, the business problems addressed with classical machine learning, should (only) be those that "(1) require prediction rather than causal inference; and (2) are sufficiently self-contained, or relatively insulated from outside influences." Yet, Christensen et al. (2016, p. 4) assert that "though it's no surprise that correlation isn't causality, we suspect that most managers have grown comfortable basing decisions on correlations."

Due to this limitation in explaining causal relationships, classical machine learning is only of limited use for managerial decision tasks. Additionally, the fact that outcomes of deep learning models cannot be easily interpreted because of the many feature layers involved in a decision (Rai, 2020), further complicates the use of prediction models for strategic business questions. As Athey (2017) points out, gaps yet persist between making a prediction and making a decision. While most existing machine learning research focuses on the relationship between data and prediction, the relationship between prediction and decision is still underdeveloped. Figure 1 provides an overview of commonly used data analysis and estimation algorithms, and shows how they map across the space of prediction-based methods, observational causal inference techniques, and experimental approaches.[2]

Figure 1: Examples of popular data analysis algorithms in statistics and econometrics, as well as machine learning and artificial intelligence, classified according to prediction and causal inference methods. Causal inference methods are further differentiated according to observational (based on ex-post observed data) and experimental approaches.



To optimize data-augmented decision-making in organizations, assumptions and limitations of prediction methods, especially with respect to the questions that can be answered, need to

[2]As the theoretical discussion in the following section will make clear, such a classification necessarily remains fuzzy. Estimation procedures such as OLS or LASSO can be employed for prediction and causal inference purposes, and the use of methods such as PSM or IV does not per se ensure causality. Nonetheless, the depicted approaches and algorithms are frequently associated with a set of key assumptions that, if valid, allow to draw causal inferences in certain contexts. We regard the taxonomy offered by Figure 1 as a useful, yet incomplete, overview of a rapidly evolving field.

be understood and considered in the decision process (Athey, 2017). As Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan (2017, p. 40) have shown, "being clear about how predictions translate to decisions can substantially influence how the prediction function is evaluated." Business problems such as pricing or inventory management that are typically addressed with prediction models, require more robust causal assumptions to guide optimal decision outcomes (Bertsimas & Kallus, 2016; Lycett, 2013). Prediction systems lack "the knowledge of causal relationships among the various variables, relying solely on past data to make decisions. Thus, [standard machine learning] cannot foresee future consequences as humans can" (Balasubramanian, Ye, & Xu, 2020, p. 14). Optimal managerial 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, Malo, Nagbøl, & Penttinen, 2020, p. 270).

Shiffrin (2016) notes that big data by itself is not very helpful to organizations, as the detection of patterns is only the first step towards causal inference. The author argues that "[e]xplaining 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. The field of causality thus presents a promising future path for business strategy, intimately connected with developments in statistics, social science, and computer science, and well suited for observational data and experiments (Blei & Smyth, 2017). Shifting attention to causal inference in data science and machine learning therefore appears as an inevitable next step. "[D]iscovering causal relations means acquiring robust knowledge that holds beyond the support of an observed data distribution [...], and it extends to situations involving forms of reasoning" (Schölkopf et al., 2021, p. 2).

Causal methods such as experimentation are already seen to significantly advance businesses' understanding of the "causal relationships between human behavior and economic value," which can provide valuable new insights into decision-making (Gillon, Brynjolfsson, Griffin, Gupta, & Mithas, 2012, p. 290) and entrepreneurial strategy (Agrawal, Gans, & Stern, 2021; Camuffo et al., 2020; Koning, Hasan, & Chatterji, 2020). Accordingly, Hartford, Lewis, Leyton-Brown, and Taddy (2016, p. 20) argue more generally that "the next generation of problems in [machine learning] 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." Illustrated by Figure 1, a growing array of not only experimental, but also observational causal inference methods in the field of machine learning enable such transition and provide the causal knowledge necessary to fully address strategic business questions. As Malone (2018, p. 258) proposes "strategy combinators" that merge human reasoning with machine learning, "could rapidly generate and evaluate various strategic possibilities". We hence propose that big data (George, Haas, & Pentland, 2014) needs to be complemented with *smart causal models* to advance data-augmented strategic decision-making in organizations.

2.3 The Epistemological Challenge of Causal Inference

The task of causal inference, to predict the 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 that environment and thereby changes it. The chain of events that would have prevailed without the action is altered, so a prediction derived from passive observation carries little information about the events that will transpire afterward. Formally, this idea can be best illustrated with the help of a causal diagram (Durand & Vaara, 2009; Pearl, 1995). Figure 2a 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 2: (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 2a 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.[3] The model 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, their presence renders the model stochastic with a probability distribution $P(\varepsilon)$ over the set of endogenous variables.[4]

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 & Pearl, 2016). The only requirement is that causal relationships must be acyclic (Pearl, 2009). That means that by tracing paths between nodes following the directed edges in the diagram (such as, e.g., $Z \rightarrow X \rightarrow Y$ in Figure 2a), 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 $A \rightarrow B \rightarrow C \rightarrow A$ are ruled out, a stipulation that captures the intuitive notion that a variable cannot be a cause of itself.[5] Due to this property of acyclicity, causal diagrams are also referred to as *directed acyclic graphs* (DAGs) in the causal inference literature (Pearl, 2009).

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

[3]Equations would not be able to capture this asymmetry since, e.g., $X = aZ$ would be equivalent to $Z = X/a$.

[4]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.

[5]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.

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 to assess the quantitative effect of such an intervention on other variables of interest in the model. 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 & Vytlacil, 2007), can easily be derived from it.

Interventions can also be illustrated graphically in a DAG. Figure 2b 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 & Bareinboim, 2021). 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* (Bareinboim et al., 2020; Pearl & Mackenzie, 2018).[6,7]

Nevertheless, under certain circumstances, $P(y|do(x))$ might be transferable into an equivalent expression that can be computed from pre-intervention information. For the graph in Figure 2a,

[6]The hierarchy states that information at one layer (*seeing*) almost always (in a measure-theoretic sense) underdetermines information at higher layers (*doing*). This difference is conceptually related to the *fundamental problem of causal inference*, as formulated by Holland (1986). 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 challenges of obtaining causal knowledge are already introduced there.

[7]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 in it. An example here would be the decision to optimally allocate storage capacity C based on seasonal demand patterns Y. If managers want to induce change in the system, 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.

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 so-called *identification* problem of causal inference (Koopmans, 1949; Pearl, 2009), since it allows the analyst to estimate post-intervention (causal) quantities purely based on passive pre-intervention observations without manipulating the treatment variable X directly (referred to as *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 2a, for example, no influence factors other than Z jointly affect X and Y, so assessing the conditional distribution Y given X for each value of Z separately, i.e., $P(y|x, z)$, 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).

