

# Independence and Conditional Independence in Causal Systems

Karim Chalak<sup>\*†</sup>      Halbert White  
Boston College      UC San Diego

September 19, 2008

## Abstract

We study the interrelations between (conditional) independence and causal relations in settable systems. We provide definitions in terms of functional dependence for direct, indirect, and total causality as well as for (indirect) causality *via* and *exclusive of* a set of variables. We then provide necessary and sufficient causal and stochastic conditions for (conditional) dependence among random vectors of interest in settable systems. Immediate corollaries ensure the validity of *Reichenbach's principle of common cause* and its informative extension, the *conditional* Reichenbach principle of common cause. We relate our results to notions of *d*-separation and *D*-separation in the artificial intelligence literature.

**Keywords:** causality, conditional independence, *d*-separation, direct effect, Reichenbach principle, settable systems.

## 1 Introduction

The concepts of independence and conditional independence (see e.g. Dawid, 1979, 1980) play a central role in the study of causal inference across a variety of disciplines including statistics (e.g. Rubin 1974; Holland, 1986; Dawid, 2000, 2002; Rosenbaum, 2002), econometrics (e.g. Granger, 1969; Heckman 2005; Chalak and White, 2007a), and artificial

---

<sup>\*</sup>Karim Chalak is Assistant Professor of Economics, Dept. of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467 (email: chalak@bc.edu); and Halbert White is Chancellor's Associates Distinguished Professor of Economics, Dept. of Economics 0508, UCSD, La Jolla, CA 92093 (email: hwhite@ucsd.edu).

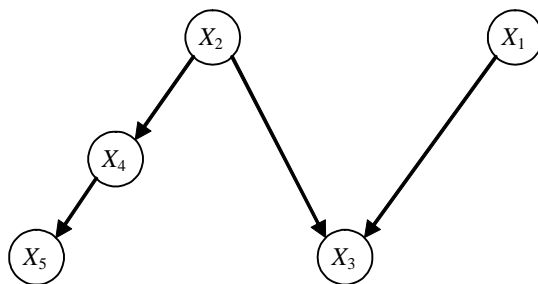
<sup>†</sup>We thank participants of the Harvard/MIT and Yale Econometrics Seminars, Julian Betts, Graham Elliott, Clive Granger, Arthur Lewbel, Mark Machina, Dimitris Politis, Sin-Miao Wang, and especially Ruth Williams for their helpful comments and suggestions. Any errors are solely the author's responsibility.

intelligence (e.g. Spirtes, Glymour, and Scheines, 1993 (SGS); Studeny 1993; Pearl 1988, 1993, 1995, 2000).

In the last two decades, the use of graphical methods to represent conditional independence relationships has been extensively studied in the artificial intelligence literature. This literature introduced graphical criteria applicable to directed acyclic graphs (DAG) to characterize independence and conditional independence among variables in "Bayesian Networks" or, more specifically, "directed Markov fields." In these DAGs, each node represents a random variable. For example, in Graph 1 we have 5 random variables  $X_1, \dots, X_5$ . A DAG is said to represent a probability distribution for these random variables when the joint density function exists and factorizes as the product of the densities of each random variable conditional on its "parents" in the graph. For example, in  $G_1$  we have:

$$p(x_1, x_2, x_3, x_4, x_5) = p_1(x_1)p_2(x_2)p_3(x_3|x_1, x_2)p_4(x_4|x_2)p_5(x_5|x_4),$$

where the left-hand term denotes the joint density and each right-hand term denotes the density of one variable conditional on the value of its "parents." Following Dawid (2002), we refer to DAGs of this kind as "probabilistic DAGs."



Graph 1 ( $G_1$ )

Lauritzen, Dawid, Larsen, and Leimer (1990, theorem 1) show that the joint density admits such a recursive factorization if and only if the collection of conditional independence statements that each variable is conditionally independent of its "non-descendants" given its "parents" in the DAG holds. Lauritzen et. al. (1990) refer to the latter property as the "directed local Markov property," Pearl (2000, theorem 1.2.7) calls it the "parental Markov condition," and SGS (p. 54) refer to this as the "causal Markov property." Using Dawid's (1979) notation  $\perp$  to denote independence,  $G_1$  implies for example that  $X_1 \perp X_2$  and  $X_3 \perp X_4 \mid (X_1, X_2)$  for any distribution represented by  $G_1$ .

A causal meaning is sometimes attributed to such DAGs. In particular, a directed arrow from  $X_2$  to  $X_3$  is interpreted to mean that " $X_2$  is a direct cause of  $X_3$ ." Nevertheless, there is no formal basis for such interpretations in probabilistic DAGs. Indeed, Dawid (2002, p. 164) states "there is absolutely nothing in the probabilistic semantics by which such graphs are supposed to be interpreted that is relevant to such causal intuitions." Although Lauritzen and Spiegelhalter (1988) refer to "causal networks" throughout, they avoid a strict causal interpretation, saying that "causality has a broad interpretation as any natural ordering in which knowledge of a parent influences opinion concerning a child – this influence could be logical, physical, temporal, or simply conceptual in that it may be most appropriate to think of the probability of children given parents" (Lauritzen and Spiegelhalter, 1988, p. 160). Indeed, there is no reference whatsoever to causality in Lauritzen, et. al. (1990); the discussion there is solely concerned with properties of conditional independence and its graphical representation.

A main goal of this paper is to fill the gap created by this lack of causal content. We do this by providing a formal framework in which a DAG can have a well-defined and natural causal meaning, consistent with the independence and conditional independence relations implied by its causal structure. To accomplish this goal, we rely on the framework of *settable systems*, proposed by White and Chalak (2008) (WC) as an extension of the Pearl Causal Model (PCM) (Pearl, 2000, pp. 202-205). Specifically, we study the interrelations between the properties of independence and conditional independence and causal relations defined from functional relationships holding within the settable systems framework. Rather than building on the structure of probabilistic DAGs and their properties (and in particular the Markov assumptions), we first introduce functional definitions of causality within the settable systems framework. We then study the interrelations between these definitions of causality and the concepts of independence and conditional independence. Although graphical representations of our definitions and related concepts emerge that are helpful to heuristic understanding, we emphasize that our analysis does not rely on properties of graphs. Instead, the analysis is driven by functional relationships holding among the various components of a given settable system.

Pearl (2000, definitions 3.2.1 and 4.5.1) provides definitions for "total" and "direct"

causal effects within the PCM. Although Pearl (2000, p. 165) states that "the notion of indirect causality has no intrinsic operational meaning apart from providing a comparison between the direct and total effects," Pearl (2001) and Avin, Shpitser, and Pearl (2005) revise this view and provide definitions for indirect effects as well as "path-specific" effects. Related notions of direct, indirect, and total effects have been proposed in Robins and Greenland (1992), SGS, Robins (2003), Didelez, Dawid, and Geneletti (2006), and Geneletti (2007); see also Rubin (2004).

We contribute to this strand of the literature by providing rigorous definitions of direct and indirect causality based on functional dependence. We refine previous definitions of indirect causality to accommodate notions of causality *via* a set of variables and *exclusive of* a set of variables in recursive systems. Although these refinements are of interest in their own right, their larger significance is that they provide suitable foundations for further developments, described shortly below.

The interrelations between causality and conditional independence relations are also central to certain strands of the philosophy literature (e.g., Spohn, 1980; Hausman and Woodward, 1999; Cartwright, 2000). At the center of much of this inquiry is Reichenbach's (1956) "principle of common cause," which states that if two variables are associated (e.g., correlated) then either one "causes" the other or they both share a third "common cause." For example, the assumption  $X_1 \perp X_2$  in  $G_1$  may be attributed to the lack of a common cause of  $X_1$  and  $X_2$ . Although this principle has intuitive appeal and despite its venerated status, its formal standing is nevertheless ambiguous. Is it an axiom or a postulate, or is it a logical consequence of assumptions as yet unformulated?

Another main goal of this paper is to answer this question. Specifically, we show that Reichenbach's principle follows as a logical consequence of the assumptions defining our settable systems framework. We then extend these results to establish the *conditional* Reichenbach principle of common cause; and we provide necessary and sufficient conditions for probabilistic conditional dependence of certain vectors of random variables in settable systems. Immediate corollaries of these results constitute causal conditions sufficient to ensure independence or conditional independence among random vectors in settable systems. Our Reichenbach-type results follow from properties of direct causality and of our refined

notions of indirect causality via and exclusive of specified variables. Our results permit, but do not require, systems with background variables analogous to the PCM. Further, if present, these background variables are allowed to be jointly independent but need not be. Thus, our results contain as special cases versions of both Markovian and non-Markovian PCMs.

