Causal Inference as an Organizational Problem*

Michael D. Ryall
University of Toronto

Olav Sorenson
UCLA

May 29, 2021

Abstract

Research on organizations generally presumes that their leaders have the ability to direct the organization towards a set of goals. That presumption, however, depends crucially on the ability of leaders to understand how particular actions or directives might influence organizational outcomes, a problem of causal inference. We develop a formal model of this problem in which managers attempt to infer the effects of possible directives based on what they can observe. Our model reveals that inference only becomes a problem under specific conditions but these conditions seem common in real-world settings. We discuss how a variety of organizational features might mitigate this inference problem.

*Jerker Denrell, Rem Koning, Jesper Sørensen, and the attendees of the the Berlin New Directions in Alliance Governance Conference, the Nagymaros Conference, the Wharton Conference on Corporate Strategy and Innovation, and the participants of seminars at Duke, Harvard, Purdue, the University of Maryland, the University of Utah, and Washington University provided insightful comments on earlier versions of this paper. The usual disclaimer applies.

Organizations have long been thought to solve a problem of coordination, allowing groups of people to cooperate to achieve ends that none of them could on their own [Parsons, 1960]. One sees this idea across the social sciences. Among early sociologists, Weber [1968], for example, viewed organizations as means to ends, analyzing the structural features that allowed them to operate effectively. Early organizational theorists, such as the administrative and scientific management schools, focused on how to optimize the structures of organizations to coordinate the activities of their employees [Taylor, 1911, Simon, 1945]. Economic theories of the firm argue that organizations facilitate joint production that might not otherwise occur when individuals trying to collaborate through trade find it too difficult to allocate their jointly-created value in a way that impels everyone to do their part [Coase, 1937, Alchian and Demsetz, 1972].

Yet the mere existence of an organization, even a formal one, does not guarantee coordination among those affiliated with it. Political perspectives on organizations, for example, highlight the fact that decisions within a firm and the resultant firm behavior arise from internal negotiations among those who hold important information and who control valuable resources [March, 1962, Pfeffer and Salancik, 1974]. Others have noted that organization members whose own goals deviate from those of the organization may divert its resources for personal gain [Ross, 1973, Jensen and Meckling, 1976, Williamson, 1981].

We call attention to another potential impediment to coordination, the absence of a discernible connection between the actions of the members of the organization and the consequences of those actions for organizational outcomes. Even if everyone has a common goal, coordination requires that the individuals involved, or at least those leading them, have a sufficient understanding of what each must do to achieve that end. In essence, leaders must understand the causal relationships between their directives and organization-level outcomes.

In contrast, nearly all existing analyses of organizations begin from the assumption that leaders understand these causal connections. Even when they have incomplete information, they effectively know what they do not know.[1] Managers, for example, might not be able

[1] The classic literature on bounded rationality meanwhile raises a somewhat different issue [March and Simon, 1958, March, 1978]. March and Simon [1958], for example, consider coordination in settings where organization members understand the underlying causal connections but where the complexity of those relationships and their interactions strains the cognitive capacity of humans to identify the ideal course of action.

to observe the amount of effort being exerted by their subordinates but they nevertheless have been assumed to understand precisely how that effort should relate to organizational outcomes. When leaders understand these causal relationships, they can usually solve the coordination problem stemming from such uncertainty through monitoring or by designing incentive systems that encourage organization members to act in ways that forward organizational goals [e.g., Grossman and Hart, 1986, Milgrom and Roberts, 1992].

But what happens if leaders cannot even determine what they do not know? To answer this question, we develop a formal model, a mathematical representation, of this problem, which allows us to specify the conditions under which managers – at least in principle – have sufficient information to infer the effectiveness of a directive or of an organizational policy.[2]

We model the organization and its environment as a (finite but potentially large) set of variables that produce some organizational outcome of interest, such as sales or profits, based on a specific pattern of causal relationships.[3] These variables represent both internal factors, such as employee actions or automated processes, and external factors, such as customer preferences or competitor activities.

Our model presents a "best-case" scenario, in many ways, for managers. They can see what workers do, thereby assuming away monitoring problems. The employees in the model do not necessarily have divergent goals, eliminating agency issues. We consider organizations as small as one employee and with only a few causal paths between the variables. Complexity therefore should not pose a problem. Our managers can also observe both the direct consequences of employee actions and overall organizational performance. If causal inference fails under such generous assumptions – and, indeed, it does – it would undoubtedly fail under more challenging conditions.

We introduce one novel element to these models, not generally present in prior research:

Organizations, by dividing goals into subgoals, by restricting the information necessary to achieve those subgoals, and by providing a structure for aggregating subgoals, have been seen as a means of ameliorating these cognitive constraints [Simon, 1945, 1962].

[2]We refer to the person trying to understand these effects as the "manager" and those carrying out the activities of the organization as "employees"; our results nevertheless extend to communities and informal structures without any formal authority.

[3]We adopt the Bayesian network approach to modeling causal systems [e.g., Pearl, 2009]. Many readers will recognize this approach as the basis for analyzing empirical identification strategies as directed, acyclic graphs [Morgan and Winship, 2007, Elwert, 2013, Elwert and Winship, 2014].

One or more of the factors that influence the organization and its performance sit outside of the awareness of the managers. They remain hidden.

We explore when these hidden factors confound accurate assessment of the causal relationships between the actions of employees and organizational performance. Even under this best-case scenario, hidden factors can confound the ability of the manager to infer these causal connections.

That hidden factors could impede causal inference seems unsurprising. What is surprising is that, regardless of the number of factors or the complexity of their relationships to production, hidden influences only prevent this inference when the causal connections have a specific local structure. From a purely mathematical perspective, our analysis reveals that the vast majority of possible patterns relating hidden influences to organizational performance still permit causal inference.

From a practical perspective, however, our analysis points to a potential problem: The specific structure in which a hidden factor stymies inference seems common – perhaps even pervasive – in real-world settings. Nearly all prior research, however, effectively presupposes that this problem of causal inference has somehow already been solved. But if managers cannot discern the effects of a directive, they cannot govern the organization, they cannot coordinate the employees.

The existence of this inference problem should not surprise social scientists. Much of the focus of empirical research and methodology over the past two decades has been aimed at understanding whether particular attributes, events, and policies have causal effects on outcomes of interest [e.g., Morgan and Winship, 2007, Angrist and Pischke, 2009]. Even with methodological advances, such as the use of instrumental variables, causal inference often proves difficult.

Curiously, however, that recognition of the difficulty of causal identification in empirical research has not carried over to our theoretical understanding of social and organizational behavior. Instead, the almost universal – but usually implicit – assumption has been that organization members understand the causal relationships leading from their actions and policies to organizational outcomes.

Despite the pessimistic predictions of this result, organizations exist and appear to coordinate their members effectively. They represent the dominant means of producing goods, services, and other forms of collective action. Most organizations therefore appear to have found solutions to this problem.

Based on our analysis, we see three possible paths for mitigating the causal inference problem: (i) through direct discovery of the effects of individual behaviors (experimentation), (ii) by giving the manager insight into the hidden factors (illumination), or (iii) by replacing the problematic effects of hidden factors with dependencies on public factors (substitution).

In the final section, we consider a variety of organizational features through this lens. These features may alleviate the managerial inference problem through experimentation, illumination or substitution. Given that organizations generally appear able to coordinate their members despite the pervasive potential for managerial inference problems, extensions of our model – and its causal systems perspective on organizations – could shed new light on how and why certain features of organizations contribute to their success.

Organizations as causal systems

Organizations come in a wide variety of forms, from the formal for-profit entities that dominate the production and distribution of goods and services to the non-profit organizations that support various communities and even to informal groups of individuals. Across all of these diverse forms, the individuals involved, the members, often attempt to coordinate their actions to attain group-level goals. One of more of these members usually facilitates this coordination by playing the role of a leader, someone with the authority to implement policies and directives to forward these goals. Choosing the right policies and directives, however, first requires leaders to understand the causal relationships between what members do and organization-level outcomes.

To understand better the nature of this inference problem, we develop a formal model based on Bayesian network theory. Over the last thirty years, our understanding of causal identification has grown by leaps and bounds. An important element in that progress has been the efforts of Judea Pearl and others to develop a mathematical language for representing and analyzing stochastic causal systems [e.g., Pearl, 1988, 2009]. The Bayesian network approach depicts the causal system as a directed, acyclic graph (DAG), in which the nodes represent random variables of interest and the directed edges indicate influence relationships. A causal relationship exists when the probability distribution of a random variable (the effect) depends upon the state of another random variable (the cause). When an edge points from a cause to an effect, a "direct" relationship exists (as opposed to an indirect effect operating through a sequence of direct relationships). A variable may have many direct causes.

Building on this infrastructure, our model treats the organization and its environment as a system of causal relationships between factors, represented by variables.[4] In any organization, a variety of factors interact to produce the organization-level outcomes. Some of these variables represent the internal operations and attributes of the organization, such as the actions of its members, the consequences of those actions, and the goals of the collective. Some variables represent factors external to the organization, those exerting an influence from the environment. Examples of environmental factors include the actions and preferences of buyers and suppliers, the institutions interacting with the organization, and the competitive and cooperative actions of other organizations.

Other than to limit them to a finite number of states, we do not restrict the distributions of these variables.[5] We simply assume that a variable has states that correspond to the actual states of the underlying factor it represents. Thus, for example, a variable could be real-valued if it represented something like the output of a machine or the dollars allocated to inventory. But variables can also represent factors with qualitative states, such as the mood of an employee or the legal environment within which the organization operates.

Although the internal versus external distinction has relevance to which factors the manager has some ability to control, it does not matter to the inference problem. We therefore do not distinguish between internal and external factors in the model.

[4] In other words, we refer to real-world objects as "factors" and to the mathematical objects representing them in our model as "variables."

[5] Our theory can be extended to continuous variables. However, this extension would come at the cost of significantly increased mathematical complexity with little added in the way of additional insights.

The key distinction for inference is whether the manager understands the (stochastic) behavior of a variable. We label as *open* (to the manager) those variables corresponding to factors about which the manager has insight. Such insight might come from the manager being able to observe the variable directly or having access to reporting on its state. But such insight might also arise in a more indirect manner. Observing an organization over a long period of time, for example, could allow a manager to develop an accurate intuition for its behavior.

Simply being unobserved, however, does not mean that we would consider a variable closed. Managers could have the ability to discern the states of variables they cannot observe from their influence on the behavior of those they can. For example, if a manager had a good sense of the distribution of a particular aptitude in the population, observing the behavior of her employees might allow her to infer the qualitative state of this variable for her own employees. We assume that the performance of the organization, in terms of meeting its objectives, belongs to the set of open variables—that is, the manager can observe it.