Experimentally manipulating a variable and measuring the 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, the capacity to carry out experimental studies does not alleviate the need for a causal model (Deaton & 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 certain 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 as *transportability* in the causal inference literature; while the social sciences commonly refer to it as *external validity* (Bareinboim & Pearl, 2016). Solving this problem requires ex-ante causal assumptions about the data generating process.[8]

Moreover, in many practical settings, directly intervening on a variable of interest is not feasible, because it would be costly, unethical and/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 & Pearl, 2012a).[9] 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, Zettelmeyer, Bhargava, & Chapsky, 2019). Instead, however, consumers can be randomly assigned to either a treatment group, who will be shown the ad once they log on to the platform, or a control group, who will only see a neutral message. That way, advertisers can effectively manipulate the ad exposure, but the assignment will remain imperfect because many consumers will never visit the platform during the field phase of the experiment. Thus, these customers are never exposed to the ad, even if they have been assigned to the treatment group—a problem called "one-sided noncompliance" in the literature (Imbens & Rubin, 2015). This 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 & Pearl, 2012a; Semadeni, Withers, & Trevis Certo, 2014), another fact that highlights the need for a model to obtain 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 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. Thus, there is a potential mismatch between the questions that are being pursued and the capacity of the methods that are employed to answer them. The purpose of this study is to assess the practical implications of this mismatch and to

[8]Often analysts implicitly assume *direct transportability*, i.e., that causal effects remain constant over time and across different populations (Pearl & Bareinboim, 2011). However, this may or may not hold true.

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

explore the use of causal inference methods in contemporary organizations.

3 Method & Analysis

Given the novelty of this research focus, we employed a mixed methods research design, combining interviews with a survey instrument in an exploratory sequential design (Creswell & Plano Clark, 2018), to derive a comprehensive understanding of the topic of causal inference in contemporary organizations (Bettis et al., 2014; Johnson, Onwuegbuzie, & Turner, 2007). Additionally, due to 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 from individuals in key positions to comprehend the topic (Aguinis & Solarino, 2019; Rowley, 2012; Vaughan, 2013). The research setting and sample were thus selected for their suitability to reveal existing relationships and underlying phenomena. They are not representative of some general population but rather chosen to facilitate the generation of new theoretical insights (Eisenhardt & Graebner, 2007). In that regard, we deemed practitioners from the field of data science and machine learning 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 literatures of computer science and economics, it is reasonable to assume that it diffuses into the industry primarily via data scientists and machine learning engineers. Second, as this study explores 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. Semi-structured interviews were held from September 2019 to May 2020 (Appendix A). The interviews took 30 to 45 minutes each and were conducted via video conference. Before each interview, the participants were informed about the research project, its procedures, and the confidentiality of their responses (Rea & Parker, 2014). To facilitate analysis, the interviews were recorded and transcribed. To extract a holistic and

Table 1: Overview of interviews

ID	Role	Industry	Country
CONSI	Data scientist	Consulting	GER
RETA1	Senior applied (data) scientist	Retail & consumer goods	GER
TECH1	Research data scientist	Technology, media, telecommunications	USA
CONS2	consultant & software engineer	Consulting	GER
CONS3	Data science consultant	Consulting	USA
MANU1	Data scientist	Industrial manufacturing	GER
TOUR1	Chief technology officer	Hospitality & tourism	GER
TECH2	Senior machine learning engineer	Technology, media, telecommunications	USA
MANU2	Senior vice president	Industrial manufacturing	GER
TECH3	Research data scientist	Technology, media, telecommunications	USA
CONS4	Data science consultant	Consulting	GER
HEALTH1	Research data scientist	Health services	ISR
TOUR2	Data scientist	Hospitality & tourism	GER
TECH4	Senior applied (data) scientist	Technology, media, telecommunications	USA
TECH5	Data scientist & machine learning engineer	Technology, media, telecommunications	CAN

descriptive account of the meaning of the textual material in light of the research questions, the transcripts were analyzed using qualitative content analysis (Mayring, 2000; Morris, 1994; Weber, 1990). Primary content categories were initially formulated based on the research focus and interview questions to determine the levels of abstraction for the subsequent inductive category development. We then open-coded the material, extracting subcodes and developing additional categories from the data, until theoretical saturation was reached. A detailed description of the coding process can be found in Appendix E. The final coding frame (Appendix B) consists of eleven main categories, each with their own subcategories formulated from the material.

Revealing important variables, clarifying relevant concepts, and establishing a common terminology, the first eight interviews provided the basis to inductively develop the survey instrument, which improves construct validity and increases the probability that the survey is relevant to the research (Bryman, 2012; Creswell & Plano Clark, 2018). A pre-test of the survey was run with three 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 (Appendix C), which was administered online in parallel with the last seven interviews. Multiple choice questions were displayed in random order. Guided by the first interviews, the target population of the survey was data scientists in organizations that emphasize big data and machine learning in their business. The respondents were understood as representatives of their fields and their organizations. Potential respondents were recruited and contacted in the same way as interview partners. In total 342 responses were recorded, of which 108 were discarded as respondents did not go beyond the first page of general questions.[10] The majority of respondents, 68 percent, who completed the survey are from the private sector and only 22 percent come from academic institutions, such as universities or publicly funded research institutes. Although data collection was focused towards data science practitioners in particular (as the description of the survey and introduction text made abundantly clear), we decided to also include the participants from academic institutions in our analysis, expecting that those responding to the survey are from institutes that are in close collaboration with industry. We thus believe that these respondents can contribute valuable and relevant insights to the research, that should not be excluded.[11] Including 33 percent of the respondents, the technology, media, and telecommunications sector is the most represented in the sample, followed by the education, research, and public sector with 17 percent and the financial service sector with 14 percent. Of the respondents, 62 percent are data and research scientists. The size of the respondents' organizations is relatively equally distributed, with 33 percent having 250 or fewer employees and 30 percent having 5000 or more employees. With 35 percent, a third of the organizations are less than 10 years old. Most organizations, 44 percent, are from Europe, closely followed by North American firms with 41 percent. Detailed descriptive statistics can be found in Appendix D. Analysis of the survey responses generally focuses on descriptive figures to extend and validate findings from the semi-structured interviews.

4 Results

Below, we present interview and survey findings in parallel to allow for the triangulation of results across cases as well as methods. The derivation of findings from different interviews is presented transparently (with codes in parentheses) and supplemented with quantitative evidence from the primary survey data to provide validity and generalizability.[12] Where appropriate and conducive to the generation of theoretical insights, relevant online resources reviewed during the research process are incorporated. It should be noted that the last seven of the 15 interviews were performed in spring 2020, during the covid–19 pandemic. While this certainly needs to be considered an unexpected

[10]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.