Using properties of conditional independence relationships (e.g. Dawid, 1979), it is possible to infer further conditional independence statements that hold among the variables represented in a probabilistic DAG. In particular, Geiger, et. al. (1990) (see also Verma and Pearl, 1988; Geiger and Pearl, 1993; Pearl, 2000) provide a graphical criterion, called "*d*-separation," that can identify exactly the conditional independence relations implied by a probabilistic DAG under a set of axioms that they refer to as the "graphoid axioms<sup>1</sup>." Lauritzen, et. al. (1990, proposition 3) provide a graphical criterion equivalent to *d*-separation and show that the implications of these criteria when applied to a probabilistic DAG are equivalent to the directed local Markov property (Lauritzen et. al., 1990, theorem 1). For example, one can further conclude from  $G_1$  that  $X_3 \perp X_4 \mid X_2$  and that  $X_3 \perp X_5 \mid X_2$ .

Implications of *d*-separation have been ascribed causal intuition (see for e.g. Pearl, 2000, p. 16-17). In example  $G_1$ , *d*-separation implies  $X_2 \perp X_5 \mid X_4$  which is interpreted to mean that conditioning on a variable  $X_4$  that fully mediates the effect of a cause  $X_2$  on a response  $X_5$  renders  $X_2$  and  $X_5$  conditionally independent. Similarly,  $X_3 \perp X_4 \mid X_2$  is interpreted to mean that conditioning on the common cause  $X_2$  of the two effects  $X_3$  and  $X_4$  renders  $X_3$  and  $X_4$  conditionally independent. Also, the fact that  $X_1 \perp X_2 \mid X_3$  is not implied by *d*-separation is attributed to the notion that conditioning on a common response  $X_3$  of causes  $X_1$  and  $X_2$  renders these conditionally dependent.

We emphasize that there is no formal basis for such causal interpretations in probabilistic DAGs. As we show, however such causal statements are fully meaningful in our settable systems framework. Moreover, although they are not generally valid, we consider special settable systems in which they do hold.

---

<sup>1</sup>In Geiger, et. al. (1990), the four graphoid axioms are properties of conditional independence relations discussed, for example, in Dawid (1979).

This paper is organized as follows. In Section 2, we introduce a version of WC’s settable systems framework. Using this, we provide rigorous definitions of direct causality based on functional dependence, as well as notions of indirect causality via a set of variables and exclusive of a set of variables in recursive systems. Section 3 introduces and proves the conditional Reichenbach principle of common cause. We also provide necessary and sufficient conditions for probabilistic (conditional) dependence of certain vectors of random variables in recursive settable systems. The traditional Reichenbach principle of common cause obtains as a corollary of these results. In Section 4, we relate our results to notions of  $d$ –separation and  $D$ –separation (discussed in Geiger et. al., 1990) and study special settable systems analogous to the Markovian and non-Markovian PCM. In particular, we provide conditions sufficient for the causal intuitions attributed to  $d$ –separation or  $D$ –separation to hold in recursive settable systems. Section 5 concludes and discusses directions for future research. Formal mathematical proofs are collected in the Mathematical Appendix.

Our results thus contribute to answering two fundamental questions of interest for the study of empirical relationships. First, what restrictions (if any) on the possible functionally defined causal relationships holding between variables of interest follow from knowledge of the probability distribution governing these variables? Conversely, what implications for their probability distribution derive from knowledge of functionally defined causal relationships between variables of interest?

## 2 Direct and Indirect Causality in Settable Systems

### 2.1 Settable Systems

WC introduced settable systems as an extension of the PCM that accommodates optimization, equilibrium, and learning. Heuristically, a *stochastic settable system* is a mathematical framework that describes an environment in which a countable number of *units* interact under uncertainty. A unit is construed broadly. It could be a neuron, person, firm, market, or a player-decision pair in game theory, for example. For each unit  $i$  there is a *settable* variable  $\mathcal{X}_i$ . A settable variable  $\mathcal{X}_i$  has a dual aspect. It can be *set* to a random variable

denoted by  $Z_i$  (the *setting*), or it can be *free* to respond to settings of all other settable variables in the system. In the latter case, it is denoted by  $Y_i$ , the *response*. The response  $Y_i$  of a settable variable  $\mathcal{X}_i$  is determined by a *response function*  $r_i$ . For example,  $r_i$  can be determined by a governing principle such as optimization, determining the response for unit  $i$  that is best in some sense, given the settings of all the other settable variables.

Formally, we work with a somewhat specialized version of WC's definition. For this, we write the positive integers as  $\mathbb{N}^+$  and let  $\bar{\mathbb{N}}^+ = \mathbb{N}^+ \cup \{\infty\}$ . When  $n = \infty$ , we interpret  $i = 1, \dots, n$  as  $i = 1, 2, \dots$

**Definition 2.1 *Elementary Partitioned Settable System*** Let  $(\Omega, \mathcal{F})$  be a measurable space such that  $\Omega$  contains at least two elements. Let the **primary setting**  $Z_0 : \Omega \rightarrow \Omega$  be the identity mapping. For  $i = 1, 2, \dots, n$ ,  $n \in \bar{\mathbb{N}}^+$ , let  $\mathbb{S}_i$  be a multi-element Borel-measurable subset of  $\mathbb{R}$  and let **settings**  $Z_i : \Omega \rightarrow \mathbb{S}_i$  be surjective measurable functions. Let  $Z_{(i)}$  be the vector including every setting except  $Z_i$  and taking values in  $\mathbb{S}_{(i)} \subseteq \Omega \times_{j \neq i} \mathbb{S}_j$ ,  $\mathbb{S}_{(i)} \neq \emptyset$ . Let **response functions**  $r_i : \mathbb{S}_{(i)} \rightarrow \mathbb{S}_i$  be measurable functions and define **responses**  $Y_i(\omega) := r_i(Z_{(i)}(\omega))$ . Define **settable variables**  $\mathcal{X}_i : \{0, 1\} \times \Omega \rightarrow \mathbb{S}_i$  as

$$\mathcal{X}_i(0, \omega) := Y_i(\omega) \quad \text{and} \quad \mathcal{X}_i(1, \omega) := Z_i(\omega), \quad \omega \in \Omega.$$

Define  $Y_0 : \Omega \rightarrow \Omega$  and  $\mathcal{X}_0 : \{0, 1\} \times \Omega \rightarrow \Omega$  by  $Y_0(\omega) := \mathcal{X}_0(0, \omega) := \mathcal{X}_0(1, \omega) := Z_0(\omega)$ ,  $\omega \in \Omega$ .

Put  $r := \{r_i\}$  and  $\mathcal{X} := \{\mathcal{X}_0, \mathcal{X}_1, \dots\}$ . The pair  $\mathcal{S} := \{(\Omega, \mathcal{F}), (r, \mathcal{X})\}$  is an **elementary partitioned settable system**.

We briefly discuss a number of specific features of settable systems. The first component of a stochastic settable system is the measurable space  $(\Omega, \mathcal{F})$ . The space  $\Omega$  is assumed to contain at least two elements, ensuring that a *primary intervention*  $\omega \rightarrow \omega^*$ , defined as a pair of distinct  $\Omega$  values  $(\omega, \omega^*)$ , exists. Generally there will be vast numbers of primary interventions.

The settings  $Z_{(i)}$  take values in  $\mathbb{S}_{(i)} \subseteq \Omega \times_{j \neq i} \mathbb{S}_j$ ; we have  $\mathbb{S}_{(i)} \subset \Omega \times_{j \neq i} \mathbb{S}_j$  if there are joint restrictions on the admissible settings values. For example, certain elements of  $\mathbb{S}_{(i)}$  might represent probabilities that must add to one. Whenever  $\mathbb{S}_{(i)}$  contains at least two

elements, it possesses an *admissible intervention* to  $\mathcal{X}_{(i)}$ ,  $z_{(i)} \rightarrow z_{(i)}^* := (z_{(i)}, z_{(i)}^*) \in \mathbb{S}_{(i)} \times \mathbb{S}_{(i)}$ . The requirements that  $\Omega$  contains at least two points and that the  $Z_i$ 's are surjective are necessary but not sufficient for the existence of an admissible intervention. These requirements do suffice, however, when  $\mathbb{S}_{(i)} = \Omega \times_{j \neq i} \mathbb{S}_j$ . When  $\mathbb{S}_{(i)}$  possesses admissible interventions, the assumed surjectivity ensures that for any admissible intervention  $z_{(i)} \rightarrow z_{(i)}^*$  there exists a primary intervention  $\omega \rightarrow \omega^*$  such that  $z_{(i)} \rightarrow z_{(i)}^* = Z_{(i)}(\omega) \rightarrow Z_{(i)}(\omega^*)$ .