By contrast, we label as *closed* variables representing factors that are hidden to the manager. These factors may prove elusive for a variety of reasons. Managers, for example, may not recognize that they exist. Even if conscious of them, they may not believe that they influence organizational behavior – paying them no attention as a result. Even if aware of the factors and their relevance, managers may have no means of collecting reliable information on them. Managers, for example, understand that other members of the organization have their own beliefs and personal priorities. Yet the primary source of insight into those beliefs are the employees themselves, who often have a vested interest in not revealing them.[6]

Together, the open and closed variables form a comprehensive system, which represents all of the internal and external factors relevant to the performance of the organization. We refer to this system of open and closed variables, connected by a set of influence relations, as the *objective model* of the organization. As explained in detail below, the objective model implies a joint probability distribution (JPD) on these variables.

[6]Agency problems – a divergence of goals between managers and employees – represent a special case of our model. As we will demonstrate, the extent to which these diverging interests prove problematic to causal identification depends on the overall pattern of influence relations in the objective model of the organization.

We also assume that the manager has some minimal appreciation of the causal ordering of factors relative to organizational performance and has awareness of the marginal JPD on open variables. The first of these assumptions seems relatively innocuous. We simply mean that the manager can distinguish which factors cause organizational performance (either directly or indirectly) versus which ones stem from it. If the manager had a more extensive understanding of the influence relationships between open variables – all the way up to having an accurate understanding of all of them – it would not change our results.

The second assumption deserves more attention. The *marginal* JPD on open variables corresponds to the JPD on all variables after "integrating out" the closed variables (i.e. summing over their states). In other words, the manager can determine the joint states of all the open variables and how frequently those joint states occur. This assumption equips managers with an unusual degree of insight. However, as noted earlier, our aim has been to build a model as favorable to the manager as possible. If inference fails under this assumption, it must surely fail under less generous ones.

Given this information, can the manager infer the causal effects of employee actions on organizational performance? If so, then the manager can determine which actions to encourage through policies and directives. We begin by illustrating the intuition behind our general analysis with three simple examples.

Three simple cases

Each of our simple cases involves a manager trying to assess how the actions of a single employee affect the performance of an organization. In each situation, the manager has the intention of directing the employee to engage in the action most beneficial to the organization's goal. Before she can decide how to direct the employee, however, she must first determine what she would want the person to do.

Each of these simple cases involve three open variables and one closed one. In each case, the open variables – the ones that the manager can observe or has insight into – are the actions of the employee, the immediate outcome related to those actions, and organizational performance. Intuitively, the actions of the employee influence the immediate outcome which, in turn, has consequences for organizational performance.

In all three cases, the closed variable, which remains hidden to the manager, influences the actions chosen by the employee. In the second and third cases, the closed variable also influence the behavior of other variables in the system.

Case I: An innocuous closed factor

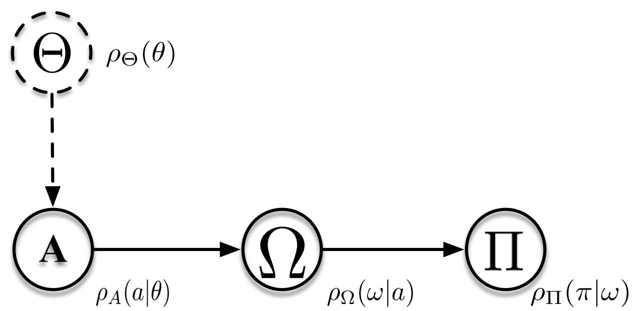

Figure 1: A graphical model of the first case.

Figure 1 provides a graphical representation of the first case. We depict the open variables as solid circles and the closed variable as a dashed one. In this example and in those following it, A denotes the employee's action, Ω the intermediate output associated with those actions, and Π organizational performance. We denote states of these variables with small letters (e.g., $a \in A$). To keep the discussion as simple as possible, assume that all of the variables have only two possible states: 0 and 1.

Although the model appears simple, these open variables could represent a plethora of internal processes. For example, the employee's action might reflect the number of orders filled on a particular day, the intermediate factor might capture the accuracy of order fulfillment, and performance might measure revenue less the cost of labor and shipment corrections. Or, the employee might need to decide whether or not to offer a discount to a customer segment, which influences the sales of the organization in that segment, which in turn determines the overall profitability of the firm.

The closed variable, Θ, represents a factor that influences the actions of the employee but which remains hidden to the manager. This variable could also represent a range of factors. It could reflect something as simple as whether the employee got stuck in traffic on the way to work, influencing the person's performance throughout the day. It could represent something about the employee's general attitudes or opinions, such as his tendency to act altruistically or his beliefs about the importance of customer satisfaction. It could capture a rule or strategy the employee has chosen to determine a course of action. In the case of the employee deciding whether to offer a discount, it might capture his beliefs about the customer's price sensitivity.

Although some of these examples involve factors that the employee might understand even if the manager does not, closed factors could remain hidden to both the manager and the employee. The employee, for example, may not appreciate that traffic in the morning affects his own performance throughout the day.

The arrows depict the *influence relationships* between the factors: $X \to Y$ means that, for at least one pair of states $x \neq x'$ of X, the associated probability distributions on the states of Y differ. For example $A \to \Omega$ means that, $a = 0$ versus $a = 1$ results in different probabilities that $\omega = 0$ (and, by implication, that $\omega = 1$). We can describe these influence relationships locally through the use of conditional probability tables (see Figure 2). We use ρ_X to represent the local stochastic law that governs the state of factor X. For example, ρ_Π describes the stochastic effect of the intermediate factor (Ω) on organizational performance (Π). In our example, when $\omega = 1$, the probability that $\pi = 0$ is 0.60.

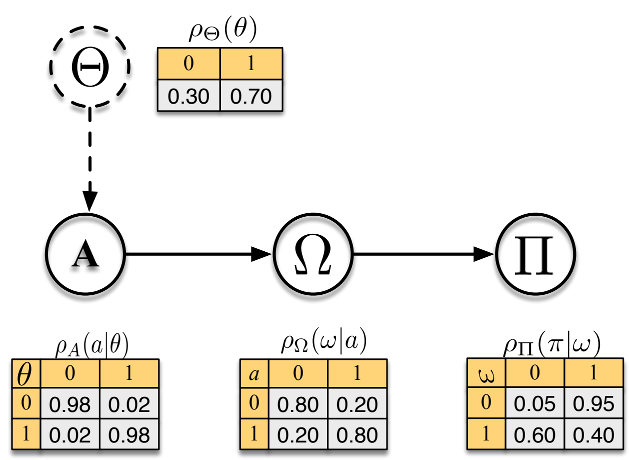

Figure 2: Parameters for the first case.

We can use the relationships defined by the conditional probability tables in Figure 2 to calculate a JPD over all of the variables contained within the system (i.e. the probabilities of every possible configuration of open and closed variables). We can compute the JPD from the following equation:

$$P(\theta, a, \omega, \pi) = \rho_\Pi(\pi|\omega)\rho_\Omega(\omega|a)\rho_A(a|\theta)\rho_\Theta(\theta). \tag{1}$$

Table 1, Part (a) reports the JPD arising from the causal system described in Figure 2. We refer to this JPD as the *status quo* distribution because it describes the behavior of the system absent any intervention by the manager.[7]

Recall, however, that the manager only has access to the marginal JPD on the *open* variables, $P(a, \omega, \pi)$, which we label as the *MDO*. Table 1, Part (b), reports the MDO for this example. To compute the MDO, first note that, by the definition of conditional probability, we can factor the JPD according to:

$$P(\theta, a, \omega, \pi) = P(\pi|\omega, a, \theta)P(\omega|a, \theta)P(a|\theta)P(\theta). \tag{2}$$

From Equation 2, we then calculate the values in Part (b) by applying the law of total probability to "integrate out" the closed variable:[8]

$$P(a, \omega, \pi) = \sum_{\theta \in \Theta} P(\pi|\omega, a, \theta)P(\omega|a, \theta)P(a|\theta)P(\theta). \tag{3}$$

To summarize the model in formal terms, the primitive in our analysis is a comprehensive causal system that explicates the influence relationships between the factors and quantifies their local stochastic laws (e.g., Figure 2). We derive the JPD on all of the factor variables from these details. Once the JPD has been elaborated, we can derive the MDO from it. We

[7]This enumeration of possibilities allows us to calculate any probability for the system. For example, the probability that the employee chooses $a = 1$ is 0.692 and the probability that the employee chooses $a = 0$ conditional on $\theta = 1$ is 0.98. Due to rounding, the probabilities reported in the JPD tables do not always sum exactly to one.

[8]The notational pivot from the local stochastic laws in Equation 1 to the conditional probabilities, computed from the JPD, in Equations 2 and 3 highlights the fact that, once the JPD has been calculated, the calculation of the MDO does not require information about the model's parameters.

do not need information on the structure of the causal system to compute the MDO. We assume that the manager knows: 1) the MDO; and 2) that the employee's action, the focus of a potential intervention, sits upstream from organizational performance in the causal chain.

(a) JPD

Θ	A	Ω	Π	P
0	0	0	0	0.0118
0	0	0	1	0.2235
0	0	1	0	0.0353
0	0	1	1	0.0235
0	1	0	0	0.0001
0	1	0	1	0.0011
0	1	1	0	0.0029
0	1	1	1	0.0019
1	0	0	0	0.0006
1	0	0	1	0.0105
1	0	1	0	0.0017
1	0	1	1	0.0011
1	1	0	0	0.0069
1	1	0	1	0.1304
1	1	1	0	0.3293
1	1	1	1	0.2196

(b) MDO

A	Ω	Π	P
0	0	0	0.0123
0	0	1	0.2340
0	1	0	0.0370
0	1	1	0.0246
1	0	0	0.0069
1	0	1	0.1315
1	1	0	0.3322
1	1	1	0.2215

(c)

$$
\begin{aligned}
P(\pi = 1) &= .6116 \\
P(\pi = 1|a = 1) &= .5100 \\
P(\pi = 1|a = 0) &= .8400
\end{aligned}
$$

Table 1: Part (a) JPD generated by the organization in Figure 2. Part (b) provides the marginal distribution on open variables (MDO) implied by the JPD. Part (c) reports the probability that $\pi = 1$, as well as the conditional probabilities that $\pi = 1$ given $a = 1$ and $a = 0$, respectively.

Can the manager manage? Assume that the manager wants to maximize the probability of $\pi = 1$ and that she has the ability to direct the actions of the employee (i.e. to mandate a specific action).[9] The manager therefore need not worry about miscommunication or a misalignment of incentives leading to a disconnect between what she wants the employee to do and what the employee actually does. This again represents a best-case scenario for the manager: she has a complete and accurate understanding of the statistical behavior of the open variables and the ability to ensure that the employee engages in a particular action.