[11]Findings are robust to excluding respondents from academic institutions from the analysis.

[12]Additional figures not provided in the text can be found in online Appendix G.

circumstance regarding the replicability of this study (Aguinis & 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 almost certainly too recent at the time of the interviews to have had a significant impact on the interview and survey responses.

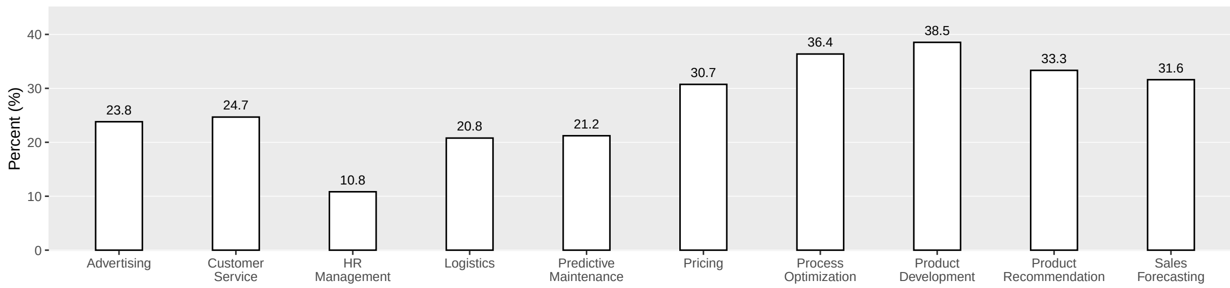
4.1 The types of questions firms address with their data science efforts

To obtain a more general understanding of the business problems practitioners are typically presented with, and to 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%[13] of the companies whose representatives we interviewed, especially of those offering web-based products or software, data science and machine learning are integral parts of their products (1.a.i), employed to ensure and optimize product functionality. This includes recommendation systems, automated pricing algorithms, and, in the case of online marketplaces and metasearch services, price predictions. Accordingly, the survey responses in Figure 3 indicate that respondents employ data science for product recommendations (33%) and pricing (31%) on their platforms. Similarly, 33% of the interviewees (1.a.iv) and 39% of the survey respondents specifically mention the application of data science and machine learning to product development. Moreover, 53% of interviewees (1.a.ii) and 36% of survey participants say that data science is applicable to optimizing business processes, such as improving the response time to a customer request or the scheduling of flights by airlines. Predictive maintenance is identified as a relevant area in the interviews (1.a.iii) and confirmed by 21% of survey respondents. A frequent application of data science and machine learning, according to 53% of interviewees (1.a.v) and 32% of survey respondents, is the forecasting of sales as well as demand and other financial figures. The 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 particular decision situation (1.a.vi).

When asked about the relevance of data science for strategic decisions in particular, all the interviewees confirmed its importance and provided practical business cases (1.b). As TECH3

[13] For the sake of brevity, we use a %-sign in the results section.

Figure 3: Use cases of data science applications in contemporaneous organizations



noted, one way data scientists contribute to the organizations is *"to help 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, 87% of interviewees mention applications of data science that seek to understand the marketplace, including to segment the customer base, to analyze revenue streams and customer churn, and to evaluate business metrics 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); 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.

A theme that emerged 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 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."* However, the problems described by the executives are generally too broadly defined to recommend concrete actions to data scientists. *"[W]ith data science and machine learning in general what I have observed is that there is very little top-down"* (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,"* such as responding to a particular problem described by an executive. Our findings highlight the central role of data scientists in determining the types of analyses the organization's analytical capacities are directed at. Thus, the approaches to strategic problems, and the solutions proposed for them, depend on, and are limited to, the methods and capabilities available to the data scientists in the organization.

4.2 Awareness of the difference between correlational and causal knowledge

To assess how far knowledge of causality is 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, a finding reflected by survey respondents, 97% of whom indicate familiarity with the distinction. 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 that actors in the broader organization will interpret results of correlation-based analyses as causal relationships (2.a).

However, despite this conceptual understanding on the part of data scientists, the degree of recognition of causal inference in their professional work varies greatly across practitioners. At one end, three of the 15 interviewees say that causal inference is not at all considered in their projects (8.a.i). Two of these three interviewees (TECH4 and CONS3) work in firms offering consulting services whose data science efforts are often restricted by their clients' demands and decisions. The third interviewee is an executive (MANU2), who appears to be inexperienced with the topic. As in each case, the interviewee's daily work is determined largely by management views and opinions, this finding seems to indicate that to practitioners in management, who typically are the decision-makers in organizations, causal inference is not yet relevant. With 60%, the majority of interviewees, however, says that the understanding of causal inference is beginning to slowly diffuse in their organization (8.a.ii) and 40% mention that they are new to the topic but very interested in learning more (8.a.iii). Hence, overall, awareness of and interest in the topic of causal inference is growing in industry. Diffusion seems very much bottom-up (8.a.iv), driven by few data scientists. As TECH3 said, *"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."* Similarly, TOUR1 noted, *"Together with one of our data scientists, I am the one who is currently pushing this topic. We are missing that view."*

And, finally, CONS3 explained, *"In particular the data scientists are really aware of it [causal inference] and are following discussions and developments."*

4.3 Importance of causal inference in business today

To explore more closely the 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 in strategic decision situations. Highlighting the relevance of causal knowledge for their work (3), interviewees generally stress the importance of causal inference for addressing a variety of questions in the business context (4). This finding is confirmed by 47% of the survey respondents, who say that causal inference is important for their data science projects. 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."*