The response  $Y_i(\cdot) := r_i(Z_{(i)}(\cdot))$  is random due to its dependence on settings  $Z_j, j \neq i, j \neq 0$ , but it may also be inherently stochastic as a consequence of its direct dependence on  $\omega$  through  $Z_0$ . Thus, a response may embody an aspect of "pure" randomness.

The setting  $Z_0$  and response  $Y_0$  of the *primary settable variable*  $\mathcal{X}_0$  are such that  $Z_0(\omega) = Y_0(\omega) = \omega$ . Therefore, the setting  $Z_0$  of the primary settable variable may directly influence all other responses in the system, whereas its response  $Y_0$  is unaffected by other settings. In this sense,  $\mathcal{X}_0$  introduces randomness to a settable system.

WC's definition explicitly accommodates *attributes*, i.e. fixed items (e.g., numbers, sets, functions) associated with  $i$ . For conciseness and without essential loss of generality, we leave attributes implicit here.

A stochastic settable system is thus composed of a "stochastic" component, i.e., the measurable space  $(\Omega, \mathcal{F})$ , and a structural or causal component  $(r, \mathcal{X})$ , resting on the stochastic component and consisting of response functions and settable variables. WC discuss in detail the relationship of settable systems to the PCM (Pearl, 2000).

In Definition 2.1, a single response  $Y_i$  is free to respond to settings of all other variables in the system. We also wish to consider systems in which responses of several settable variables jointly respond to settings of other variables in the system. This can occur, for example, when responses are determined as a solution to a joint optimization problem. Such specifications are formally implemented in settable systems by *partitioning* the system under study to group certain variables into specific blocks. The system in Definition 2.1 is called "elementary," as every unit  $i$  forms a block by itself. We now introduce a general definition of partitioned settable system.

**Definition 2.2 *Partitioned Settable System*** Let  $\mathcal{S}^e$  be an elementary settable system. Let  $\Pi = \{\Pi_b\}$  be a partition of  $\{1, \dots, n\}$ ,  $n \in \bar{\mathbb{N}}^+$ , with cardinality  $B \in \bar{\mathbb{N}}^+$  ( $B := \#\Pi$ ). For

$i = 1, 2, \dots, n$ , let  $Z_i^\Pi$  be settings and let  $Z_{(b)}^\Pi$  be the vector containing  $Z_0$  and  $Z_i^\Pi, i \notin \Pi_b$ , and taking values in  $\mathbb{S}_{(b)}^\Pi \subseteq \Omega \times_{i \notin \Pi_b} \mathbb{S}_i, \mathbb{S}_{(b)}^\Pi \neq \emptyset, b = 1, \dots, B$ . Suppose there exist measurable functions  $r_i^\Pi : \mathbb{S}_{(b)}^\Pi \rightarrow \mathbb{S}_i$  specific to  $\Pi$  such that responses  $Y_i^\Pi(\omega)$  are determined as

$$Y_i^\Pi(\omega) := r_i^\Pi(Z_{(b)}^\Pi(\omega)), \text{ for } i \in \Pi_b, \quad b = 1, \dots, B.$$

Define the settable variables  $\mathcal{X}_i^\Pi : \{0, 1\} \times \Omega \rightarrow \mathbb{S}_i$  as

$$\mathcal{X}_i^\Pi(0, \omega) := Y_i^\Pi(\omega) \quad \text{and} \quad \mathcal{X}_i^\Pi(1, \omega) := Z_i^\Pi(\omega) \quad \omega \in \Omega.$$

Put  $r^\Pi := \{r_i^\Pi\}$  and  $\mathcal{X}^\Pi := \{\mathcal{X}_0, \mathcal{X}_1^\Pi, \mathcal{X}_2^\Pi \dots\}$ . The pair  $\mathcal{S} := \{(\Omega, \mathcal{F}), (\Pi, r^\Pi, \mathcal{X}^\Pi)\}$  is a **partitioned settable system**.

Observe that response functions and responses are partition-specific. Thus, in Definition 2.2, the response  $Y_i^\Pi$  for unit  $i$  in block  $\Pi_b$  is determined by the function  $r_i^\Pi$  of settings  $Z_{(b)}^\Pi$  outside of block  $\Pi_b$ . For the remainder of the paper, we refer to  $i = 0$  as the *primary unit* and we denote by  $\Pi_0 = \{0\}$  the block corresponding to the primary settable variable.

## 2.2 Direct Causality and Direct Causality Graphs

Settable systems provide a suitable framework for the study of causality. We next give a definition of direct causality within this framework, based on functional dependence. Of particular note is that we define causality in terms of settable variables rather than random variables or events, as is typical elsewhere. For notational convenience, we suppress explicit reference in what follows to the superscript  $\Pi$  in  $Z_i^\Pi, r_i^\Pi, Y_i^\Pi$ , and  $\mathcal{X}_i^\Pi$  unless absolutely necessary; it should nevertheless be borne in mind that these functions are partition-specific.

Heuristically, we say that a settable variable  $\mathcal{X}_i, i \notin \Pi_b$ , *directly causes*  $\mathcal{X}_j, j \in \Pi_b$ , in  $\mathcal{S}$  when the response for  $\mathcal{X}_j$  differs for different settings in  $\mathcal{X}_i$ , while holding all other variables corresponding to units outside of  $\Pi_b$  to the same setting values. There are two main ingredients to this notion of direct causality. Let  $z_{(b)(i)}$  denote the vector containing all elements of setting values  $z_{(b)}$  except  $z_i$ . The first ingredient is an admissible intervention  $(z_{(b)(i)}, z_i) \rightarrow (z_{(b)(i)}, z_i^*)$ . The intervention references only setting values corresponding to units outside of  $\Pi_b$ . Note also that it differs only in the final component. The second ingredient is the behavior of the response to this intervention.

We formalize this notion of direct causality as follows.

**Definition 2.3 *Direct Causality*** Let  $\mathcal{S}$  be a partitioned settable system. For given positive integer  $b$ , let  $j \in \Pi_b$ . (i) For given  $i \notin \Pi_b$ ,  $\mathcal{X}_i$  **directly causes**  $\mathcal{X}_j$  in  $\mathcal{S}$  if there exists an admissible intervention  $(z_{(b)(i)}, z_i) \rightarrow (z_{(b)(i)}, z_i^*)$  such that

$$r_j(z_{(b)(i)}, z_i^*) - r_j(z_{(b)(i)}, z_i) \neq 0,$$

and we write  $\mathcal{X}_i \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_j$ . Otherwise, we say  $\mathcal{X}_i$  **does not directly cause**  $\mathcal{X}_j$  in  $\mathcal{S}$  and write  $\mathcal{X}_i \not\xrightarrow{D}_{\mathcal{S}} \mathcal{X}_j$ . (ii) For  $i, j \in \Pi_b$ ,  $\mathcal{X}_i \not\xrightarrow{D}_{\mathcal{S}} \mathcal{X}_j$ .

We emphasize that even though we follow the literature in referring to "interventions," with their mechanistic or manipulative connotations, the mathematical concept only involves the properties of a response function on its domain.

According to this definition, direct causality may fail either because the set  $\mathbb{S}_{(b)}^{\Pi}$  is so constrained that it does not possess an admissible intervention of the desired structure, or because it does, but the response is the same for both elements of every admissible intervention of the specified form. The latter is perhaps the more common or intuitively appealing possibility, but we need not distinguish further between these possibilities.

Note that, by definition, variables within the same block do not directly cause each other. In particular  $\mathcal{X}_i \not\xrightarrow{D}_{\mathcal{S}} \mathcal{X}_i$ . Also, Definition 2.3 permits *mutual causality*, so that  $\mathcal{X}_i \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_j$  and  $\mathcal{X}_j \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_i$  without contradiction for  $i$  and  $j$  in different blocks. Mutual causality is ruled out in SGS (p. 42), for example, where it is an axiom that if  $A$  causes  $B$  then  $B$  does not cause  $A$ .

We call the response value difference in Definition 2.3 the *direct effect of  $\mathcal{X}_i$  on  $\mathcal{X}_j$*  of the specified intervention. This corresponds to the notion of "controlled" direct effect in Pearl (2001). Nevertheless, the PCM requires a unique fixed point, a requirement absent here; the PCM also does not have a notion of partitioning, so the PCM notion pertains only to elementary partitions; and the PCM does not account for possible joint restrictions on setting values, and thus assumes that  $\mathbb{S}_{(b)} = \Omega \times_{i \neq j} \mathbb{S}_i$ .