[9]The model can readily be extended to cases where the manager chooses policies to influence the employee's choice of action instead of directly decreeing his actions.

Does she have sufficient information to assess the effect of directing the employee to engage in a particular action? Yes, she does. But why?

The MDO suggests that directing the employee to do $a = 0$ should increase the probability of the desired outcome $\pi = 1$, whereas mandating $a = 1$ would decrease it. Part (c) of Table 1 reports the probability of $\pi = 1$, both overall and conditional on the actions of the employee *as computed from the MDO.*

Will the conditional probabilities rom the MDO reliably predict the expected effect of a managerial interventions? Not necessarily. The conditional probabilities in the MDO reflect the default behavior of the system – the status quo – *absent* any intervention. Interventions change the underlying system. In some cases, those changes mean that the conditional probabilities from the status quo system would not predict the effects of an intervention.

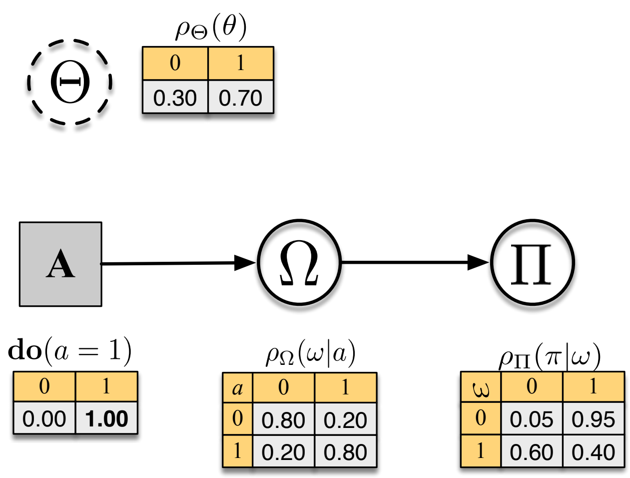

Figure 3: The effect of the intervention $\mathbf{do}(a = 1)$.

Figure 3, depicts the system when the manager directs the employee to do $a = 1$. We represent this intervention as a gray box. In the conditional probability table, we replace the local stochastic law parameters with the bold **do** operator, indicating that the employee does $a = 1$ with a probability of one. The directive removes the influence of Θ on the employee's actions (A) since the employee must do $a = 1$ regardless of the state of Θ.

With the system adjusted to reflect the intervention, we can calculate the JPD ($P_{\mathbf{do}(a=1)}$ in our notation). From it, we can also calculate the MDO for this modified system. Table 2 reports this JPD, in Part (a), and its associated MDO, in Part (b). We can use this information to calculate the effect of the intervention: $P_{\mathbf{do}(a=1)}(\pi = 1) = .5100$.

(a) JPD for $\mathbf{do}(a = 1)$

Θ	A	Ω	Π	$P_{\mathbf{do}(a=1)}$
0	0	0	0	0.0000
0	0	0	1	0.0000
0	0	1	0	0.0000
0	0	1	1	0.0000
0	1	0	0	0.0030
0	1	0	1	0.0570
0	1	1	0	0.1440
0	1	1	1	0.0960
1	0	0	0	0.0000
1	0	0	1	0.0000
1	0	1	0	0.0000
1	0	1	1	0.0000
1	1	0	0	0.0070
1	1	0	1	0.1330
1	1	1	0	0.3360
1	1	1	1	0.2240

(b) MDO for $\mathbf{do}(a = 1)$

A	Ω	Π	$P_{\mathbf{do}(a=1)}$
0	0	0	0.000
0	0	1	0.000
0	1	0	0.000
0	1	1	0.000
1	0	0	0.010
1	0	1	0.190
1	1	0	0.480
1	1	1	0.320

(c)

$P_{\mathbf{do}(a=1)}(\pi = 1) = .5100$

Table 2: Part (a) details the JPD generated by the intervention shown in Figure 3. Part (b) provides the corresponding MDO. Part (c) reports the probability that $\pi = 1$.

Returning to our original question: Does the manager have sufficient information to predict the effects of an intervention? Yes, she does. Compare the conditional probabilities in Part (c) of Table 1 with the result shown in Part (c) of Table 2. Notice that that $P_{\mathbf{do}(a=1)}(\pi = 1) = P(\pi = 1|a = 1)$. The conditional probability of π given $a = 1$ in the status quo system correctly predicts the effect on π of directing the employee to do $a = 1$. Going through a similar analysis, we can demonstrate that the effect of $\mathbf{do}(a = 0)$ also matches the conditional probability in the status quo MDO ($P_{\mathbf{do}(a=0)}(\pi = 1) = P(\pi = 1|a = 0)$). Even though the system includes a hidden factor, the data on open variables allows for an accurate assessment of the effects of directing the employee to engage in a specific action on organizational performance .

Although this example uses only binary variables with a specific set of conditional probabilities, this result generalizes to variables with any (finite) number of states and any set of conditional probabilities. The result stems from the structure of the system rather than from a careful choice of parameters. To understand better why the employee actions have the same effect when mandated by the manager as predicted by the MDO, let us return to the status quo system. Without an intervention, organizational performance varies with the closed variable: $P(\pi = 1|\theta = 1) = 0.9985$ but $P(\pi = 1|\theta = 0) = 0.8335$. However, this dependence only arises indirectly through the sequential causal chain $\Theta \rightarrow A \rightarrow \Omega \rightarrow \Pi$. Because this indirect influence flows through the actions of the employee, Π is *conditionally independent of* Θ *given* A in the JPD generated by the status quo system. Computing the probability of Π given A from the MDO therefore provides an accurate assessment of the causal effect of the employee's action on organizational performance. Even though Θ remains closed to the manager, the MDO provides sufficient information for her to assess the effect of directing the employee to do a particular action. Consistent with the literature on Bayesian Network theory, we label the effect of A on Π as *identifiable.*[10] We summarize this claim in our first proposition (for the proof, see the Appendix).

Proposition 1. Given a collection of variables with an arbitrary number of $K \geq 2$ states each and parameterized by any strictly positive stochastic laws consistent with Figure 1, the effect of an intervention $\mathbf{do}(a)$ on organizational performance π is identifiable for all $a \in A$.

Case II: Confounding common causes

Although Proposition 1 seems like good news for the manager – indeed, it is – that result depends on a specific pattern of influence relationships between the open and closed variables: one in which the closed variable only influences performance through the actions of the employee. Consider instead a case in which both the actions of the employee and the immediate consequences of these actions depend on a common closed variable.

Figure 4 depicts such a system.[11] In this case, the employee might represent a purchasing agent who has private information about the reliability and timeliness of a particular supplier. Based on that information, the employee might change levels of inventory or open

[10] See, for example, Pearl [2009].

[11] This structure also corresponds to the setting used in many economic analyses of "principal-agent" problems – the employee, the agent, has private information about the efficacy of his actions on intermediate factors, such as product quality, cost, or quantity which, in turn, affect profitability.

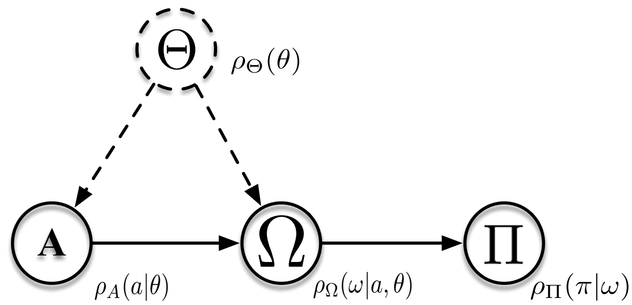

Figure 4: An organization with a confounding variable.

relationships with additional suppliers. The reliability and timeliness of the supplier, in turn, could influence the costs associated with maintaining a particular inventory level or the benefits of spreading purchasing across more than one supplier.

Can the manager still predict the effect of an intervention in this case? No, she cannot. Although the causal structure in Figure 4 only differs from the previous example by the addition of a single edge (connecting Θ to Ω), the favorable result of Proposition 1 no longer holds. In this case, Θ exerts a *confounding* effect on the system. The manager can no longer predict the effect of intervening on the employee's actions using the MDO.

To illustrate the intuition behind this result, let us, once again, consider a specific numerical example. Figure 5 reports a set of local stochastic laws that describe the stochastic behavior of this organization. We can again use these conditional probabilities to construct the overall JPD:

$$P(\theta, a, \omega, \pi) = \rho_\Pi(\pi|\omega)\rho_\Omega(\omega|a,\theta)\rho_A(a|\theta)\rho_\Theta(\theta). \tag{4}$$

Although this equation appears similar to that of (1), it differs from it in one crucial respect: the closed variable θ now appears as an argument in the stochastic law ρ_Ω.

The manager no longer has the ability to predict the effects of an intervention. The problem stems from the fact that both the employee actions and their immediate outcomes now depend directly on the hidden factor. In the JPD generated by this system, Π is no longer independent of Θ given A. The MDO therefore no longer provides sufficient information to isolate the effects of the employee actions.

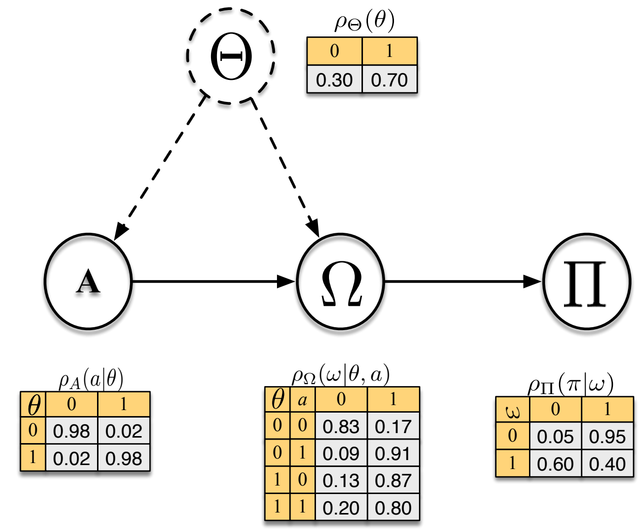

Figure 5: Parameters for the organization in Figure 4.

In the interest of brevity, we will not go through all of steps in the calculations again. However, although the JPD associated with this system differs from the one presented in Table 1 Part (a), *it has an identical MDO* (Table 1 Part (b)). Thus, referring to Part (c) of that table, $P(\pi = 1) = 0.6116$, $P(\pi = 1|a = 1) = .5100$, and $P(\pi = 1|a = 0) = .8400$.

What happens when the manager intervenes now? By proceeding as we did in the previous case, we can calculate that $P_{\mathbf{do}(a=1)}(\pi = 1) = 0.4927$ and $P_{\mathbf{do}(a=0)}(\pi = 1) = 0.5871$. As in the previous example, the MDO suggests that, of the three options available to the manager ($\mathbf{do}(a = 1)$, $\mathbf{do}(a = 0)$, or no intervention), the manager should choose $\mathbf{do}(a = 0)$. In reality, however, because $P(\pi = 1) = 0.6116$ under the status quo, the organization would perform better if the manager did not intervene.