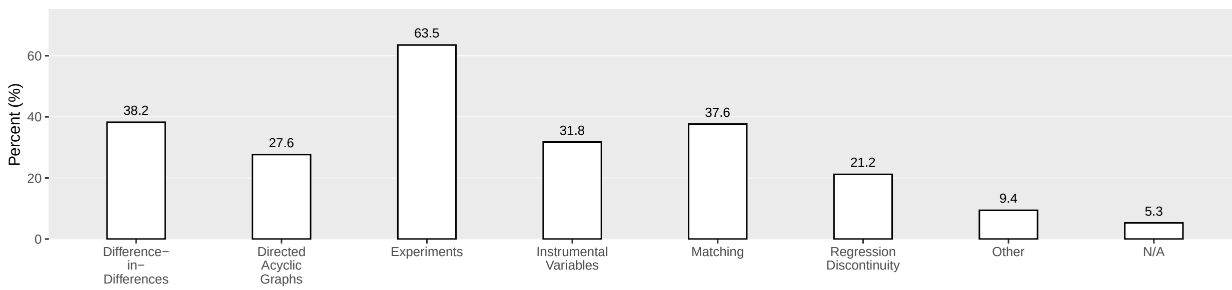
Interviewees provide several examples of business problems where they see the potential to apply causal inference to better understand the relevant causal forces at play in the business environment (4.a) (see also online Appendix F). TECH2 said, *"One area where we are still interested in doing observational 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 say that in practice, causal inference is useful for deriving more robust predictions of business metrics (4.b); increasing operational efficiency (4.c); solving particularly complex problems that require a more fundamental, generally applicable model of reality (4.d); and evaluating the performance of specific interventions such as product changes (4.e). TECH6 explained, *"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."* The majority of respondents speak of 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 (4.f). Stressing the importance of causal inference for strategic decision situations in particular, TOUR1 stated, *"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 still find pure prediction to be more common in their data science projects. 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 the methods available to data scientists often determine the types of questions that can be answered, the strict predominance of correlational methods indicated by interviewees (2.b) implies that causal questions are insufficiently answered or not addressed at all. Indeed, 73% of interviewees realize that the mostly 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, RETA1 said that this means that their organization *"might be optimizing for something that doesn't really cause a change in [customer] behavior."* Emphasizing the importance of raising awareness to and resolving this mismatch in firms' data scientific approaches, CONS4 concluded, *"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."*

4.4 Diffusion of causal inference methods and techniques

Survey results in Figure 4 show that practical causal inference techniques and approaches are unevenly distributed across organizations. By far the most prominent causal inference technique employed by firms today is experimentation (5.h). The majority, 73%, of interviewees and 64% of survey respondents indicate that they apply A/B tests, bandits, or reinforcement learning for causal inference at their organization. In contrast, observational causal inference techniques—those based on ex-post data analysis without active manipulation or randomization—are less widespread. Those mentioned most often by interviewees and survey respondents respectively are difference-in-differences (33% (5.a), 38%), matching (27% (5.c), 38%) and instrumental variable estimation (20% (5.b), 32%). Distinguishing between experimental and observational causal inference techniques, we can derive several findings about the use of causal inference methods and techniques in contemporary organizations' data science efforts.

Figure 4: Usage of causal inference methods in data science applications



As implied above, experiments are the default causal inference technique for most data science practitioners (5.h.i). The majority of participants says 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 using here, but the massive hammer that tech companies swing around when it comes to causal inference is running experiments."* Especially respondents in organizations with web-based business models say they are continuously conducting a large number of experiments (5.h.ii). TECH2 noted, *"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, will go through A/B testing in its final stage."* This finding is well in line with recent discussions in the literature, demonstrating that the relevance of experimental methods in the business domain is growing (Bojinov, Sait-Jacques, & Tingley, 2020; King, 2020; Thomke, 2020). The popularity of experiments in the business context appears to come primarily 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 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.

However, while experiments appear to be the preferred choice to answer causal queries in the business context, several drawbacks, including the A/B testing pitfalls identified by Bojinov et al. (2020), render them impractical in many problem spaces. Of our interviewees, 73% mention difficulties concerning the practical applicability (6.a) of A/B tests, relating to situations in which the business environment, the data availability, or the parameters of interest are unsuitable for an experimental approach. Practitioners add that experiments are often time-consuming, rendering them impractical for pressing business decisions; that control and treatment cannot be sufficiently administered in certain situations (as it might not be possible to exclude people from the treatment); and that the data that the experiment can collect is not suitable for answering the business question. The practitioners we interviewed explain further that experiments cannot be easily and safely set up in environments such as manufacturing and medical services 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, and/or prices to different customers present an important shortcoming to 27% of interviewees (6.a.i) and 36% of survey participants. According to 40% of interviewees (6.b) and 47% of survey respondents, experiments lead to inferior user experience for customers and high costs for the firm. Moreover, interviewees identify particular technical shortcomings (6.c) that impair the reliability of the measured effects; the lack of suitable outcome metrics appears to be the most significant of these, as indicated by 51% of survey respondents. In other words, the outcomes of interest that decision-makers would like to affect are often difficult to observe, requiring analysts to rely on proxy metrics. As TECH3 noted, *"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."*

Finally, 40% of interviewees and 41% of survey respondents mention external validity, that is, the applicability of experimental results in different settings, as a big concern (6.d). In practice, this becomes relevant when experimental results obtained in one market (for example, a geographic area) are supposed to be used in another. The generalizability and applicability of experimental results, as well as the value of data collected, are thus limited within experimental approaches to causal inference. While our 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, previous experimental results might not hold anymore. As TECH2 explained, *"There, external validity becomes a concern. When we rely on 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 most 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. The global pandemic has caused drastic changes in the world, altering people's 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 are presented with the choice to either extrapolate from pre-pandemic results or to rerun experiments, which, as TECH2 confirms, can entail substantial resource investments.

Given those drawbacks of experimental methods, and the advantages practitioners perceive in respect to observational causal inference methods, such as difference-in-differences, DAGs, or matching (Figure 4), some practitioners identify observational methods as relevant alternatives or complements to experiments when the latter are not feasible (3.d). Indeed, survey results reveal that respondents value observational causal inference methods for being based on actual field data (64%), for their ease of implementation (55%), and for the high external validity of their 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 industry identify particular obstacles to using observational approaches. The biggest challenge appears to be the understandability and applicability of observational approaches to practical business cases (7.b), as practitioners perceive them as more complex than standard machine learning techniques and experiments. Interviewees remark that their analysis requires numerous untestable assumptions which need to be based on a valid 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. Similarly, the survey results show that observational causal inference methods are indeed seen as difficult to implement (34%) and explain (37%). Interviewees further state that analyses with observational methods entail lengthy and expensive deployment efforts, making them impractical in fast-moving business environments (7.a), a position 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"* (CONS1) to customers as well as non-data-scientists in management. TECH1 stated, *"Even if it's valid, it's much harder to convince a businessperson based on such a complicated analysis."* From the interviews, it can therefore be derived that more industry examples from different sectors and organizations 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. In fact, the diffusion of practical causal inference tools and software libraries also appears to be in its infancy, as 35% of survey participants do not apply practical causal inference tools. As Table 2 suggests, most libraries were only published in mid-2019, with three industrial open-source libraries leading in engagement on GitHub. Similarly, 47% of interviewees (7.e) are either not aware of observational causal inference methods, are unsure of their practical application, or are not using any external tools but developing their own. Meanwhile, expert users assert that the existing software has significant shortcomings (7.d). Some of them find it not fully developed (7.d.iii). Others say that the user experience is not interactive enough and is still too theoretical; that the tools provided lack graphical ability to make the models more understandable or lack the ability to provide users with comprehensive what-if scenarios (7.d.ii); or, generally, that suitable, standard tools (especially in Python) are not yet available (7.d.i). Survey results confirm that many practitioners, although generally interested, perceive hurdles to adopting observational causal inference methods: only 27% find existing tools and software suitable for their purposes.