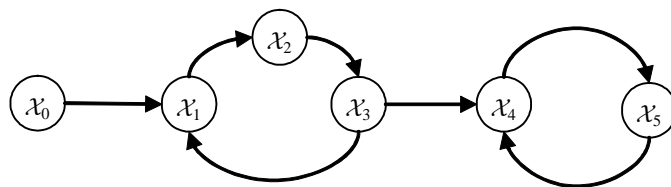
We now introduce notions of paths, successors, predecessors, and intercessors, adapting graph theoretic concepts discussed, for example, by Bang-Jensen and Gutin (2001) (BG).

**Definition 2.4 *Paths, Successors, Predecessors, and Intercessors*** Let  $\mathcal{S}$  be a partitioned settable system. For given positive integer  $b$  let  $j \in \Pi_b$  and  $i \notin \Pi_b$ . We call the

collection of settable variables  $\{\mathcal{X}_i, \mathcal{X}_{i_1}, \dots, \mathcal{X}_{i_m}, \mathcal{X}_j\}$  an  $(\mathcal{X}_i, \mathcal{X}_j)$ -walk of length  $m + 1$  if  $\mathcal{X}_i \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_{i_1} \xrightarrow{D}_{\mathcal{S}} \dots \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_{i_m} \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_j$ . When the elements of an  $(\mathcal{X}_i, \mathcal{X}_j)$ -walk are distinct, we call it an  $(\mathcal{X}_i, \mathcal{X}_j)$ -path. We say  $\mathcal{X}_i$  **precedes**  $\mathcal{X}_j$  or  $\mathcal{X}_j$  **succeeds**  $\mathcal{X}_i$  if there exists at least one  $(\mathcal{X}_i, \mathcal{X}_j)$ -path of positive length. If  $\mathcal{X}_i$  precedes  $\mathcal{X}_j$ , we call  $\mathcal{X}_i$  a **predecessor** of  $\mathcal{X}_j$ , and we call  $\mathcal{X}_j$  a **successor** of  $\mathcal{X}_i$ . If  $\mathcal{X}_i$  precedes  $\mathcal{X}_j$  and  $\mathcal{X}_i$  succeeds  $\mathcal{X}_j$ , we say  $\mathcal{X}_i$  and  $\mathcal{X}_j$  belong to a **cycle**. If  $\mathcal{X}_i$  and  $\mathcal{X}_j$  do not belong to a cycle,  $\mathcal{X}_k$  succeeds  $\mathcal{X}_i$ , and  $\mathcal{X}_k$  precedes  $\mathcal{X}_j$ , we say  $\mathcal{X}_k$  **intercedes**  $\mathcal{X}_i$  and  $\mathcal{X}_j$ . If  $\mathcal{X}_k$  intercedes  $\mathcal{X}_i$  and  $\mathcal{X}_j$ , we call  $\mathcal{X}_k$  an  $(\mathcal{X}_i, \mathcal{X}_j)$  **intercessor**. We denote by  $\mathcal{I}_{i:j}$  the set of  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors.

The *direct causality graphs* associated with settable systems are directed graphs. Specifically, the direct causality graph for a given partitioned settable system  $\mathcal{S}$  is a directed graph  $G := (V, E)$  with a non-empty countable set of vertices  $V = \{\mathcal{X}_i : i = 1, \dots, n\}$  and a set of arcs  $E \subset V \times V$  of ordered pairs of distinct vertices such that an arc  $(\mathcal{X}_i, \mathcal{X}_j)$  belongs to  $E$  if and only if  $\mathcal{X}_i \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_j$ . From Definition 2.3, there exists at most one  $(\mathcal{X}_i, \mathcal{X}_j)$  arc, so  $G$  need not contain nor can it contain “parallel arcs.” Since  $\mathcal{X}_i \not\xrightarrow{D}_{\mathcal{S}} \mathcal{X}_i$ , there can be no arc  $(\mathcal{X}_i, \mathcal{X}_i)$  in  $E$ , so  $G$  need not and can not contain self-directed arcs or “loops.”<sup>2</sup>

Graph  $G_2$  illustrates the concepts of Definition 2.4. We have that  $\{\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_4\}$  and  $\{\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_4\}$  are an  $(\mathcal{X}_0, \mathcal{X}_4)$ -walk of length 7 and an  $(\mathcal{X}_0, \mathcal{X}_4)$ -path of length 4, respectively. We also have that  $\mathcal{X}_0$  precedes  $\mathcal{X}_4$ ,  $\mathcal{X}_3$  succeeds  $\mathcal{X}_1$ , and that  $\mathcal{X}_1$  and  $\mathcal{X}_3$  belong to a cycle, as do  $\mathcal{X}_4$  and  $\mathcal{X}_5$ . The set of  $(\mathcal{X}_1, \mathcal{X}_4)$  intercessors is given by  $\mathcal{I}_{1:4} = \{\mathcal{X}_2, \mathcal{X}_3\}$ . We use the term “intercessor” instead of the possible descriptor “mediator,” as the latter may connote transmission; we want to avoid this, because intercessors need not transmit effects, as we explain further below.



Graph 2 ( $G_2$ )

<sup>2</sup>Loops and parallel arcs can nevertheless be useful in other contexts; see, for example, Golubitsky and Stewart, 2006. With loops or parallel arcs permitted, one may have a “directed pseudograph” or a “directed multigraph” (see BG, p. 4). These are not relevant here.

We emphasize that these direct causality graphs are different from other graphs in the literature. Nodes in direct causality graphs represent settable variables and not random variables or events; arcs represent direct causality relations, rather than functional dependence or probabilistic dependence.

### 2.3 Direct and Indirect Causality in Recursive Settable Systems

In what follows, we focus on *recursive* partitioned settable systems, defined next. For  $0 \leq a \leq b$ , we define  $\Pi_{[a:b]} := \Pi_a \cup \dots \cup \Pi_{b-1} \cup \Pi_b$ . (For  $a < b$ ,  $\Pi_{[b:a]} := \emptyset$ .)

**Definition 2.5 *Recursive Partitioned Settable System*** *Let  $\mathcal{S}$  be a partitioned settable system. For  $b = 0, 1, \dots, B$ , let  $Z_{[0:b]}$  denote the vector containing the settings  $Z_i$  for  $i \in \Pi_{[0:b]}$  and taking values in  $\mathbb{S}_{[0:b]} \subseteq \Omega \times_{i \in \Pi_{[1:b]}} \mathbb{S}_i$ ,  $\mathbb{S}_{[0:b]} \neq \emptyset$ . Suppose that  $r := \{r_i\}$  is such that the responses  $Y_i = \mathcal{X}_i(1, \cdot)$  are determined as*

$$Y_i := r_i(Z_{[0:b-1]}), \text{ for } i \in \Pi_b, \quad b = 1, \dots, B.$$

*Then we say that  $r$  is **recursive**,  $\Pi$  is a **recursive partition**, and the pair  $\mathcal{S} := \{(\Omega, \mathcal{F}), (\Pi, r, \mathcal{X})\}$  is a **recursive partitioned settable system** or simply that  $\mathcal{S}$  is **recursive**.*

We employ the convenient structure of recursive systems to provide definitions of total and indirect causality. This also facilitates the comparison between our results and the DAG-related literature. We leave the study of the interrelations between (conditional) independence and total and indirect causal relationships in non-recursive systems (see, e.g., Lauritzen and Richardson, 2002) for other work.

#### 2.3.1 Direct Causality in Recursive Settable Systems

We now consider how Definition 2.3 (direct causality) specializes to recursive systems. For this, let  $i \in \Pi_{b_1}$  and  $j \in \Pi_{b_2}$  with  $0 \leq b_1 < b_2$ . We write values of settings corresponding to  $\Pi_{[a:b]}$  as  $z_{[a:b]}$ . We also let  $z_{[0:b](i)}$  denote a vector of values for settings for all settable variables corresponding to  $\Pi_{[0:b]}$  except  $\mathcal{X}_i$ . Since  $\mathcal{S}$  is recursive, we can express response values for  $\mathcal{X}_j$  as  $r_j(z_{[0:b_2-1]})$ . We abuse notation somewhat to permute the arguments of  $r_j$

in a way that emphasizes their recursive relation to the argument corresponding to  $\mathcal{X}_i$ . In particular, we write

$$r_j(z_{[0:b_1]}(i), z_i, z_{[b_1+1:b_2-1]}) = r_j(z_{[0:b_2-1]}).$$

Definition 2.3 then concludes that  $\mathcal{X}_i \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_j$  if there exists an admissible intervention  $(z_{[0:b_1]}(i), z_i, z_{[b_1+1:b_2-1]}) \rightarrow (z_{[0:b_1]}(i), z_i^*, z_{[b_1+1:b_2-1]})$  such that

$$r_j(z_{[0:b_1]}(i), z_i^*, z_{[b_1+1:b_2-1]}) - r_j(z_{[0:b_1]}(i), z_i, z_{[b_1+1:b_2-1]}) \neq 0.$$