Why would the manager do better by not intervening? In the status quo system, the employee reacts to private information about a closed factor that partially determines the effectiveness of his actions. Importantly, the employee responds to this information in a manner beneficial to the organization. Any policy or directive to the employee that does not depend similarly to the state of the hidden factor cannot do as well. The MDO, however, does not allow the manager to separate out these effects. The manager cannot isolate the effect of A on Ω.

Even the fact that the manager would prefer not to intervene, however, stems only from a specific set of conditional probabilities. Figure 6, for example, presents an alternative set of parameters that generates exactly the same pattern from the perspective of the manager – the same MDO – as the system in Figure 5. In this case, however, an intervention *would* improve performance. This indeterminacy holds for any organizational system with the structure shown in Figure 4. We summarize this result in our second proposition (for the proof, see the Appendix).

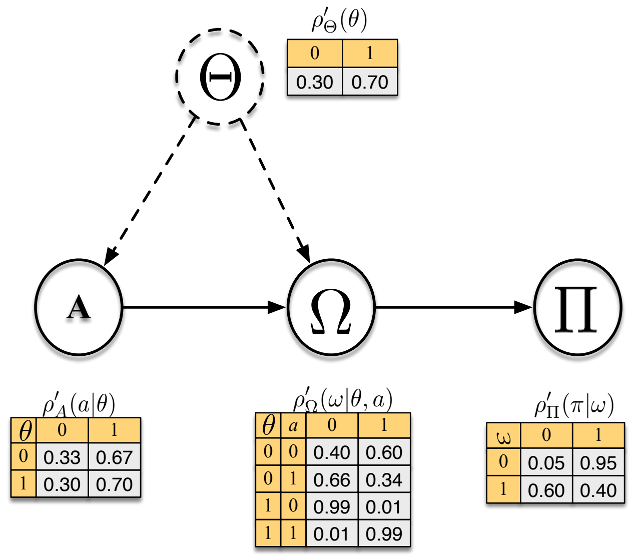

Figure 6: A second set of parameters for the organization in Figure 4.

Proposition 2. Given a collection of variables with an arbitrary number of $K \geq 2$ states each and parameterized by any strictly positive stochastic laws consistent with Figure 4, the effect of an intervention $\mathbf{do}(a)$ on organizational performance is not identifiable for any $a \in A$.

Case III: Non-confounding common causes

Given the previous case, one might expect that any closed common cause – that is, any case in which a closed variable influences organizational performance through more than one path – would preclude causal identification. But that intuition turns out to be wrong.

Consider the system depicted in Figure 7, with closed factor Θ that directly influences both the employee's action and organizational performance. Following the same procedures

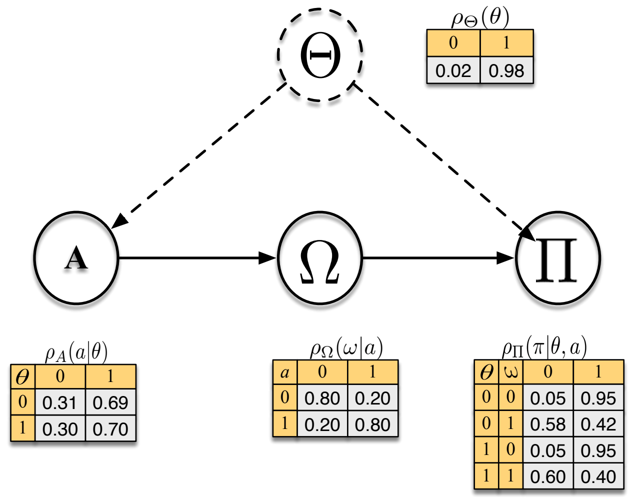

Figure 7: An organization with a closed common cause on A and Π.

as above, we can demonstrate that the JPD produced by this system again has the same MDO as the two earlier cases (Table 1 Part (b)). Here, however, the manager *can* deduce the effects of the employee's actions on performance from the MDO.

The JPD generated by the status quo system in this case differs in two crucial ways from those in the last case (Figures 5 and 6). First, in this case, conditional on the employee's action, the probability of the immediate outcome of that action does not depend on the closed variable. Second, conditional on the intermediate variable, the probability of organizational performance does not depend on the employee's actions. The conditional probability of Π given A therefore accurately predicts the effects of interventions on A.

Put differently, regardless of the employee's action, the organization has the highest expected performance when $\omega = 0$. Since, conditional on A, Ω does not depend on Θ, the employee should choose an action to maximize the probability that $\omega = 0$. In other words, he should do $a = 0$. As in the first case, then, the conditional probabilities the manager can derive from the MDO accurately predicts the effects of an intervention. We formalize this intuition in the following proposition.

Proposition 3. Given a collection of variables with an arbitrary number of $K \geq 2$ states each and parameterized by any strictly positive stochastic laws consistent with Figure 7, the effect of an intervention $\mathbf{do}(a)$ on organizational performance is identifiable for all $a \in A$.

Review

These three cases provide some intuition into the causal inference problem of interest. In all three cases, the manager wants to maximize the probability that the organization meets its objective by directing an employee to engage in a specific action. First, however, the manager must determine what she wants the employee to do. Determining that, however, requires her to understand the causal effects of the employee's action on performance.

In attempting to determine these causal effects, she can only use information from the open variables. It the system did not have any closed variables – any factors hidden to her – then causal inference would not pose a problem. When closed variables do influence the system, however, whether or not she can predict the effects of a policy or directive depends on seemingly-minor differences in the causal system.

Again, these cases provide a best-case scenario for the manager in many ways. All of the example organizations have simple causal structures. Most of the challenges that a manager would face in the real world have been assumed away. The manager can observe what the employee does and the immediate consequences of those actions. In fact, the manager has perfect information on the behavior of all of the open variables. Moreover, the employee does whatever the manager wants him to do.

The only challenge comes from the existence of a single closed factor, a factor hidden from the manager. These cases only vary in how that factor influences the open factors. In the first case, the closed factor only directly influences the actions chosen by the employee. In the second, it affects both those actions and the immediate consequences of those actions. In the third, it influences both those actions and the overall performance of the organization.

One might see these results as half-empty or half-full. In two of the cases, the manager could correctly predict the expected effect of an intervention. In one, she could not.

If the closed factor only influences employee behavior, then the effects of those actions on performance are identifiable from the information available to the manager (the MDO). However, when the closed factor impacts both employee behavior and the immediate outcome of that behavior – when both depend on a common cause – the two become confounded. The second case therefore suggests that hidden factors that act as common causes may produce a managerial inference problems. Yet, the third case, also including a closed common cause, does *not* preclude inference. Common causes, therefore, do not always create inference problems.

These cases also raise another issue. Each case, despite having a different influence structure, had exactly the same MDO. Managers observe the MDO, not the system itself or even the full JPD. So, even though a manager could actually predict the consequences of an intervention if presiding over a system with the same structure as the first or third case, the manager cannot determine whether the system she manages has that structure. In other words, from the MDO, she could not discern whether the first, second, or third case represents her situation.

Although the three cases provide a sense of the intuition behind the inference problem, they have their own shortcomings. In real-world organizations, even departmental managers might attend to hundreds or thousands of open factors, which themselves may depend on any number of hidden factors. The set of possible influence structures, moreover, increases exponentially with the number of factors in the system.

How many of these possible influence structures might create inference problems for the manager? Notice that, in all three simple cases, identification depended only on the qualitative structure of the causal system – the pattern of relationships connecting the open and closed variables – rather than the specific parameter values quantifying the local stochastic laws. This property holds even for complicated cases, a result that we build upon in characterizing identifiable interventions more generally.

A general analysis

We begin by developing some mathematical infrastructure for the analysis. As in the examples above, an organization consists of a set of factors, represented by variables. The behavior of those factors depends on the structure of the set of influence relations, which are quantified as local stochastic laws.

Variables: Although our simple examples included only four variables each, we now allow the model to include any finite number of variables. Let $\mathcal{V}$, indexed by $N \equiv \{1, \ldots, n\}$, with elements V_j ($j \in N$), represent the set of *variables* in the objective model. We refer to an element $v_j \in V_j$ as a *state* of V_j.[12] When a variable takes on a specific state, v_j, we write $V_j = v_j$ and refer to v_j as an *instantiation* of V_j; $v = (v_1, \ldots, v_n) \in V \equiv \prod_{j=1}^{n} V_j$ denotes an *instantiation* of $\mathcal{V}$, where V (without a subscript) denotes the set of all such instantiations.

Each organization consists of one or more *members* (employees), indexed by $I \equiv \{1, \ldots, m\}$, $m < n$. Each member $i \in I$ has an associated set of *beliefs* (B_i) and a set of *actions* (A_i).[13] Beliefs capture ways in which information available to an employee might influence that person's choice of action. Importantly, we assume that a member's actions depend only on his or her beliefs and, conversely, a member's beliefs influence only his or her actions. Members can influence each other but any such influence must flow through the effects that their actions, or the consequences of those actions, have on the beliefs of other members. We denote the classes of action and belief variables as $\mathcal{A} \equiv \{A_1, \ldots, A_m\}$ and $\mathcal{B} \equiv \{B_1, \ldots, B_m\}$, respectively.

The model can include a variety of other variables. These variables can capture factors external to the organization, such as the actions of suppliers, consumer preferences, and the choices of competitors. They can also capture factors internal to the organization, such as intermediate work product, capital equipment, contracts, culture, and so on. $\mathcal{V}$ also always includes a variable, Π (with real-valued states), that represents *organizational performance*.[14]

We can partition $\mathcal{V}$ into two subsets. The set of *open variables*, labeled $\mathcal{V}_O$, and the set of *closed variables*, labeled $\mathcal{V}_C$. We assume $\mathcal{V}_O$ includes all the action variables ($\mathcal{A}$), other variables representing factors open to the manager (e.g., the intermediate outcomes, Ω, in our simple cases), and the organizational performance variable (Π). The set of *closed variables*,

[12] We restrict all variables to a finite number of states.

[13] Although we restrict actions and beliefs to a single variable each per employee, these variables could easily represent the compounds of many independent activities or of beliefs across multiple dimensions. This assumption therefore does not limit the generality of the model.

[14] Treating the outcomes of interest as a single performance variable does not imply single-dimensional measures of performance. The states themselves may represent multidimensional outcomes. We nevertheless assume that the manager can at least partially order the states of the performance variable according to their desirability.

meanwhile, comprises both the beliefs of employees ($\mathcal{B}$) and other variables representing factors hidden to the manager.