Table 2: Software libraries on causal inference (selected examples, May 09, 2021). *Contributors* and *Stars* refer to repository statistics at GitHub.

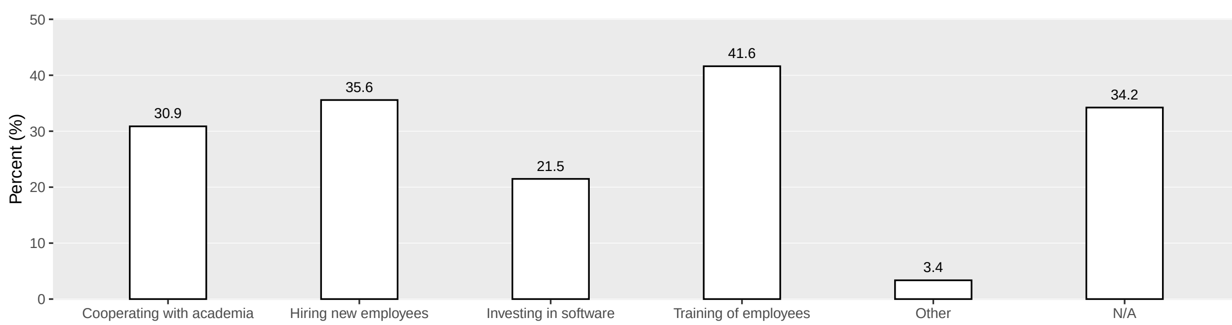
Entity	Library	Release	Source	Contrib.	Stars	Lang.
Microsoft	DoWhy	July 15, 2019	Public, GitHub	39	2,883	Python
Uber	CausalML	July 10, 2019	Public, GitHub	25	1,934	Python
Google	Causal Impact	August 2, 2014	Public, GitHub	7	1,278	R
QuantumBlack	CausalNex	January 28, 2020	Public, GitHub	18	1,098	Python
Academic	ggdag	October 9, 2019	Public, CRAN	1	320	R
Academic	dagitty	August 26, 2016	Public, CRAN	5	139	R
IBM	Causal Inference	July 12, 2019	Public, GitHub	4	138	Python
Academic	DoubleML	November 16, 2020	Public, GitHub	2	40	Python/R
Netflix XP	Causal Models	April 29, 2019	Private, Inhouse	NA	NA	Python/R

4.5 What is the future of causal inference in industry?

As the theory and results presented up to this point suggest that causal inference is an emerging topic among data scientists in industry, in the following, we focus on deriving a future outlook of where the industry is moving towards. 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 for business intelligence and their analytical capacity, many practitioners doubt their benefits for strategic decision-making (9.a). 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."* Our survey results show that while 78% of the practitioners feel positively about the impact of data science on the quality of decision-making in their company, 22% are not sure or even disagree completely. Yet the application of causal inference is seen to have the potential to overcome some of the shortcomings of machine learning in strategic decision-making (9.c). *"I think causal inference [...] has the potential to improve decision-making, at least more than machine learning does,"* TECH1 said. 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, 83% of the practitioners we interviewed believe 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 in 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). Moreover, 40% of interviewees (10.d) and 42% of survey respondents (Figure 5) want to train existing data science employees more intensively in causal inference. *"We try to level people up by teaching and providing lots of best practices and examples."* (TECH3). Of the practitioners we surveyed, 36% plan to expand their team's capabilities by hiring suitable talent. In that regard, TECH3 specified that such causal inference experts come *"almost invariably from economics."* 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 an ongoing trend in the tech sector to increasingly hire economists for data science jobs due to their specific skill set (Athey & Luca, 2019).

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



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 still 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 [that is active in causal inference], but it's not so common and that's why it's hard for us to decide whether it is worth investing in it."* Second, in light of the lack of awareness and of the unavailability of applicable, standardized tools (8.b.iv), practical causal inference methods need to become more accessible to practitioners. 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."* This gap needs to be closed so that causal inference can be applied more broadly to business decisions. Finally, interviewees stress the need to overcome important structural challenges (8.b.i) in contemporary organizations that obstruct such broader diffusion. These challenges include a lack of established processes, missing incentive structures, missing training, and the pressure on data scientists to deliver fast results in practical business environments. Importantly, this last challenge restricts the data science approaches that can be 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 it's good enough and making some impact, you go with it."* (TECH2) Given the finding that the broader organization 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 currently don't have processes to 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, therefore, our results highlight the need for skill development, better integration of methodological and business knowledge, and the establishment of suitable processes for causal data science approaches to improve organizational decision-making.

5 Discussion

The main point addressed in this study is epistemological. Strategic and organizational decision-making require choices between different courses of action (Simon, 1964), which in turn depends on causal knowledge to predict the likely outcomes of the managerial initiatives under consideration. Due to ground-breaking technological progress in the last decade, machine learning and artificial intelligence have the potential to become important inputs for optimized decision-making in modern organizations (Brynjolfsson & McElheran, 2019). However, when learning about individual cause-and-effect relationships is the goal, an adequate methodology is needed (George et al., 2014; George, Osinga, Lavie, & Scott, 2016). 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 (Bareinboim et al., 2020; Hünermund & Bareinboim, 2021; Pearl & Mackenzie, 2018). In particular, it is now 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 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"*.

Every causal model needs to be based on theory. Causal models originate in 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 provide information 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 can directly intervene in the environment she is observing, such as through A/B tests or reinforcement learning algorithms (Forney, Pearl, & Bareinboim, 2017). The problem of transportability (or external validity) remains if experimental results are to be used in contexts (e.g., temporal or geographical) that differ even slightly from the ones they have been obtained in. Solving this problem requires bringing 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 (Mullainathan & Spiess, 2017; Pearl, 2019). 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 said: "The biggest challenge [with causal inference] is that you don't believe the results. Unlike with 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 a conclusion that not just our organization, but many organizations are reaching."

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 poorly aligned with organizational goals. This development is still at its beginning, however, and knowledge about causal inference methods needs to be shared more widely outside of a relatively small group of specialists. Our empirical findings indicate that most respondents plan to invest more into causal inference capabilities. Main channels thereby constitute training measures as well as the hiring of new employees with educational backgrounds from statistics, computer science, and economics, who can contribute the required methodological skills to the organization (Athey & Luca, 2019).