It follows that if  $\mathcal{S}$  is recursive,  $i \in \Pi_{b_1}$ , and  $j \in \Pi_{b_2}$  with  $b_2 < b_1$  then  $\mathcal{X}_i \not\xrightarrow{D}_{\mathcal{S}} \mathcal{X}_j$ : successors do not directly cause predecessors. In particular, if  $\mathcal{X}_i \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_j$  then  $\mathcal{X}_j \not\xrightarrow{D}_{\mathcal{S}} \mathcal{X}_i$ . Thus, recursive systems do not admit mutual causality. For the graph, this means that we cannot have both arcs  $(\mathcal{X}_i, \mathcal{X}_j)$  and  $(\mathcal{X}_j, \mathcal{X}_i)$  belonging to  $E$ . In addition, a recursive system  $\mathcal{S}$  is acyclic: it does not admit cycles of the form  $\mathcal{X}_i \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_{i_1} \xrightarrow{D}_{\mathcal{S}} \dots \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_{i_m} \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_i$ . Thus, when  $\mathcal{S}$  is recursive, its corresponding direct causality graph  $G$  is a DAG (see proposition 1.4.3 in BG). BG's (p. 175) DFSA algorithm outputs an acyclic ordering of the vertices of any DAG.

In the expression above for recursive settable system direct causality, the values for successors to  $\mathcal{X}_i$  (corresponding to blocks  $\Pi_{[b_1+1:b_2-1]}$ ) are set to the same arbitrary value  $z_{[b_1+1:b_2-1]}$  in both argument lists. Sometimes it is of interest to evaluate the direct effect of  $\mathcal{X}_i$  on  $\mathcal{X}_j$  when successor values are set in both argument lists to the response value that obtains when  $\mathcal{X}_i$  is set to  $z_i$ . We refer to a setting that is determined as a response to its predecessors' settings as a *canonical setting*. In recursive systems, a canonical setting for settable variable  $\mathcal{X}_i$ ,  $i \in \Pi_b$ , is given by

$$Z_i = Y_i := r_i(Z_{[0:b-1]}).$$

Canonical settings are particularly relevant in observational studies where control is not feasible.

For example, we can let settings values  $z_{[b_1+1:b_2-1]}$  be determined as responses  $y_{[b_1+1:b_2-1]}$  to the admissible values of their predecessors' settings, written as

$$y_{[b_1+1:b_2-1]} = r_{[b_1+1:b_2-1]}(z_{[0:b_1]}).$$

The elements of this response vector obtain by recursive substitution. Any given element of this vector depends only on its corresponding predecessors. The direct effect associated with this configuration is then evaluated as

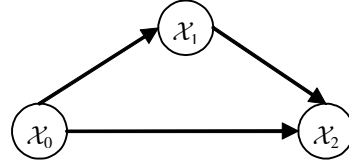
$$r_j(z_{[0:b_1](i)}, z_i^*, y_{[b_1+1:b_2-1]}) - r_j(z_{[0:b_1](i)}, z_i, y_{[b_1+1:b_2-1]}).$$

Pearl (2001) refers to this as the "natural" direct effect. Although Pearl (2001) does not assume recursiveness, he employs the PCM, with its unique fixed point requirement. As mentioned above, we do not require a fixed point, unique or otherwise, so just as for our prior notion of direct causality, this concept of direct causality does not depend on this.

Now consider the following system of three settable variables (see Graph  $G_3$ ) to illustrate the relationships between our Definition 2.3 of direct causality and several notions of direct effects discussed in the literature.

$$\mathcal{X}_1(0, \cdot) = r_1(\mathcal{X}_0(1, \cdot)), \text{ and}$$

$$\mathcal{X}_2(0, \cdot) = r_2(\mathcal{X}_0(1, \cdot), \mathcal{X}_1(1, \cdot)).$$



Graph 3 ( $G_3$ )

Definition 2.3 concludes that  $\mathcal{X}_0 \xrightarrow{D}_S \mathcal{X}_2$  if there exists an admissible intervention  $(z_0, z_1) \rightarrow (z_0^*, z_1)$  such that

$$r_2(z_0^*, z_1) - r_2(z_0, z_1) \neq 0.$$

When this difference is non-zero, it justifies the link from  $\mathcal{X}_0$  to  $\mathcal{X}_2$ . This difference corresponds to the notion of "controlled direct effect" in Pearl (2001).

If  $z_1$  is restricted to a specific suitable value, then we obtain a notion in the spirit of the "standardized direct effect" of Didelez, Dawid, and Geneletti (2006) and Geneletti (2007). In particular, the canonical choice  $z_1 = r_1(z_0)$  yields Pearl's (2001) previously mentioned natural direct effect

$$r_2(z_0^*, r_1(z_0)) - r_2(z_0, r_1(z_0)).$$

This also is what Robins and Greenland (1992) and Robins (2003) call the "pure" direct effect. These same authors refer to

$$r_2(z_0^*, r_1(z_0^*)) - r_2(z_0, r_1(z_0^*))$$

as the "total direct effect."

In other cases, the literature considers notions of direct effects defined as a contrast in some aspect of the distributions of responses for different settings. For example, let  $P$  be a probability measure on  $(\Omega, \mathcal{F})$ ; then the "average" direct effect of  $\mathcal{X}_1$  on  $\mathcal{X}_2$  in the above example is given by

$$E[r_2(Z_0, z_1^*) - r_2(Z_0, z_1)],$$

where  $E$  is the expectation operator associated with  $P$ .

Here, we consider direct effects to be differences in response values for any admissible intervention of the specified form. As Holland (1986) notes, these effects need not be identifiable absent other assumptions. Nevertheless, the direct causality concept of Definition 2.3 is in a precise sense the simplest and most general of the alternatives discussed. It is simplest, in that direct causality is well defined even in the absence of recursive structure or fixed points. It is most general, as it is necessary but not sufficient for the others.

### 2.3.2 Indirect Causality in Recursive Settable Systems

We next define notions of indirect causality for recursive systems. We distinguish notions of indirect causality *via* and *exclusive of* specified variables. These definitions extend notions of indirect causality in Robins and Greenland (1992), SGS, Pearl (2001), Robins (2003), Didelez, Dawid, and Geneletti (2006), and Geneletti (2007), and notions of "path-specific" effects in Pearl (2001) and Avin, Shpitser, and Pearl (2005). Although these extensions are of interest in their own right, their greater significance is that they provide appropriate tools for establishing the conditional Reichenbach principle of common cause, as well as later results on  $d$ -separation and  $D$ -separation.

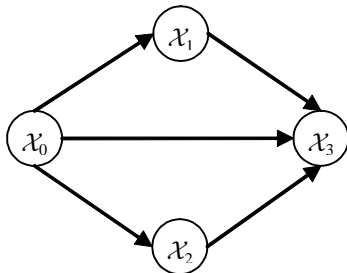
**Indirect Causality Via Given Variables** The basic idea of indirect causality adopted here is straightforward. Consider, for example, the system illustrated in  $G_3$ . There,  $\mathcal{X}_0$  indirectly causes  $\mathcal{X}_2$  via  $\mathcal{X}_1$  if there exists an admissible intervention  $(z_0, r_1(z_0)) \rightarrow (z_0, r_1(z_0^*))$  such that

$$r_2(z_0, r_1(z_0^*)) - r_2(z_0, r_1(z_0)) \neq 0.$$

In the first case,  $Z_0$  is set to the value  $z_0$  and  $Z_1$  to the canonical value  $r_1(z_0)$ . In the second case,  $Z_0$  is set to the value  $z_0$  and  $Z_1$  is set to the canonical value  $r_1(z_0^*)$  that obtains when  $Z_0$  is set to  $z_0^*$ . This corresponds to the notion of "natural indirect effect" in Pearl (2001) and Didelez, Dawid, and Geneletti (2006) and to the notion of "pure indirect effect" in Robins and Greenland (1992) and Robins (2003).