Causal structure: As in the simple cases, the behavior of a variable may depend on other variables. Let G be a directed, acyclic graph, the nodes of which are the elements of $\mathcal{V}$ and the directed edges of which indicate direct influence relationships. We label the set of variables that directly influence V_j as $\mathcal{D}_j$; i.e., $V_k \rightarrow V_j$ if and only if $V_k \in \mathcal{D}_j$. Then, $D_j \equiv \prod_{V_k \in \mathcal{D}_j} V_k$ denotes the set of all possible instantiations of the direct influences on V_j. Assuming that influence relationships only operate in one direction and that paths of influence relationships do not cycle prevents causes from becoming their own effects.

A set of local *stochastic laws* govern the behavior of every variable. Specifically, for each variable, $V_j \in \mathcal{V}$, the local stochastic law, denoted ρ_j, is a function from D_j to strictly positive probability distributions on the states of V_j.[15] Then, when $D_j = d_j$, $\rho_j(v_j|d_j)$ represents the probability that $V_j = v_j$.

We assume that closed variables, aside from those associated with the beliefs of employees, have no antecedents – that is, they do not depend on other variables in the model. This assumption simplifies our representation without much loss of generality. One can imagine, for example, chains of closed variables or that closed variables might sit between other sorts of variables. We implicitly concatenate such chains. Suppose $U_{(\cdot)} \in \mathcal{V}_C$ and $V_{(\cdot)} \in \mathcal{V}_O$ represent closed and open variables, respectively. For example, without loss of generality, we could represent $V_j \leftarrow U_g \leftarrow \cdots \leftarrow U_h \rightarrow \cdots \rightarrow U_k \rightarrow V_i$ as $V_j \leftarrow U'_h \rightarrow V_i$, where the elements of U'_h elaborate all the joint states of $U_g, \ldots, U_h, \ldots, U_k$. Similarly, we could simplify $V_j \rightarrow U_h \rightarrow V_i$ to $V_j \rightarrow V_i$. In cases where the instantiation of a variable does not depend on any other variables, we write $\rho_j(v_j)$.

The organizational model: An *organizational model* $M \equiv (G, \mathcal{V}_C, \mathcal{V}_O, \rho_1, \ldots, \rho_n)$ consists of the graph G which identifies the variables and their influence relations, the sets identifying open and closed variables, $\mathcal{V}_C$ and $\mathcal{V}_O$, and the accompanying stochastic laws,

[15]That is, every state in V_j occurs with some positive probability, perhaps vanishingly small. This assumption avoids division-by-zero complications.

ρ_js. An organizational model M implies a JPD, $P_M(v)$, constructed as follows:

$$P_M(v) = \prod_{j=1}^{n} \rho_j(v_j|d_j). \tag{5}$$

When the context clearly implies M, we omit its subscript. This construction implies:

$$P_M(v) = \prod_{j=1}^{n} P_M(v_j|d_j). \tag{6}$$

As in the examples, we frequently focus on the marginal joint probability distribution over the open variables, $P_M(\mathcal{V}_O)$, the MDO.

Interventions: We assume that the manager can direct any organizational member to **do** any action and that the member will follow that directive without error. For each member i, let $A_i^+ \equiv A_i \cup \{\text{idle}\}$ denote the set of interventions available to the manager ("idle" means that the manager gives member i no specific direction). Then, for any $a_i \in A_i^+$, the notation $\mathbf{do}(a_i)$ means that the manager intervenes in the system by giving member i the instruction to do a_i.

Given M and $\mathbf{do}(a_i) \in A_i^+$, for each ρ_j in M, define the *intervened law* $\rho_{j|\mathbf{do}(a_i)}$ as:

$$\rho_{j|\mathbf{do}(a_i)}(v_j|d_j) \equiv \begin{cases} 1 & \text{if } v_j = a_i \\ \rho_j(v_j|d_j) & \text{if } v_j \in A_i \text{ and } \mathbf{do}(a_i) = \text{idle} \\ 0 & \text{if } v_j \in A_i, v_j \neq a_i \text{ and } \mathbf{do}(a_i) \neq \text{idle} \\ \rho_j(v_j|d_{j|\mathbf{do}(a_i)}) & \text{if } j \neq i \end{cases}, \tag{7}$$

where $d_{j|\mathbf{do}(a_i)}$ represents d_j with any component involving A_i replaced by $\mathbf{do}(a_i)$. As one would expect intuitively, this definition simply indicates that the intervention $\mathbf{do}(a_i)$ alters the stochastic process such that a_i occurs with probability one. The JPD induced by an intervention is:

$$P_{M|\mathbf{do}(a_i)}(v) = \prod_{j=1}^{n} \rho_{j|\mathbf{do}(a_i)}(v_j|d_j). \tag{8}$$

When the context clearly implies M, we simply write $P_{\mathbf{do}(a_i)}$ and refer to the intervened MDO as $P_{\mathbf{do}(a_i)}(\mathcal{V}_O)$.

Causal identification

In this setup, what does it mean to say that the MDO would allow the manager to assess the effects of a potential intervention on organizational performance? We define this ability as follows (adapted from Pearl [2009]).

Definition 1 (Identifiability)**.** Given an organizational model M and some $i \in I$, the intervention $\mathbf{do}(a_i) \in A_i^+$ is *identifiable* if $P_{M|\mathbf{do}(a_i)}(\mathcal{V}_O) = P_{M'|\mathbf{do}(a_i)}(\mathcal{V}_O)$ for all alternative models M' that satisfy: (i) $G = G'$; and (ii) $P_M(\mathcal{V}_O) = P_{M'}(\mathcal{V}_O)$.

In other words, a managerial intervention is identifiable if all possible parameterizations of the local stochastic laws for an organizational model that produce the same MDO, $P_M(\mathcal{V}_O)$, also have the same marginal distribution over open variables after an intervention, $P_{M|\mathbf{do}(a_i)}(\mathcal{V}_O)$. Given this definition, we can prove the following result by removing the effects of the closed beliefs through the law of total probability and by then applying Tian and Pearl [2002] to our context.

Proposition 4. Given an organizational model M, an intervention $\mathbf{do}(a_i)$ is identifiable if and only if no closed, bi-directed path connects A_i to any V_j such that $A_i \to V_j$.

A *path* exists between two variables when a sequence of influence relationships connects them (regardless of the directionality of those relationships). A *closed, bi-directed path* is a path that includes one or more closed variables in which the edges at both ends point outward. For example, $A_i \leftarrow B_i \leftarrow U_g \to B_j \to A_j$ and $A_i \leftarrow B_i \leftarrow U_g \to V_j$ (where $V_j \in \mathcal{V}_O$) are both closed, bi-directed paths.

When applying this proposition, all of the potential paths between two variables must be considered. For example, in Figure 4, two paths connect the actions of the employee to the intermediate outcomes: the direct path, $A \rightarrow \Omega$, and an indirect path, $A \leftarrow \Theta \rightarrow \Omega$. This indirect path includes a closed variable (Θ) and is bi-directed. Consistent with the proposition, the effect of an intervention in that example is not identifiable.

Given the wide range of possible organizational models, this result has two surprising features. First, identifiability depends only on the qualitative structure of influence relations, summarized by G. The specific numerical parameters do not matter. Second, identification only fails when a confounding common cause exists with respect to an employee's action and one of its immediate consequences. Any other pattern of influence relationships allows identification. In the set of all possible organizational models, almost none belong to the subset of problematic cases.

Although that fact may sound promising, the relevant question is not what proportion of structures prove problematic but rather how prevalent those problematic structures might be in practice. Unfortunately the answer is: probably very common. One can easily imagine problematic cases. For example, any situation in which an employee receives private information about something that affects the immediate consequences of his or her actions would prove problematic. In real-world settings, these cases seem more the rule than the exception.

Solutions to the inference problem

On the one hand, the real world appears rife with problematic situations. On the other hand, organizations exist in vast numbers. They produce most goods and services. They coordinate individuals pursuing a wide variety of shared goals. Although few would claim that managers have an easy job, they can and do organize and lead people effectively. If the causal inference problem is as common as it is serious, then this fact suggests that organizations must find ways around it.

How might organizations circumvent this problem? Notice that the problematic cases have a specific form: they stem from situations in which closed variables, such as variables about which employees have private information, influence both employee actions as well as the direct consequences of those actions.

We see three potential ways of getting around this problematic structure: experimentation, illumination, and substitution. Experimentation means learning the effects of an intervention directly by doing it and then observing its results. Illumination refers to gaining insight into the closed variables through actions or processes that effectively shift them from being closed to being open. Substitution involves exchanging the influences of closed variables with those of open variables.

A number of organizational features may serve to resolve causal inference problems by implementing some combination of experimentation, illumination or substitution. Although we believe that each of the following conjectures can be proven formally, doing so would require more space that this article allows. All of these features have been discussed at length in the organizations literature. The causal inference perspective nevertheless adds a novel dimension to how and why they might contribute to organizational effectiveness.

Experimentation

The most obvious solution to the inference problem involves experimentation. Experimentation allows managers to infer causal relationships within organizations for the same reason that it allows scientists to infer these relationships in the world. Experimentation, in this case by directing the employee to engage in a range of actions at different times, removes the dependency of the employee's choice of action on any closed variables [Pearl, 2009], thereby allowing for accurate assessment of the relationship between the employee's behavior and the consequences of that behavior.

Organizations increasingly use experiments as a formal tool. A-B testing, for example, has become a popular approach for evaluating software product designs [Koning et al., 2019, Deniz, 2020]. Many companies have also implemented their own randomized clinical trials (RCTs) to test policies ranging from job design to incentive schemes [e.g., Orpen, 1979, Lazear, 2000]. For many types of employee actions, however, formal experiments end up being impractical or unethical. Even when possible, the costs of designing and implementing them can prove prohibitive.

Of course, experimentation may serve as a solution to the inference problem even in cases where managers have not explicitly designed and implemented such tests. Learning-by-doing, for example, can occur even when the variation in what an actor does stems entirely from errors—in our setting, engaging in actions that would have suboptimal outcomes based on the expectations of the employees [Muth, 1986, Zangwill and Kantor, 1998].

Learning-by-doing, however, has its own complications. Noise, for example, can stymie the ability of the learner to understand the behavior of the system [e.g., Lounamaa and March, 1987, March, 1991]. Complex interdependencies in a causal system, moreover, can lead myopic learners to become trapped on local optima [Kauffman, 1993, Levinthal, 1997]. Levinthal and March [1981] felicitously labeled this phenomenon the "competency" trap.