Several of our interview partners further expressed their opinion that moving away from a purely correlation-based framework will be a major trend in the data science community in the next few years. Examples of causal inference initiatives in major tech firms illustrate where the industry is heading. The video streaming platform Netflix employs causal inference methods in its recommendation systems (Raimond, 2018) and rigorously runs experiments for any product change considered before it becomes a default component in the user experience (Urban, Sreenivasan, & Kannan, 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 is dedicating an increasing amount of its resources to implementing causal inference approaches as a means to improve their user experience (Harinen & Li, 2019). And the tech giant Google is using causal inference to assess how effective online advertising campaigns are in influencing search-related site visits (Brodersen, Gallusser, Koehler, Remy, & Scott, 2015; Varian, 2016).

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, Gambardella, Stern, & Zenger, 2020; Felin, Gambardella, & Zenger, 2020; Felin & Zenger, 2009, 2017). Accurate predictions of the outcomes of future actions, instrumental for effective organizational decision-making and strategic foresight, require managers to build theories. Simple data-driven approaches relying on readily observable evidence and performance feedback alone are not sufficient for developing value-creating strategies (Felin & 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, Gambardella, & Zenger, 2020; Zenger, 2016). The importance of theories—*"abstract, causal representation[s] of the world"* (Felin & Zenger, 2017, p. 262)—as an input for causal learning underscores the truth value of *"no causes in, no causes out"* (Cartwright, 1989)

At the same time, within the theory-based view, the origins of viable theories and their relations to empirical evidence and experimentation are topics that have not yet been well researched (Felin, Gambardella, Stern, & Zenger, 2020; Gavetti & 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.[14] 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 (Bareinboim et al., 2020; Pearl, 2019). As a guiding principle, this delineation

[14]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 (2021); Pearl (2009); Pearl, Glymour, and Jewell (2016); Pearl and Mackenzie (2018); Peters, Janzing, and Schölkopf (2017).

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 to inform theory (Hünermund & Bareinboim, 2021). It offers remedies if data is imperfect due to limited perception and selective observation (Bareinboim & Pearl, 2012b; Correa, Tian, & Bareinboim, 2019). And it proposes tools for managers to transport insights between various contexts (Bareinboim & Pearl, 2012c; Lee, Correa, & Bareinboim, 2020; Pearl & Bareinboim, 2011, 2014), necessary for effective theorizing.

The need for theoretical causal modeling to establish information transfer across the layers of the causal hierarchy underscores 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 critical for leveraging existing decision-making algorithms. As the bi-directionality of human and machine decisions poses challenges to organizations employing decision-making algorithms (Shrestha, Ben-Menahem, & von Krogh, 2019), the question arises of how human sensemaking and machine learning can work together to improve the generation of insights from business analytics (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 need to engage in sensemaking and interpretation of the prediction output offered by algorithms, to connect needs, problems, and alternative solutions. Similarly, Athey (2018) sates that automated prediction algorithms cannot leave domain experts out of the loop. Among scholars and practitioners, concerns remain about the identifiability of causal effects, about the confounders measured in a particular setting, about selecting the right outcome variables, and about deriving accurate strategies from (causal) relationships. In light of these challenges, causal modeling could in fact play a crucial role in the so-called "automation-augmentation paradox" (Raisch & Krakowski, 2021) that artificial intelligence poses to the management domain, and thus address concerns about the future role of managers under machine intelligence (Balasubramanian et al., 2020; Ghosh, Thomke, & Pourkhalkhali, 2020). Because of their characteristic that the crucial assumptions on which conclusions rest need to be made explicit ex-ante, causal AI methods are also able to avoid the concerns about explainability and potential fairness that come with the existing approaches to automated decision-making (Shrestha et al., 2021; Zhang & Bareinboim, 2018).

Data has become a strategic resource (Hartmann & Henkel, 2020), and this study argues that causal inference might, therefore, emerge as an important organizational capability. However, to fully develop it, top management—the originators of value-creating theories—and data scientists with their relevant technical expertise will need to work more closely together than they currently do. The results of our interviews indicate that data analytics, although appreciated for its general utility in business intelligence (Shrestha et al., 2021), is not yet well integrated into the process of organizational strategy formulation. It seems that top management often 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 (Besharov & Smith, 2014; Dunn & Jones, 2010). Thus, there is a need for interdisciplinarily trained individuals who can span the boundary between the two domains (Argote, McEvily, & Reagans, 2003; Gittelman & Kogut, 2003). This growing demand for combining deep business knowledge and strong data science skills is likely to affect business school education, which needs to incorporate more training in machine learning and causal inference methodology, including outside of dedicated business analytics programs, to develop the kind of holistic competencies that are necessary for effective data-augmented decision-making.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Press.

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

Agrawal, A., Gans, J. S., & Stern, S. (2021). Enabling entrepreneurial choice. *Management Science*. (Forthcoming)

Aguinis, H., & 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., & 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., & Penttinen, E. (2020). Challenges of explaining the behavior of black-box ai systems. *MIS Quarterly Executive*, *19*(4), 259–278.

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* (pp. 507–547). University of Chicago Press.

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

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

Athey, S., & 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., & Mangematin, V. (2013). Business models: A challenging agenda. *Strategic Organization*, *11*(4), 418–427.

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

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

Balasubramanian, N., Ye, Y., & 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., & Icard, T. (2020, July). *On Pearl's hierarchy and the foundations of causal inference* (Tech. Rep. No. R-60). Columbia University. (Forthcoming in ACM special volume in honor of Judea Pearl)

Bareinboim, E., & Pearl, J. (2012a). Causal inference by surrogate experiments: z-identifiability. In *28th conference on uncertainty in articial intelligence* (pp. 113–120).

Bareinboim, E., & Pearl, J. (2012b). Controlling selection bias in causal inference. In *Proceedings of the fifteenth international conference on artificial intelligence and statistics* (pp. 100–108).

Bareinboim, E., & Pearl, J. (2012c). Transportability of causal effects: Completeness results. In *Proceedings of the twenty-sixth AAAI conference on artificial intelligence* (pp. 698–704).

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

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

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

Besharov, M. L., & 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., & Mitchell, W. (2014). Editorial: Qualitative empirical research in strategic management. *Strategic Management Journal*, *36*(4), 637–639.

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

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

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

Bojinov, I. I., Sait-Jacques, G., & 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., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, *9*(1), 247–274.

Bryman, A. (2012). *Social research methods* (Vol. 4). Oxford University Press: Oxford.

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

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

Brynjolfsson, E., & 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., & 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., & Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. *American Economic Review*, *106*(5), 124–27.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165–1188.

Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, *41*(8), 1381–1411.

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

Correa, J. D., Tian, J., & Bareinboim, E. (2019). Identification of causal effects in the presence of selection bias. In *Proceedings of the 33rd AAAI conference on artificial intelligence* (Vol. 33, pp. 2744–2751).

Creswell, J. W., & Plano Clark, V. (2018). *Designing and conducting mixed methods research.*

third. SAGE Publications.

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

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

De Leeuw, E. D., Hox, J., & Dillmann, D. (2008). *International handbook of survey methodology* (1st ed.). Routledge.

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

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

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

Dunn, M. B., & 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., & Vaara, E. (2009). Causation, counterfactuals, and competitive advantage. *Strategic Management Journal*, *30*(12), 1245–1264.

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

Eisenhardt, K. M., & 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., & Zenger, T. (2020). Lean startup and the business model: Experimentation revisited. *Long Range Planning*, *53*(4). (In Press)

Felin, T., Gambardella, A., & Zenger, T. (2020). Value lab: A tool for entrepreneurial strategy. *Management & Business Review*. (Forthcoming)

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

Forney, A., Pearl, J., & Bareinboim, E. (2017). Counterfactual data-fusion for online reinforcement learners. In *Proceedings of the 34th international conference on machine learning* (Vol. 70, pp. 1156–1164).

Furnari, S. (2015). A cognitive mapping approach to business models: Representing causal structures and mechanisms. In *Business models and modelling (advances in strategic management, vol. 33)* (pp. 207–239). Emerald Group Publishing Limited.

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

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

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

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57*(2), 321–326.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, *59*(5), 1493–1507.

Ghosh, S., Thomke, S., & Pourkhalkhali, H. (2020). The effects of hierarchy on learning and performance in business experimentation. *Proceedings of the Academy of Management 2020*, *2020*(1), 20500.

Gillon, K., Brynjolfsson, E., Griffin, J., Gupta, M., & Mithas, S. (2012). Panel-business analytics: Radical shift or incremental change. In *Proceedings of the 32nd international conference on information systems.*

Gittelman, M., & 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., & Teodoridis, F. (2020). *Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings.* (Available at SSRN 3468822)

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

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica, Journal of the Econometric Society*, *11*(1), 1–12.

Harinen, T., & 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., & Taddy, M. (2016). Counterfactual prediction with deep instrumental variables networks. *arXiv preprint arXiv:1612.09596*.

Hartmann, P., & 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., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction.* Springer Science & Business Media.

Heckman, J. J., & 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., & 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., & Pearman, A. D. (1999). Breaking the frame: An analysis of strategic cognition and decision making under uncertainty. *Strategic Management Journal*, *20*(10), 977–985.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960.

Hünermund, P., & Bareinboim, E. (2021). Causal inference and data-fusion in econometrics. *arXiv preprint arXiv:1912.09104*.

Iansiti, M., & Lakhani, K. R. (2020). Competing in the age of AI. *Harvard Business Review*.

Imbens, G. W., & Angrist, J. (1994). Estimation and identification of local average treatment effects. *Econometrica*, *62*(2), 467–475.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

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

Kaminski, J. C., & 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 these techniques is only getting stronger"–a conversation with Pinterest's jeremy king. *Harvard Business Review*, 16–17.

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

Koning, R., Hasan, S., & Chatterji, A. (2020). Digital experimentation and startup performance: Evidence from a/b testing. *Available at SSRN 3440291*.

Koopmans, T. C. (1949). Identification problems in economic model construction. *Econometrica, Journal of the Econometric Society*, *17*(2), 125–144.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & 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., & Bareinboim, E. (2020). General transportability – synthesizing observations and experiments from heterogeneous domains. In *Proceedings of the 34th AAAI conference on artificial intelligence* (pp. 10210–10217).

Lu, C., & 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., & Bown, N. J. (2003). Cognitive mapping of causal reasoning in strategic decision making. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making* (pp. 253–272). West Sussex, England: Wiley Online Library.

Mayring, P. (2000). Qualitative content analysis. *A Companion to Qualitative Research*, *1*(2), 159–176.

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

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

Mitchell, J. R., Shepherd, D. A., & 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., & 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., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

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

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

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

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

Pearl, J., & Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. In *25th AAAI conference on artificial intelligence* (pp. 247–254). Menlo Park, CA: AAAI Press.

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

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

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York, NY: Basic Books.

Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference*. The MIT Press.

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

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, October). *Time, context and causality in recommender systems.* Netflix. (Presented October 2018 at France is AI, Paris, France)

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192–210.

Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide*. John Wiley & Sons.

Rowley, J. (2012). Conducting research interviews. *Management Research Review*, *35*(3), 260–271.

Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. *Proceedings of the IEEE*, *109*(5), 612–634.

Semadeni, M., Withers, M. C., & 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., & 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., & 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., & 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., & 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), 1–22.

Strotz, R. H., & 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*, *28*(2), 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*, *98*(2), 40–47.

Tidhar, R., & 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., & 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., & 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., & 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., ..., ., & 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*, *4*(4), 404–409.

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* (Vol. 2). SAGE Publications.

Woodward, J. F. (2003). *Making things happen: A theory of causal explanation.* Oxford, UK: Oxford University Press.

Zenger, T. (2016). *Beyond competitive advantage: How to solve the puzzle of sustaining growth while creating value.* Harvard Business Review Press.

Zhang, J., & Bareinboim, E. (2018). Fairness in decision-making – the causal explanation formula. In *Proceedings of the 32nd AAAI conference on artificial intelligence.*

Appendix

A 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?

B 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
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

– Continued from previous page –

ID	Subcode	Definition	Freq.
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
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

– Continued from previous page –

ID	Subcode	Definition	Freq.
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
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
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

– Continued from previous page –

ID	Subcode	Definition	Freq.
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
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

– Continued from previous page –

ID	**Subcode**	**Definition**	**Freq.**
	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

C 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	
Other (please specify)	

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	

Correlation versus causation

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

	Selection
Yes	
No	
Not sure	

Causal inference in data science

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 design	
Difference-in-differences	
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 specify)	

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 specify)	

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)	

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

Q26: 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	

Q27: 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	

D Descriptives

Characteristics of survey respondents ($n = 234$)

		Frequency	Proportion
Job Affiliation	Academic institution	50	21.5%
	Private sector	159	68.2%
	Both	13	5.6%
	Other	11	4.7%
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%
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%
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%

E Technical appendix on coding process

All interviewees were deemed as equally important to the research. Interviews were conducted in English[15]. A guide of twelve open-ended questions with one to four sub-questions each was 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. 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 (Mayring, 2000; Morris, 1994; Weber, 1990). 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 coding frame in Appendix B) 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 (Cho & Lee, 2014; Guthrie, Petty, Yongvanich, & Ricceri, 2004). 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 (Mayring, 2000; Weber, 1990). The final coding frame (Appendix B) 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 researchers in the process. Disagreements were discussed and resolved by conceptual clarification.