It is necessary but not sufficient for our notion of indirect causality that  $\mathcal{X}_0$  directly cause  $\mathcal{X}_1$  and that  $\mathcal{X}_1$  directly cause  $\mathcal{X}_2$ . We emphasize that transitivity of causation is not guaranteed here, unlike classical treatments such as SGS (p. 42), where transitivity of causation is axiomatic. Instead, transitivity depends on the response functions. For example, if  $r_1(z_0) = \max(z_0, 0)$  and  $r_2(z_0, z_1) = \min(z_1, 0)$ , then  $\mathcal{X}_0 \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_1$  and  $\mathcal{X}_1 \xrightarrow{D}_{\mathcal{S}} \mathcal{X}_2$ , but  $\mathcal{X}_0$  does not indirectly cause  $\mathcal{X}_2$ , as  $r_2(z_0, r_1(z_0^*)) = \min(\max(z_0^*, 0), 0) = 0$  for all  $z_0^*$ . With transitivity,  $\mathcal{X}_i$  is an indirect cause of  $\mathcal{X}_j$  if there exists an  $(\mathcal{X}_i, \mathcal{X}_j)$ -path of length greater than 2 (SGS, pp. 44-45). Although this example conveys the basic idea, we work with more refined notions of indirect causality, elaborated below.

In  $G_3$ ,  $\mathcal{X}_1$  is the only  $(\mathcal{X}_0, \mathcal{X}_2)$  intercessor. In the presence of multiple intercessors, we may be interested in indirect causality via one specified variable. Consider, for example, the system illustrated in  $G_4$ .



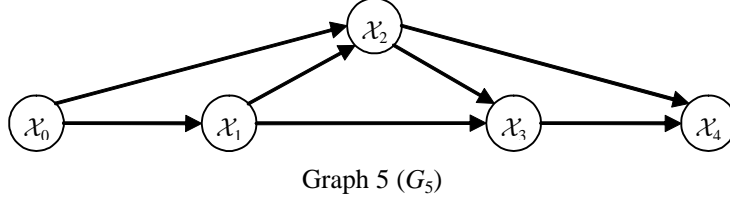
Graph 4 ( $G_4$ )

We say that  $\mathcal{X}_0$  indirectly causes  $\mathcal{X}_3$  via  $\mathcal{X}_1$  if there exists an admissible intervention  $(z_0, r_1(z_0), z_2) \rightarrow (z_0, r_1(z_0^*), z_2)$  such that

$$r_3(z_0, r_1(z_0^*), z_2) - r_3(z_0, r_1(z_0), z_2) \neq 0.$$

If we restrict  $z_2$  to the value  $r_2(z_0)$  in the above difference, we essentially obtain the "path-specific effect transmitted through the path  $\{\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_3\}$ " in Pearl (2001) and Avin, Shpitser, and Pearl (2005).

In the examples above, we considered notions of indirect causality via a single settable variable. More generally, we consider notions of indirect causality via a collection of settable variables, as illustrated in  $G_5$ .



Here, we say that  $\mathcal{X}_0$  indirectly causes  $\mathcal{X}_4$  via  $\mathcal{X}_1$  or  $\mathcal{X}_3$  if there exist an admissible intervention  $(r_2(z_0, r_1(z_0)), r_3(r_1(z_0), r_2(z_0, r_1(z_0)))) \rightarrow (r_2(z_0, r_1(z_0^*)), r_3(r_1(z_0^*), r_2(z_0^*, r_1(z_0^*))))$  such that

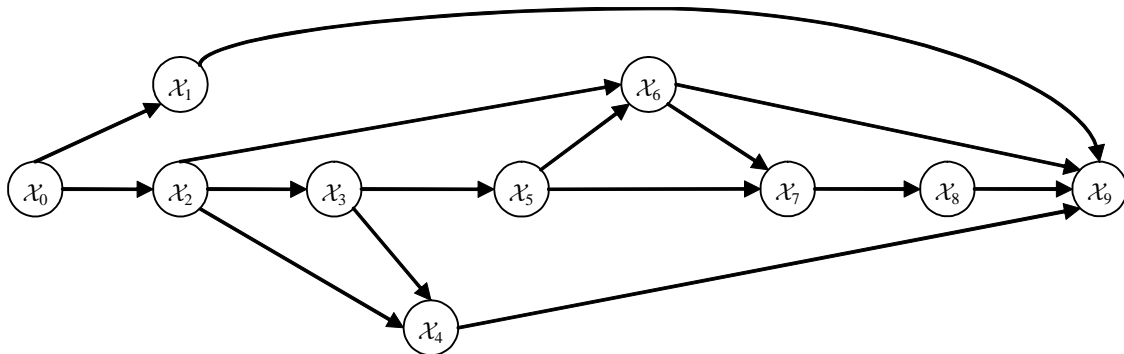
$$r_4(r_2(z_0, r_1(z_0^*)), r_3(r_1(z_0^*), r_2(z_0^*, r_1(z_0^*)))) - r_4(r_2(z_0, r_1(z_0)), r_3(r_1(z_0), r_2(z_0, r_1(z_0)))) \neq 0.$$

(Note that here and elsewhere we simplify notation by omitting response function arguments corresponding to variables that are not direct causes of the specified response.) Setting the first arguments of  $r_4$  to  $r_2(z_0, r_1(z_0^*))$  and  $r_2(z_0, r_1(z_0))$  in the response functions above ensures that the difference in the response for  $\mathcal{X}_4$  is not due to effects transmitted through the path  $\{\mathcal{X}_0, \mathcal{X}_2, \mathcal{X}_4\}$ .

In the general case, the idea underlying indirect causality in recursive systems is essentially the same as in these examples, but to express this rigorously requires particular care and non-trivial further notation. Roughly speaking, we say that  $\mathcal{X}_i$  indirectly causes  $\mathcal{X}_j$  via  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors  $\mathcal{X}_A$ , if the response of  $\mathcal{X}_j$  differs when the effects of setting  $\mathcal{X}_i$  to the value  $z_i$  as opposed to  $z_i^*$  are not transmitted directly, but only through  $\mathcal{X}_A$ .

In order to study the response of  $\mathcal{X}_j$  under the relevant scenarios, we partition the  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors in a recursive manner relative to  $\mathcal{X}_A$ . We distinguish (a) the  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors that belong to paths through  $\mathcal{X}_A$  and (b) those that don't. Among the former, we distinguish: (a.i) the variables that strictly precede  $\mathcal{X}_A$ ; (a.ii)  $\mathcal{X}_A$ ; (a.iii) the variables that intercede elements of  $\mathcal{X}_A$ ; and (a.iv) the variables that strictly succeed  $\mathcal{X}_A$ . The structure of this partition of the  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors enables us to provide general definitions for (indirect) causality via  $\mathcal{X}_A$ .

For illustration, we employ system  $\mathcal{S}_6$ , illustrated in graph  $G_6$ , where  $\Pi_1 = \{1, 2\}$  and  $\Pi_b = \{b + 1\}$  for  $b = 2, \dots, 8$ . The complexity of this example is not capricious. This is the simplest system permitting a full illustration of the relationships that must be considered in a general definition of indirect causality.



Graph 6 ( $G_6$ )

To begin the illustration, take  $b_1 < b_2$ ,  $i \in \Pi_{b_1}$ ,  $j \in \Pi_{b_2}$ . For example, in  $\mathcal{S}_6$ , let  $b_1 = 1$  and  $b_2 = 8$ , let  $i = 2$  (the second element of  $\Pi_1 = \{1, 2\}$ ), and let  $j = 9$  (the sole element of  $\Pi_8$ ). We denote by  $ind(\mathcal{I}_{i:j})$  the indexes of the elements of the  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors  $\mathcal{I}_{i:j}$ . For example, in  $\mathcal{S}_6$ , we have  $ind(\mathcal{I}_{2:9}) = \{3, 4, 5, 6, 7, 8\}$ . We treat elements of  $\Pi_{[0:b_2]}$  that do not correspond to  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors as elements of  $\Pi_{[0:b_1]}$  or  $\Pi_{b_2}$  without loss of generality. Thus,  $ind(\mathcal{I}_{i:j}) = \Pi_{[b_1+1:b_2-1]}$ .

Let  $A$  be a subset of  $ind(\mathcal{I}_{i:j})$ . In  $\mathcal{S}_6$ , we can let  $A = \{5, 7\}$ , say. In what follows we permute the arguments of response values  $r_j(z_{[0:b_2-1]})$  for  $\mathcal{X}_j$  to emphasize their recursive ordering in relation to  $\mathcal{X}_i$  and  $\mathcal{X}_A$ .

For given  $k \in A$ , let  $\mathcal{I}_{i:j}^k := \mathcal{I}_{i:k} \cup \{\mathcal{X}_k\} \cup \mathcal{I}_{k:j}$  denote the  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors for paths through  $\mathcal{X}_k$ , and for  $\mathcal{X}_A := \cup_{k \in A} \{\mathcal{X}_k\}$ , let  $\mathcal{I}_{i:j}^A := \cup_{k \in A} \mathcal{I}_{i:j}^k$  denote the  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors for paths through  $\mathcal{X}_A$ . (For  $A = \emptyset$  we let  $\mathcal{I}_{i:j}^A = \emptyset$ .) Thus, in  $\mathcal{S}_6$  we have  $ind(\mathcal{I}_{2:9}^5) = ind(\mathcal{I}_{2:9}^7) = \{3, 5, 6, 7, 8\}$  and it follows that  $ind(\mathcal{I}_{2:9}^A) = \{3, 5, 6, 7, 8\}$  as well.