Our analysis also raises another issue with learning by doing. Being able to learn from errors, for example, requires that one understands that they had not been intended or expected in the first place. If surprising, they serve as an exogenous source of variation in the employee's actions—much as an instrumental variable provides another solution to the inference problem [Pearl, 2009]. Employees can therefore learn from their own experiences. Yet, the manager may not have the same insight into whether the actions of the employee had been intended or unintended. Thus, learning-by-doing may not extend to learning-by-observing.

The equivalent to learning-by-doing for the manager would involve only variation in the manager's actions. Sometimes the manager, for example, might intervene to direct the employee to engage in an action that she did not intend them to do. These accidents in managerial direction might then provide a source of exogenous variation that could allow the manager to develop a better understanding of the effects of her interventions.

Illumination

A second solution to the inference problem involves giving the manager insight into the closed variable that influences both the employee's actions and the consequences of those actions. In other words, transforming the problematic closed variable into an open one. If the manager has insight into that variable, it becomes part of the MDO. The manager may then have sufficient information to determine the causal relationships between the employee's actions and the outcomes of those actions.

But how might managers gain insight into these closed variables? We see at least two ways in which that might happen. The first stems from the training of the manager. The second involves the boundaries of the firm.

Internal labor markets: Scholars of work and organizations have long been interested in internal labor markets, the idea that employees would have careers within the firm [Doeringer and Piore, 1971]. Even at the level of the CEO, many of the best performing organizations promote from within.

Three dominant explanations have been offered for the functionality of internal promotions. The first has to do with incentives. If employees believe that those who do well will get promoted to the next level, presumably with commensurate gains in salary and stature, they should work harder [Malcomson, 1984]. The second stems from selection. Organizations should have more and better information about the abilities and attributes of their current employees. Promoting from within may therefore allow them to choose better managers than they could when hiring from outside [Bidwell, 2011].

The final benefit has to do with the accumulation of human and social capital. Employees who rise through the ranks within a firm should invest more in building firm-specific human and social capital [Doeringer and Piore, 1971, Groysberg et al., 2008]. That firm-specific human and social capital can also benefit the organization to the extent that it allows it to operate more efficiently or to differentiate itself from rivals [Groysberg et al., 2008, Tate and Yang, 2015].

Understanding the organization as a causal system adds another potential explanation for this practice: Internal promotions may help to solve the assessment of potentially-confounded interventions, especially in organizations with a function-based division of labor. Promoting from within can ameliorate the inference problem because those promoted have experience in the role of the employees that they manage. That experience may equip them to understand the local information available to those employees and how it influences their behaviors.

Rotational training systems may similarly serve as a means of developing managers for higher rungs on the corporate ladder. One of the limitations, however, with promoting from within as a solution to the managerial inference problem comes from the fact that, at some point, a sequence of promotions will result in the manager having oversight of jobs that she never held. Even so, if managers have had the opportunity to rotate through different functions or divisions early in their careers, they may gain important insights into the problematic cases that they will face later on.

Firm boundaries: A causal systems perspective could also form the basis for a theory of the firm. Suppose that, rather than being a challenge for organizations to overcome, confounding factors help to explain the existence of the organization in the first place. In other words, organizations might exist, in part, because they eliminate inference problems.

Effective management requires an understanding of the effects of managerial interventions. Managers, however, can generally only intervene in factors that reside within the organization. Many of the mechanisms described above for addressing the inference problem require some measure of control. Experimentation, for example, relies on the ability of the manager to direct the employee to engage in particular actions as a means of examining the effects of that direction. Formal bureaucratic rules also rely on the premise that employees have some incentive for following them. In this sense, solving managerial inference problems often involves some combination of bringing an activity within the boundary of the organization with some other means of dealing with the problematic closed variable.

Expanding the scope of the firm may even provide a direct solution to the inference problem. Returning to one of our examples above, imagine that the employee must determine inventory levels based on their information about a supplier. When that supplier resides outside the firm, the manager may need to rely entirely on her employee's actions to infer the reliability of the supplier. The manager therefore cannot evaluate the potential effects of an intervention with regard to the inventory levels. If the organization brought the supplier in-house, however, then the manager *would* have direct access to the relevant information. The problematic bi-directed path would disappear.

This explanation for the scope of the firm bears some similarity to that proposed by resource dependence theory [Pfeffer and Salancik, 1978]. In that perspective, firms expand their boundaries to gain control over uncertain sources of supply. But a causal systems perspective also differs substantially from resource dependence theory because it focuses on uncertainty in the causal relationships between actions and outcomes themselves rather than in the levels of some input available to the firm.

Substitution

A final solution to the inference problem involves replacing the closed factors influencing employee behavior with substitute influences that the manager can observe. Much as the case with illumination, this substitution shifts the factors determining employee behavior from outside to inside the MDO.

We can think of at least two organizational features that might act as such substitutes. Organizational culture – at least defined in a particular way – could replace myriad hidden factors driving employee actions with public norms understood by all. Similarly, the introduction of bureaucratic processes – rules and routines – could also replace private rules of action with public ones.

Culture: Strong culture organizations appear to perform better than their rivals [Kotter and Heskett, 1992, Gordon and DiTomaso, 1992, Boyce et al., 2015]. Two major items have been discussed as the probable sources of this advantage: shared values and more efficient communication. Organizations with strong cultures have been seen as better able to coordinate because they develop shared schema and specialized vocabularies [Srivastava et al., 2018]. Organizations with strong cultures have also been seen as having members with relatively homogeneous goals and values [Kotter and Heskett, 1992, O'Reilly and Chatman, 1996], whether through the selective recruitment and retention of members more closely aligned with the values of the organization or through members becoming more similar to each other due to social influence [March, 1991, Srivastava et al., 2018].

Neither of these attributes solve the managerial inference problem. Shared language can prevent problems of miscommunication and by doing so facilitate coordination. But it does not change the causal structure of the organizational system. Personal values, meanwhile, remain fundamentally unobservable. These values, like employee beliefs, fall into the broader category of attitudes, all of which may influence employee actions while remaining hidden to the manager. From the perspective of our theory, all such factors produce problematic structures, to the extent that these closed variables also influence the consequences of employee decisions.

However, a recent stream in the field of social ontology points to a way in which culture could solve such problems. Gilbert [2014] argues that groups can coordinate on a goal through joint commitment. Joint commitment involves individual group members committing to enact their individual roles in a manner consistent with a group-level entity holding these shared intentions. This perspective therefore suggests that an organization *can* value diversity or "believe that the customer is always right" in the sense that all members of the group act as if it did. Employees need not actually hold these intentions themselves, they only need to enact their roles in a fashion consistent with such intentions at the group level.

Importantly, this joint commitment substitutes for individual intentions (i.e. values, beliefs, goals). Also importantly, members of the organization, including the manager, must understand the organization-level intention. It must be public. The collective belief therefore becomes part of the set of states and variables observed by the manager, the MDO, thereby eliminating the problem of confounded common causes.

Bureaucracy: As organizations mature, they typically become more bureaucratic. Standardized rules and routines govern ever larger shares of their operations [Weber, 1968, Bendix, 1956, Blau and Schoenherr, 1971]. Although these rules and routines have been denigrated as stultifying, they also allow organizations to operate reliably and accountably [Hannan and Freeman, 1984, Adler and Borys, 1996].

Such policies and procedures may also clarify the implications of managerial interventions. In the absence of bureaucratic rules and procedures, employees must depend on their own discretion [Canales, 2014]. When employees act on private information about factors that influence the direct consequences of their actions, they may become unmanageable. This problem arose in the cases above, even when employees hoped to act in the best interests of the organization.

When organizations adopt policies and procedures, they restrict employee discretion. If-then rules provide strict guidance as to how an employee should respond to a particular situation (i.e. to a specific configuration of factor states). These rules, moreover, are public. They appear in manuals, handbooks, and other documents [March et al., 2000]. They become encoded in software and in operating procedures. Regardless of how they become public, these bureaucratic policies and procedures replace problematic hidden production factors with rules understood by the manager

Rules may even have the ability to solve the managerial inference problem if they remain tacit and informal—"routines" in much of the literature on organizations [e.g., March and Simon, 1958, Cyert and March, 1963, Nelson and Winter, 1982]. To the extent that "everyone" understands the nature of these if-then rules, routines would also represent variables open to the manager. They, too, could therefore solve the managerial inference problem through substitution.

Conclusion

Our results characterize the nature of an important causal inference problem for managers, when and why it exists. Causal inference only poses a problem when some closed variable – some factor of production of which the manager remains unaware or unable to infer – influences both employee actions and the direct consequences of those actions. In the space of all possible organizational systems, such structures are a rarity. In the real world, however, such circumstances seem common. For example, they arise whenever an organization member has private information about one or more factors that have a direct effect on the consequences of his actions.

If this problem proved insurmountable, organizations could not function. But they do. In practice, therefore, organizations must have means of solving or circumventing this problem.

We suggest three classes of potential solutions: experimentation, illumination, and substitution. In particular, a number of well-studied features of organizations, such as bureaucracy and organizational culture, may serve to eliminate the confounding effects of closed variables. A causal systems perspective on organizations can therefore shed new light on a number of organizational features.

Although our examples and language have been focused on formal organizations, relying on the language of managers and employees, these issues also have relevance for informal organizations and communities. When these groups include some informal authority structure, those vested with such authority essentially act as managers while those following them play the role of employees. To mobilize a group effectively, leaders must understand the probable consequences of any directions that they might give.

Our results could even extend to informal communities without any explicit or implicit hierarchy. Coordination in such groups, to some extent, requires each member to understand the consequences of their own actions in the context of the actions of others. In the structure of our model, each individual in such a community acts as both manager and employee. Effective interventions (actions) therefore often depend on each member of the community being able to infer the causal relationships between actions and consequences.

Although a number of perspectives have highlighted a variety of factors that can stymie coordination, the inference issue represents an under-appreciated problem. It can occur even in the absence of cognitive constraints. It can arise even without a divergence in interests between principals and agents.

Social scientists have long been aware of the challenge of causal identification in their own empirical research. Our understanding of and appreciation for these issues has progressed rapidly over the past two decades. We believe that recognition of this inference problem should carry over into our theoretical understanding of organizational and social behavior. This paper offers a first step, hopefully one of many, in that direction.

Appendices

Proof of Proposition 1

By the Chain Rule, one can factor *any* JPD, $P(\theta, a, \omega, \pi)$ as:

$$P(\theta, a, \omega, \pi) = P(\pi|\omega, a, \theta)P(\omega|a, \theta)P(a|\theta)P(\theta). \tag{9}$$

Together equations (1) and (9) imply that $P(\omega|a, \theta) = P(\omega|a)$. In other words, conditional on a, ω does not depend in any way on θ. This conclusion is essential for our first finding. Its generalization, the well-known "d-separation" result of Verma and Pearl [1988], will appear throughout our analysis.