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

F Practical causal inference applications

ID	Excerpt
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?
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 percent of the cases which drive 80 percent of the value?"
MANU1	"For example, I have a large typical 99 percent accuracy where my system works and I want to analyze the 1 percent 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."

– Continued from previous page –

ID	Excerpt
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 membership program. We can't actually A/B test a membership 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."

G Additional survey results

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

Mean (SD)	Not at all important	Slightly important	Moderately important	Very important	Extremely important
3.19 (1.3)	12%	18%	26%	26%	18%

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"?

Mean (SD)	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
3.09 (1.3)	12%	30%	11%	29%	17%

Figure 8: (Q16-19) How important are the following (dis)advantages of observational / experimental methods for your data science projects?

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

	Mean (SD)	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Relatively cheap	3.27 (1.2)	11.1%	13.0%	29.0%	31.5%	15.4%
Large sample size possible	3.45 (1.1)	8.0%	11.7%	23.5%	40.7%	16.0%
High external validity	3.38 (1.1)	6.9%	12.5%	30.0%	37.5%	13.1%
Easy to implement	3.40 (1.2)	6.8%	17.9%	20.4%	38.9%	16.0%
Based on actual field data	3.71 (1.1)	4.9%	10.5%	20.4%	37.0%	27.2%

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

	Mean (SD)	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Too many assumptions	3.43 (1.1)	6.2%	13.8%	28.7%	33.1%	18.1%
Time−consuming	3.16 (1.2)	9.7%	17.4%	32.9%	27.1%	12.9%
Require specific skills	3.06 (1.2)	11.9%	21.4%	27.0%	28.3%	11.3%
Require specific data sets	3.33 (1.2)	8.3%	17.8%	23.6%	33.1%	17.2%
Difficult to implement	2.89 (1.1)	12.0%	26.6%	27.2%	29.1%	5.1%
Difficult to explain	3.04 (1.2)	11.0%	23.2%	29.0%	24.5%	12.3%

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

	Mean (SD)	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Require no specific skills	2.89 (1.2)	16.3%	17.6%	35.3%	22.2%	8.5%
Require few assumptions	3.30 (1.1)	9.2%	11.8%	34.0%	30.1%	15.0%
Easy to interpret	3.65 (1.2)	7.8%	8.5%	17.0%	43.8%	22.9%
Easy to implement	3.20 (1.2)	10.6%	18.5%	23.2%	35.8%	11.9%

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

	Mean (SD)	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Relatively costly	3.27 (1.2)	11.1%	15.7%	26.1%	28.8%	18.3%
Not possible in our domain	3.06 (1.5)	23.1%	10.9%	26.3%	16.7%	23.1%
Lack suitable outcome metrics	3.38 (1.1)	6.9%	12.5%	30.0%	37.5%	13.1%
Lack of external validity	3.03 (1.2)	15.5%	17.4%	25.8%	31.6%	9.7%
Ethical concerns	2.83 (1.4)	23.7%	22.4%	17.8%	19.7%	16.4%

Figure 9: (Q21) Which causal inference tools / software are you aware of?

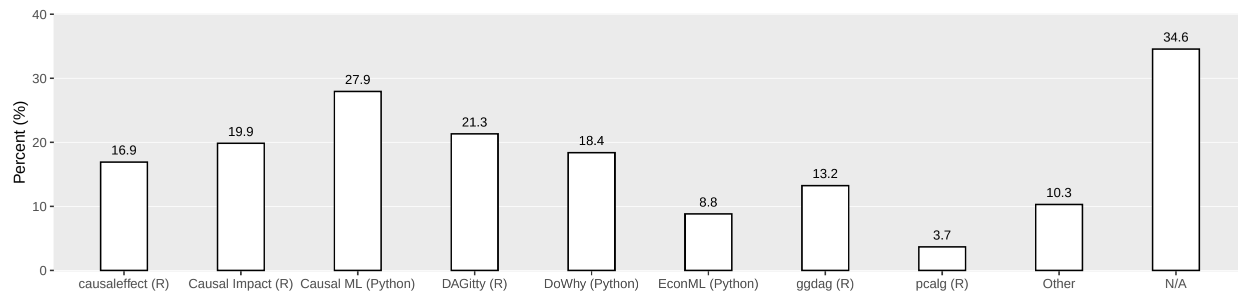


Figure 10: (Q22) How suitable do you find existing causal inference tools / software libraries for your purposes?

Mean (SD)	Not at all suitable	Slightly suitable	Moderately suitable	Very suitable	Extremely suitable
2.83 (0.94)	8.67%	27.33%	36.67%	26.67%	0.67%

Figure 11: (Q25) How important are the following disciplines / educational backgrounds of employees for improving the causal inference capabilities of your organization?

	Mean (SD)	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Statistics	4.15 (0.72)	0.0%	2.0%	13.4%	51.7%	32.9%
Social sciences	3.01 (1.21)	11.9%	25.2%	25.2%	25.8%	11.9%
Psychology	2.28 (1.21)	35.1%	25.7%	18.2%	17.6%	3.4%
Natural sciences	2.40 (1.18)	31.3%	20.0%	28.7%	17.3%	2.7%
Mathematics	3.33 (1.03)	4.1%	16.2%	35.1%	31.8%	12.8%
Engineering	2.94 (1.26)	16.0%	22.7%	24.0%	26.0%	11.3%
Economics	3.38 (1.32)	12.7%	14.0%	19.3%	30.7%	23.3%
Computer science	3.41 (1.16)	6.7%	15.4%	27.5%	30.9%	19.5%

Figure 12: (Q26) Do you think that data science is improving human decision making in your organization?

Mean (SD)	Definitely not	Probably not	Might or might not	Probably yes	Definitely yes
4 (0.86)	1.3%	3.9%	17.0%	49.0%	28.8%

Figure 13: (Q27) 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"?

Mean (SD)	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
4.24 (0.99)	2.6%	4.6%	9.8%	32.7%	50.3%

References

Bryman, A. (2012). *Social research methods* (Vol. 4). Oxford University Press: Oxford.

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

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

Mayring, P. (2000). Qualitative content analysis. *A Companion to Qualitative Research*, *1*(2), 159–176.

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

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research.* SAGE Publications.

Weber, R. P. (1990). *Basic content analysis* (Vol. 2). SAGE Publications.