Let  $\mathcal{X}_{\underline{A}} := \mathcal{I}_{i:j} \setminus \mathcal{I}_{i:j}^A$  denote the  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors not belonging to paths through  $\mathcal{X}_A$  and let  $\underline{A}$  denote the set of indexes of the elements of  $\mathcal{X}_{\underline{A}}$ . In system  $\mathcal{S}_6$ ,  $\underline{A} = ind(\mathcal{I}_{2:9}) \setminus ind(\mathcal{I}_{2:9}^A) = \{3, 4, 5, 6, 7, 8\} \setminus \{3, 5, 6, 7, 8\} = \{4\}$ . Thus, we have  $ind(\mathcal{I}_{i:j}) = ind(\mathcal{I}_{i:j}^A) \cup \underline{A}$  and  $ind(\mathcal{I}_{i:j}^A) \cap \underline{A} = \emptyset$ .

We now partition  $\text{ind}(\mathcal{I}_{i;j}^A)$  into four mutually exclusive and collectively exhaustive subsets. First, Let  $\mathcal{X}_{\bar{A}} := \cup_{k,l \in A} \mathcal{I}_{k;l} \setminus \mathcal{X}_A$  denote the *inter- $\mathcal{X}_A$  intercessors excluded from  $\mathcal{X}_A$* , and let  $\bar{A}$  denote the set of indexes of the elements of  $\mathcal{X}_{\bar{A}}$ . In  $\mathcal{S}_6$ , we have  $\bar{A} = \{6\}$ .

Next, we distinguish between the  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors for paths through  $\mathcal{X}_A$  that *strictly* precede or succeed  $\mathcal{X}_A$ . We define the  $\mathcal{X}_A$  *predecessors excluded from  $\mathcal{X}_A \cup \mathcal{X}_{\bar{A}}$*  :

$$\mathcal{P}_{i;j}^A := \cup_{k \in A} \{\mathcal{X}_l \in \mathcal{I}_{i;j}^A \text{ and } \mathcal{X}_l \notin (\mathcal{X}_A \cup \mathcal{X}_{\bar{A}}) : \mathcal{X}_l \text{ precedes } \mathcal{X}_k\},$$

and the  $\mathcal{X}_A$  *successors excluded from  $\mathcal{X}_A \cup \mathcal{X}_{\bar{A}}$*  :

$$\mathcal{S}_{i;j}^A := \cup_{k \in A} \{\mathcal{X}_l \in \mathcal{I}_{i;j}^A \text{ and } \mathcal{X}_l \notin (\mathcal{X}_A \cup \mathcal{X}_{\bar{A}}) : \mathcal{X}_l \text{ succeeds } \mathcal{X}_k\}.$$

In the example illustrated in  $G_6$ , we have  $\text{ind}(\mathcal{P}_{2;9}^A) = \{3\}$  and  $\text{ind}(\mathcal{S}_{2;9}^A) = \{8\}$ .

By construction,  $\text{ind}(\mathcal{I}_{i;j}^A) = \text{ind}(\mathcal{P}_{i;j}^A) \cup A \cup \bar{A} \cup \text{ind}(\mathcal{S}_{i;j}^A)$ , and these subsets are mutually exclusive. We verify this equality in our example, where  $\text{ind}(\mathcal{I}_{2;9}^A) = \{3, 5, 6, 7, 8\}$  and  $\text{ind}(\mathcal{P}_{2;9}^A) \cup A \cup \bar{A} \cup \text{ind}(\mathcal{S}_{2;9}^A) = \{3\} \cup \{5, 7\} \cup \{6\} \cup \{8\}$ . Thus,  $\text{ind}(\mathcal{P}_{i;j}^A)$ ,  $\underline{A}$ ,  $A$ ,  $\bar{A}$ , and  $\text{ind}(\mathcal{S}_{i;j}^A)$  form a partition of  $\text{ind}(\mathcal{I}_{i;j}^A)$ , mutually exclusive and collectively exhaustive.

We now use this partition to represent response values for  $\mathcal{X}_j$  in a convenient form. Recall that  $z_{[0;b_1](i)}$  denotes a vector of values for settings for the vector of settable variables  $\mathcal{X}_{[0;b_1](i)}$  corresponding to  $\Pi_{[0;b_1]} \setminus \{i\}$ . Thus, in  $\mathcal{S}_6$ ,  $z_{[0;1](2)}$  denotes values of settings for  $\mathcal{X}_0$  and  $\mathcal{X}_1$ . Similarly, let  $z_{i:A}$ ,  $z_{\underline{A}}$ ,  $z_A$ ,  $z_{\bar{A}}$ , and  $z_{A;j}$  denote vectors of values of settings for elements of  $\mathcal{P}_{i;j}^A$ ,  $\mathcal{X}_{\underline{A}}$ ,  $\mathcal{X}_A$ ,  $\mathcal{X}_{\bar{A}}$ , and  $\mathcal{S}_{i;j}^A$  respectively. We now slightly abuse notation to represent response values for  $\mathcal{X}_j$  (recall  $j \in \Pi_{b_2}$ ) as

$$r_j(z_{[0;b_1](i)}, z_i, z_{i:A}, z_{\underline{A}}, z_A, z_{\bar{A}}, z_{A;j}) = r_j(z_{[0;b_2-1]}),$$

where the arguments of  $r_j$  have been permuted in a particular way, so as to focus attention on settings of  $\mathcal{X}_i$  and  $\mathcal{X}_A$ .

Observe that when  $A = \text{ind}(\mathcal{I}_{i;j})$ , the sets  $\text{ind}(\mathcal{P}_{i;j}^A)$ ,  $\underline{A}$ ,  $\bar{A}$ , and  $\text{ind}(\mathcal{S}_{i;j}^A)$  are empty and we write  $r_j(z_{[0;b_1](i)}, z_i, z_A) = r_j(z_{[0;b_2-1]})$ . Alternatively, when  $A = \emptyset$ , the sets  $\text{ind}(\mathcal{P}_{i;j}^A)$ ,  $\bar{A}$ , and  $\text{ind}(\mathcal{S}_{i;j}^A)$  are empty, whereas  $\underline{A} = \text{ind}(\mathcal{I}_{i;j})$ , and we write  $r_j(z_{[0;b_1](i)}, z_i, z_{\underline{A}}) = r_j(z_{[0;b_2-1]})$ .

We make use of the recursiveness of  $\mathcal{S}$  and the definitions above to represent vectors of response values for elements of  $\mathcal{P}_{i;j}^A$ ,  $\mathcal{X}_{\underline{A}}$ ,  $\mathcal{X}_A$ ,  $\mathcal{X}_{\bar{A}}$ , and  $\mathcal{S}_{i;j}^A$  respectively in the following form

useful for general definitions of indirect causality:

$$\begin{aligned}
& r_{i:A}(z_{[0:b_1]}(i), z_i), \\
& r_{\underline{A}}(z_{[0:b_1]}(i), z_i, z_{i:A}), \\
& r_A(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\overline{A}}), \\
& r_{\overline{A}}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_A), \text{ and} \\
& r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, z_A, z_{\overline{A}}).
\end{aligned}$$

Here too, we let the elements of these response vectors obtain by recursive substitution. Any given element of one of these vectors depends only on its corresponding predecessors. Thus, although  $z_A$  appears as an argument in  $r_{\overline{A}}$ , only the predecessor elements of  $z_A$  for a given response determine that response. Observe that by definition, an element of  $\mathcal{X}_{\underline{A}}$  can not directly cause elements of  $\mathcal{P}_{i;j}^A$ ,  $\mathcal{X}_A$ , or  $\mathcal{X}_{\overline{A}}$  nor can it be directly caused by elements of  $\mathcal{X}_A$ ,  $\mathcal{X}_{\overline{A}}$ , or  $\mathcal{S}_{i;j}^A$ .

Finally, we introduce a notation for canonical settings defined as responses to specific setting values:

$$\begin{aligned}
y_{i:A} &= r_{i:A}(z_{[0:b_1]}(i), z_i), & y_{i:A}^* &= r_{i:A}(z_{[0:b_1]}(i), z_i^*), \\
y_{\underline{A}} &= r_{\underline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}), & y_{\underline{A}}^* &= r_{\underline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*), \\
& & \text{and} & \\
y_A &= r_A(z_{[0:b_1]}(i), z_i, y_{i:A}, y_{\overline{A}}), & y_A^* &= r_A(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, y_{\overline{A}}^*), \\
y_{\overline{A}} &= r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, y_A), & y_{\overline{A}}^* &= r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, y_A^*).
\end{aligned}$$

We can now state our first definition of indirect causality.