Define the indicator function $1_{\mathbf{do}(a)}$ such that $1_{\mathbf{do}(a)} = 1$ if $a = \mathbf{do}(a)$ and zero otherwise. An intervention $\mathbf{do}(a)$ results in a new JPD, denoted $P_{\mathbf{do}(a)}$, and constructed as in (1) but with $1_{\mathbf{do}(a)}$ substituted for $\rho_A(a|\theta)$:

$$P_{\mathbf{do}(a)}(\theta, a, \omega, \pi) = \rho_\Pi(\pi|\omega)\rho_\Omega(\omega|a)1_{\mathbf{do}(a)}\rho_\Theta(\theta), \tag{10}$$

which is nonzero only if $a = \mathbf{do}(a)$. One can therefore factor (10) as:

$$P_{\mathbf{do}(a)}(\theta, a, \omega, \pi) = P_{\mathbf{do}(a)}(\pi|\omega)P_{\mathbf{do}(a)}(\omega|a)P_{\mathbf{do}(a)}(a)P_{\mathbf{do}(a)}(\theta). \tag{11}$$

Suppose the Manager directs some intervention $\mathbf{do}(a)$. From the equivalence between (1) and (11), it follows that

$$P_{\mathbf{do}(a)}(\theta, a, \omega, \pi) = \begin{cases} P(\pi|\omega)P(\omega|a)P(\theta) & \text{if } a = \mathbf{do}(a), \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, by the law of total probability [e.g., Pearl, 2009, Ch. 1, loc. 468, Kindle version],

$$\begin{aligned} P_{\mathbf{do}(a)}(\omega, \pi) &= \sum_{\theta \in \Theta} P(\pi|\omega) P(\omega|a) P(\theta), \\ &= P(\omega|a) P(\pi|\omega) \sum_{\theta \in \Theta} P(\theta), \\ &= P(\omega|a) P(\pi|\omega). \end{aligned} \tag{12}$$

Since the quantities in (12) can be computed from $P(\mathcal{A}, \Omega, \Pi)$, the MDO, the effect of the invention is identifiable. Given that the distributions and the intervention have been chosen artibitrarily, moreover, this conclusion will hold for any potential intervention and for any pair P and P' such that $P(\mathcal{A}, \Omega, \Pi) = P'(\mathcal{A}, \Omega, \Pi)$.

We can say more. By the Chain Rule, one can factor $P(\mathcal{A}, \Omega, \Pi)$ as:

$$P(a, \omega, \pi) = P(a) P(\omega|a) P(\pi|\omega, a). \tag{13}$$

By Bayes' Rule,

$$P(\omega, \pi|a) = \frac{P(a, \omega, \pi)}{P(a)}. \tag{14}$$

Combining (13) and (14),

$$P(\omega, \pi|a) = P(\omega|a) P(\pi|\omega, a).$$

As noted earlier, π is conditionally independent of a given ω. Therefore,

$$P(\omega, \pi|a) = P(\omega|a) P(\pi|\omega). \tag{15}$$

Thus, $P_{\mathbf{do}(a)}(\omega, \pi) = P(\omega, \pi|a)$. As in the numerical example, an intervention $\mathbf{do}(a)$ yields the same outcome as the conditional probability of (ω, π) given a calculated from $P(A, \Omega, \Pi)$.

Proof of Proposition 2

Consider any model with a JPD generated by the system depicted in Figure 4 and parameterized by stochastic laws $\rho_\Pi, \rho_\Omega, \rho_A, \rho_\Theta$. For identification to fail for some intervention, $\mathbf{do}(a)$, there must be some alternative set of stochastic laws, $\rho'_\Pi, \rho'_\Omega, \rho'_A, \rho'_\Theta$, such that the MDOs under P and P' are equivalent, yet $P_{\mathbf{do}(a)}(\pi, \omega) \neq P'_{\mathbf{do}(a)}(\pi, \omega)$.

Begin by setting $\rho'_\Pi = \rho_\Pi$ and $\rho'_\Theta = \rho_\Theta$. Select a collection of parameters $\{\epsilon_\theta\}_{\theta\in\Theta}$ with the following properties:

1. $\forall\theta \in \Theta, \epsilon_\theta \neq 0$,

2. $\forall\theta \in \Theta, \rho_A(a|\theta) + \epsilon_\theta/\rho_\Theta(\theta) > 0$,

3. $\sum_{\theta\in\Theta} \epsilon_\theta = 0$,

4. $\sum_{\theta\in\Theta} \rho_\Omega(\omega|a,\theta)\rho_\Theta(\theta)\big(\frac{\rho_A(a|\theta)\rho_\Theta(\theta)}{\rho_A(a|\theta)+\epsilon_\theta}\big) \neq \sum_{\theta\in\Theta} \rho_\Omega(\omega|a,\theta)\rho_\Theta(\theta)$.

These conditions can be satisfied with as few as two parameters.

Next, for $\forall\theta \in \Theta$, set

$$\rho'_A(a|\theta) = \rho_A(a|\theta) + \frac{\epsilon_\theta}{\rho_\Theta(\theta)}, \tag{16}$$

and

$$\rho'_\Omega(\omega|a,\theta) = \rho_\Omega(\omega|a,\theta)\frac{\rho_A(a|\theta)}{\rho'_A(a|\theta)}. \tag{17}$$

Under this construction,

$$\begin{aligned} P(a) &= \sum_{\theta\in\Theta} \rho_A(a|\theta)\rho_\Theta(\theta) & (18)\\ &= \sum_{\theta\in\Theta} \rho'_A(a|\theta)\rho'_\Theta(\theta) & (19)\\ &= P'(a). & (20) \end{aligned}$$

We also have,

$$
\begin{aligned}
P'(\omega, a) &= \sum_{\theta \in \Theta} \rho'_\Omega(\omega|a, \theta)\rho'_A(a|\theta)\rho'_\Theta(\theta) && (21)\\
&= \sum_{\theta \in \Theta} \rho_\Omega(\omega|a, \theta)\frac{\rho_A(a|\theta)}{\rho'_A(a|\theta)}\rho'_A(a|\theta)\rho'_\Theta(\theta) && (22)\\
&= \sum_{\theta \in \Theta} \rho_\Omega(\omega|a, \theta)\rho_A(a|\theta)\rho_\Theta(\theta) && (23)\\
&= P(\omega, a). && (24)
\end{aligned}
$$

Finally, since we set $\rho'_\Pi = \rho_\Pi$, for all $\pi \in \Pi, \omega \in \Omega$, we have

$$
\begin{aligned}
P(\pi, \omega, a) &= P(\pi|\omega)P(\omega|a)P(a) && (25)\\
&= P'(\pi|\omega)P'(\omega|a)P'(a) && (26)\\
&= P'(\pi, \omega, a). && (27)
\end{aligned}
$$

The MDOs under P and P' are therefore equivalent.

Now consider the intervention $\mathbf{do}(a)$.

$$
\begin{aligned}
P_{\mathbf{do}(a)}(\pi, \omega) &= P_{\mathbf{do}(a)}(\pi|\omega)P_{\mathbf{do}(a)}(\omega), && (28)\\
&= P(\pi|\omega)\sum_{\theta \in \Theta} P(\omega|a, \theta)P(\theta) && (29)\\
&= \rho_\Pi(\pi|\omega)\sum_{\theta \in \Theta} \rho_\Omega(\omega|a, \theta)\rho_\Theta(\theta). && (30)
\end{aligned}
$$

Then, for $P'_{\mathbf{do}(a)}(\pi, \omega)$, with substitution from above,

$$
\begin{aligned}
P'_{\mathbf{do}(a)}(\pi, \omega) &= \rho'_\Pi(\pi|\omega)\sum_{\theta \in \Theta} \rho'_\Omega(\omega|a, \theta)\rho'_\Theta(\theta) && (31)\\
&= \rho_\Pi(\pi|\omega)\sum_{\theta \in \Theta} \rho_\Omega(\omega|a, \theta)\rho_\Theta(\theta)\left(\frac{\rho_A(a|\theta)\rho_\Theta(\theta)}{\rho_A(a|\theta) + \epsilon_\theta}\right). && (32)
\end{aligned}
$$

By condition 4 for the selection of the ϵ_θs, $P_{\mathbf{do}(a)}(\pi, \omega) \neq P'_{\mathbf{do}(a)}(\pi, \omega)$.

Proof of Proposition 3

In this system, P is generated according to:

$$P(\theta, a, \omega, \pi) = \rho_\Pi(\pi|\omega, \theta)\rho_\Omega(\omega|a)\rho_A(a|\theta)\rho_\Theta(\theta). \tag{33}$$

Given an intervention $\hat{a}$,

$$P_{\hat{a}}(\theta, a, \omega, \pi) = \begin{cases} \rho_\Pi(\pi|\omega, \theta)\rho_\Omega(\omega|\hat{a})\rho_\Theta(\theta) & \text{if } a = \hat{a}, \\ 0 & \text{otherwise} \end{cases}. \tag{34}$$

Consider an arbitrary parameterization of the organizational model. From (34),

$$\begin{aligned} P_{\hat{a}}(\omega, \pi) &= \sum_{\theta\in\Theta} P(\pi|\omega, \theta)P(\omega|\hat{a})P(\theta), \qquad (35) \\ &= P(\omega|\hat{a}) \sum_{\theta\in\Theta} P(\pi|\omega, \theta)P(\theta), \\ &= P(\omega|\hat{a}) \sum_{a\in A} \sum_{\theta\in\Theta} P(\pi|\omega, \theta)P(\theta|a)P(a). \end{aligned}$$

From the conditional independence inherent in this organizational structure (see Appendix), the following equivalences are true:

$$P(\theta|a) = P(\theta|a, \omega), \text{ and} \tag{36}$$

$$P(\pi|\omega, \theta) = P(\pi|\theta, \omega, a). \tag{37}$$

Substituting (36) and (37) into (35) yields

$$\begin{aligned} P_{\hat{a}}(\omega, \pi) &= P(\omega|\hat{a}) \sum_{a\in A} P(a) \sum_{\theta\in\Theta} P(\pi|\theta, \omega, a)P(\theta|a, \omega), \\ &= P(\omega|\hat{a}) \sum_{a\in A} P(\pi|\omega, a)P(a). \qquad (38) \end{aligned}$$

Noting that the quantities on the right hand side of (38) are all computable from the MDO completes the result.

Proof of Proposition 4

Lemma: Begin with an organizational model $M \equiv (G, \mathcal{V}_C, \mathcal{V}_O, \rho_1, \ldots, \rho_n)$. Consider an arbitrary actor $i \in I$ and construct a model $M' \equiv (\mathcal{V}', \rho'_1, \ldots, \rho'_{n-1})$, identical to M with the following exceptions: First, remove the belief variables:

$$\mathcal{V}' = \mathcal{V} \backslash \bigcup_{i \in I} B_i.$$

Next, for all $i \in I$, set $\mathcal{D}'_{a_i} \equiv \mathcal{D}_{b_i}$ and define

$$\rho'_{A_i}(a_i|d_{a_i}) = \sum_{b_i \in B_i} \rho_{A_i}(a_i|b_i)\rho_{B_i}(b_i|d_{B_i}). \tag{39}$$

Then, $P_{M|\hat{a}}(\mathcal{V}_O) = P_{M'|\hat{a}}(\mathcal{V}_O)$.

Proof. Recall our notational convention that $b \in \mathcal{B}$ is used to mean $b \in \mathsf{X}_{i \in I} B_i$, etc., so that we write $P(v) = P(\pi, k, a, b, u)$, $P(o) = P(\pi, k, a)$, and so on. Then,

$$\begin{aligned}
P_M(\pi, k, a, u) &= \sum_{b \in \mathcal{B}} P_M(\pi, k, a, b, u) \\
&= \sum_{b \in \mathcal{B}} \left[P_M(\pi|d_\pi) P_M(u) \prod_{i \in I} P_M(a_i|b_i) P_M(b_i|d_{B_i}) \prod_{K_j \in \mathcal{K}} P_M(k_j|d_{K_j}) \right], \\
&= P_M(\pi|d_\pi) P_M(u) \prod_{K_j \in \mathcal{K}} P_M(k_j|d_{K_j}) \sum_{b \in \mathcal{B}} \prod_{i \in I} P_M(a_i|b_i) P_M(b_i|d_{B_i}) \\
&= P_M(\pi|d_\pi) P_M(u) \prod_{K_j \in \mathcal{K}} P_M(k_j|d_{K_j}) \prod_{i \in I} \sum_{b_i \in B_i} P_M(a_i|b_i) P_M(b_i|d_{B_i})
\end{aligned} \tag{40}$$

The last two steps stem from the fact that, by assumption, beliefs only influence an actor's own actions. Turning to M',

$$\begin{aligned}
P_{M'}(\pi, k, a, u) &= P_{M'}(\pi|d_\pi) P_{M'}(u) \prod_{K_j \in \mathcal{K}} P_{M'}(k_j|d_{K_j}) \prod_{i \in I} P_{M'}(a_i|d_{A_i}) \\
&= P_M(\pi|d_\pi) P_M(u) \prod_{K_j \in \mathcal{K}} P_M(k_j|d_{K_j}) \prod_{i \in I} \sum_{b_i \in B_i} P_M(a_i|b_i) P_M(b_i|d_{B_i}) && (41) \\
&= P_M(\pi, k, a, u). && (42)
\end{aligned}$$

Step (41) results from the definition (39) and step (42) from (40). Let a_{-i} denote the action profile $a \in \mathcal{A}$ with a_i removed. From the definition of an intervention (7) and the preceding result,

$$\begin{aligned}
P_{M'|\hat{a}_i}(\mathcal{V}_O) &= \sum_{u \in \mathcal{U}} P_{M'}(\pi|d_\pi) \prod_{K_j \in \mathcal{K}} P_{M'}(k_j|d_{K_j}) \prod_{j \in I \setminus \{i\}} P_{M'}(a_j|d_{A_j}) P_{M'}(u) \\
&= \sum_{u \in \mathcal{U}} P_M(\pi|d_\pi) \prod_{K_j \in \mathcal{K}} P_M(k_j|d_{K_j}) \prod_{j \in I \setminus \{i\}} \sum_{b_j \in B_j} P_M(a_j|b_j) P_M(b_j|d_{B_j}) P_M(u) \\
&= P_{M|\hat{a}_i}(\mathcal{V}_O).
\end{aligned}$$

□

Proof: The model M' meets all the criteria for Tian and Pearl [2002], Theorem 3, which also provides the formula for $P_{M'|\hat{a}_i}$ when the intervention is identifiable.

References

Paul S. Adler and Bryan Borys. Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly*, 41:61–89, 1996.

Armen A. Alchian and Harold Demsetz. Production, information costs, and economic organization. *American Economic Review*, 62:777–795, 1972.

Joshua D. Angrist and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ, 2009.

Reinhard Bendix. *Work and Authority in Industry*. Wiley, New York, 1956.

Matthew J. Bidwell. Paying more to get less: The effects of external hiring versus internal mobility. *Administrative Science Quarterly*, 56:369–407, 2011.

Peter M. Blau and Richard A. Schoenherr. *The Structure of Organizations*. Basic Books (AZ), 1971.

Anthony S. Boyce, Levi R. G. Nieminen, Michael A. Gillespie, Ann Marie Ryan, and Danial R. Denison. Which comes first, organizational culture or performance? A longitudinal study of causal priority with automobile dealerships. *Journal of Organizational Behavior*, 36(3):339–359, 2015.

Rodrigo Canales. Weaving straw into gold: Managing organizational tensions between standardization and flexibility in microfinance. *Organization Science*, 25(1):1–28, 2014.

Ronald H. Coase. The nature of the firm. *Economica*, 16:386–405, 1937.

Richard M. Cyert and James G. March. *A Behavioral Theory of the Firm*. Blackwell Publishers, Oxford, 1963.

Berk Can Deniz. Experimentation and incrementalism: The impact of the adoption of A/B testing. Working paper, Stanford University, 2020.

Peter B. Doeringer and Michael J. Piore. *Internal Labor Markets and Manpower Analysis*. Heath, Lexington, MA, 1971.

Felix Elwert. *Handbook of Causal Analysis for Social Research*, chapter Graphical causal models, pages 245–273. Springer Science+Business Media, Dordrecht, 2013.

Felix Elwert and Christopher Winship. Endogenous selection bias: The problem of conditioning of a collider variable. *Annual Review of Sociology*, 40:31–53, 2014.

Margaret Gilbert. *Joint Commitment: How We Make the Social World*. Oxford University Press, New York, 2014.

George G. Gordon and Nancy DiTomaso. Predicting corporate performance from organizational culture. *Journal of Management Studies*, 29:783–799, 1992.

Sanford J. Grossman and Oliver D. Hart. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4):691–719, 1986.

Boris Groysberg, Linda-Eling Lee, and Ashish Nanda. Can they take it with them? The portability of star knowledge workers' performance. *Management Science*, 54:1213–1230, 2008.

Michael T. Hannan and John Freeman. Structural inertia and organizational change. *American Sociological Review*, 49(2):149–64, 1984.

Michael C. Jensen and William H. Meckling. Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics*, 3(4):305–360, 1976.

Stuart A. Kauffman. *The Origins of Order*. Oxford University Press, New York, 1993.

Rembrand Koning, Sharique Hasan, and Aaron Chatterji. Experimentation and startup performance: Evidence from A/B testing. NBER Working Paper No. 26278, 2019.

John P. Kotter and James L. Heskett. *Corporate Culture and Performance*. Free Press, New York, 1992.

Edward P. Lazear. Performance pay and productivity. *American Economic Review*, 90(5): 1346–1361, 2000.

Daniel A. Levinthal. Adaptation on rugged landscapes. *Management Science*, 43:943–950, 1997.

Daniel A. Levinthal and James G. March. A model of adaptive organizational search. *Journal of Economic Behavior and Organization*, 2(4):307–333, 1981.

Pertti H. Lounamaa and James G. March. Adaptive coordination of a learning team. *Management Science*, 33(1):107–123, 1987.

James M. Malcomson. Work incentives, hierarchy, and internal labor markets. *Journal of Political Economy*, 92(3):486–507, 1984.

James G. March. The business firm as a political coalition. *Journal of Politics*, 24(4): 662–678, 1962.

James G. March. Bounded rationality, ambiguity, and the engineering of choice. *The Bell Journal of Economics*, 9(2):587–608, 1978.

James G. March. Exploration and exploitation in organizational learning. *Organization Science*, 2:71–87, 1991.

James G. March and Herbert A. Simon. *Organizations*. John Wiley, New York, 1958.

James G. March, Martin Schulz, and Xueguang Zhou. *The Dynamics of Rules: Change in Written Organizational Codes*. Stanford University Press, Stanford, CA, 2000.

Paul R. Milgrom and John Donald Roberts. *Economics, Organization, and Management*. Prentice-Hall, Englewood Cliffs, NJ, 1992.

Stephen Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, Cambridge, 2007.

John F. Muth. Search theory and the manufacturing progress function. *Management Science*, 32(8):948–962, 1986.

Richard R. Nelson and Sidney G. Winter. *An Evolutionary Theory of Economic Change*. Belknap press, 1982.

Charles A. O'Reilly and Jennifer A. Chatman. Culture as social control: Corporations, culture and commitment. *Research in Organizational Behavior*, 18:157–2000, 1996.

Christopher Orpen. The effects of job enrichment on employee satisfaction, motivation, involvement, and performance: A field experiment. *Human Relations*, 32(3):189–217, 1979.

Talcott Parsons. *Structure and Process in Modern Societies*. Free Press, Glencoe, IL, 1960.

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., San Francisco, CA, 2 edition, 1988.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, second edition, 2009.

Jeffrey Pfeffer and Gerald R. Salancik. Organizational decision making as a political process: The case of a university budget. *Administrative Science Quarterly*, 19(2):135–151, 1974.

Jeffrey Pfeffer and Gerald R. Salancik. *The External Control of Organizations: A Resource Dependence Perspective*. Harper & Row, Stanford, CA, 1978.

Stephen A. Ross. The economic theory of agency: The principal's problem. *American Economic Review*, 63(2):134–139, 1973.

Herbert A. Simon. *Administrative Behavior*. Macmillan, New York, 1945.

Herbert A. Simon. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106:467–482, 1962.

Sameer B. Srivastava, Amir Goldberg, V. Govind Manian, and Christopher Potts. Enculturation trajectories: Language, cultural adaptation, and individual outcomes in organizations. *Management Science*, 64(3):1348–1364, 2018.

Geoffrey Tate and Liu Yang. The bright side of corporate diversification: Evidence from internal labor markets. *Review of Financial Studies*, 28(8):2203–2249, 2015.

Frederick W. Taylor. *The Principles of Scientific Management*. Harper, New York, 1911.

Jin Tian and Judea Pearl. A general identification condition for causal effects. *AAAI/IAAI*, pages 567–573, 2002.

Tom Verma and Judea Pearl. Causal Networks: Semantics and Expressiveness. In *Proceedings of the Fourth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-88)*, pages 352–59, Corvallis, Oregon, 1988.

Max Weber. *Economy and Society*. 1968.

Oliver E. Williamson. The economics of organization: The transaction cost approach. *American Journal of Sociology*, 87(3):548–577, 1981.

Willard I. Zangwill and Paul B. Kantor. Toward a theory of continuous improvement and the learning curve. *Management Science*, 44(7):910–920, 1998.