**Definition 2.6** *Indirect Causality via  $\mathcal{X}_A$*  Let  $\mathcal{S}$  be recursive. For given non-negative integers  $b_1$  and  $b_2$  with  $b_1 < b_2$ , let  $i \in \Pi_{b_1}$ , let  $j \in \Pi_{b_2}$ , and let  $A$  be a subset of  $\text{ind}(\mathcal{I}_{i;j})$ . Then  $\mathcal{X}_i$  **indirectly causes  $\mathcal{X}_j$  via  $\mathcal{X}_A$  in  $\mathcal{S}$**  if there exists an admissible intervention to  $(\mathcal{X}_{[0:b_1]}(i), \mathcal{X}_i, \mathcal{P}_{i;j}^A, \mathcal{X}_{\underline{A}}, \mathcal{X}_A, \mathcal{X}_{\overline{A}}, \mathcal{S}_{i;j}^A)$  with corresponding responses for  $\mathcal{X}_j$  such that

$$\begin{aligned}
& r_j(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, z_{i:A}, y_A^*), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, z_{i:A}, y_A^*))) \\
& - r_j(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, z_{i:A}, y_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, z_{i:A}, y_A))) \neq 0;
\end{aligned}$$

and we write  $\mathcal{X}_i \stackrel{I[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ . Otherwise, we say that  $\mathcal{X}_i$  **does not indirectly cause**  $\mathcal{X}_j$  via  $\mathcal{X}_A$  in  $\mathcal{S}$  and we write  $\mathcal{X}_i \not\stackrel{I[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ . When  $A = \text{ind}(\mathcal{I}_{i;j})$  and  $\mathcal{X}_i \stackrel{I[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ , we say  $\mathcal{X}_i$  **indirectly causes**  $\mathcal{X}_j$  in  $\mathcal{S}$  and we write  $\mathcal{X}_i \stackrel{I}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ ; when  $A = \text{ind}(\mathcal{I}_{i;j})$  and  $\mathcal{X}_i \not\stackrel{I[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ , we say that  $\mathcal{X}_i$  **does not indirectly cause**  $\mathcal{X}_j$  in  $\mathcal{S}$  and we write  $\mathcal{X}_i \not\stackrel{I}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ .

Consider system  $\mathcal{S}_6$  illustrated in  $G_6$ , for example. With  $A = \{5, 7\}$ , Definition 2.6 states that  $\mathcal{X}_2 \stackrel{I[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_9$  if there exists an admissible intervention  $(z_1, z_4, r_6(z_2, y_5), r_8(y_7)) \rightarrow (z_1, z_4, r_6(z_2, y_5^*), r_8(y_7^*))$  such that

$$r_9(z_1, z_4, r_6(z_2, y_5^*), r_8(y_7^*)) - r_9(z_1, z_4, r_6(z_2, y_5), r_8(y_7)) \neq 0.$$

Intuitively, Definition 2.6 concludes that  $\mathcal{X}_2 \stackrel{I[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_9$  if the response of  $\mathcal{X}_9$  differs when the effects of setting  $\mathcal{X}_2$  to the value  $z_2$  as opposed to  $z_2^*$  are not transmitted directly, but only through  $\mathcal{X}_A$ . Thus, setting values for  $\mathcal{X}_1$  and  $\mathcal{X}_4$  are  $z_1$  and  $z_4$  in both responses of  $\mathcal{X}_9$ . On the other hand, setting values for  $\mathcal{X}_6$  and  $\mathcal{X}_8$  differ across the two responses of  $\mathcal{X}_9$  only in response to different settings of  $(\mathcal{X}_5, \mathcal{X}_7)$ .

When  $\mathcal{X}_i \stackrel{I}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ , it follows that for some non-empty  $A \subset \mathcal{I}_{i;j}$  we have  $\mathcal{X}_i \stackrel{I[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ . The converse need not hold, because  $\mathcal{X}_i$  can indirectly cause  $\mathcal{X}_j$  through each of two distinct intercessors whose associated effects may cancel each other. For example, it may be that  $\mathcal{X}_2$  indirectly causes  $\mathcal{X}_9$  via  $\mathcal{X}_4$  as well as via  $\mathcal{X}_6$  in  $\mathcal{S}_5$  but that  $\mathcal{X}_2$  does not indirectly cause  $\mathcal{X}_9$  via  $\{\mathcal{X}_4, \mathcal{X}_6\}$  in  $\mathcal{S}_5$ .

**Indirect Causality *Exclusive of Given Variables*** We now introduce an indirect causality concept complementary to that above. For example, in the system illustrated in  $G_4$ , we say that  $\mathcal{X}_0$  indirectly causes  $\mathcal{X}_3$  exclusive of  $\mathcal{X}_1$  if there exists an admissible intervention  $(z_0, z_1, r_2(z_0)) \rightarrow (z_0, z_1, r_2(z_0^*))$  such that

$$r_3(z_0, z_1, r_2(z_0^*)) - r_3(z_0, z_1, r_2(z_0)) \neq 0.$$

Similarly, in the system illustrated in  $G_5$  we say that  $\mathcal{X}_0$  indirectly causes  $\mathcal{X}_4$  exclusive of  $\mathcal{X}_1$  and  $\mathcal{X}_3$  if there exists an admissible intervention  $(r_2(z_0, z_1), z_3) \rightarrow (r_2(z_0^*, z_1), z_3)$  such that

$$r_4(r_2(z_0^*, z_1), z_3) - r_4(r_2(z_0, z_1), z_3) \neq 0.$$

More generally, we say that  $\mathcal{X}_i$  indirectly causes  $\mathcal{X}_j$  exclusive of  $(\mathcal{X}_i, \mathcal{X}_j)$  intercessors  $\mathcal{X}_A$  if the response of  $\mathcal{X}_j$  differs when the effects of setting  $\mathcal{X}_i$  to the value  $z_i$  as opposed to  $z_i^*$  are transmitted indirectly through all succeeding variables except  $\mathcal{X}_A$ .

These examples are instances of the following definition.

**Definition 2.7 Indirect Causality Exclusive of  $\mathcal{X}_A$**  Let  $\mathcal{S}$  and  $A$  be as Definition 2.6. Then  $\mathcal{X}_i$  *indirectly causes  $\mathcal{X}_j$  exclusive of  $\mathcal{X}_A$  in  $\mathcal{S}$*  if there exists an admissible intervention to  $(\mathcal{X}_{[0:b_1](i)}, \mathcal{X}_i, \mathcal{P}_{i;j}^A, \mathcal{X}_{\underline{A}}, \mathcal{X}_A, \mathcal{X}_{\overline{A}}, \mathcal{S}_{i;j}^A)$  with corresponding responses for  $\mathcal{X}_j$  such that

$$\begin{aligned} & r_j(z_{[0:b_1](i)}, z_i, y_{i:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_{[0:b_1](i)}, z_i^*, y_{i:A}^*, z_A), \\ & \quad r_{A;j}(z_{[0:b_1](i)}, z_i^*, y_{i:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_{[0:b_1](i)}, z_i^*, y_{i:A}^*, z_A))) \\ & - r_j(z_{[0:b_1](i)}, z_i, y_{i:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_{[0:b_1](i)}, z_i, y_{i:A}, z_A), \\ & \quad r_{A;j}(z_{[0:b_1](i)}, z_i, y_{i:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_{[0:b_1](i)}, z_i, y_{i:A}, z_A))) \neq 0; \end{aligned}$$

and we write  $\mathcal{X}_i \stackrel{I[\sim A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ . Otherwise, we say that  $\mathcal{X}_i$  *does not indirectly cause  $\mathcal{X}_j$  exclusive of  $\mathcal{X}_A$  in  $\mathcal{S}$*  and we write  $\mathcal{X}_i \not\stackrel{I[\sim A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ .

In system  $\mathcal{S}_6$ , Definition 2.7 says that  $\mathcal{X}_2 \stackrel{I[\sim A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_9$  (recall that  $A = \{5, 7\}$ ) if there exists an admissible intervention  $(z_1, y_4, r_6(z_2, z_5), r_8(z_7)) \rightarrow (z_1, y_4^*, r_6(z_2^*, z_5), r_8(z_7))$  such that

$$r_9(z_1, y_4^*, r_6(z_2^*, z_5), r_8(z_7)) - r_9(z_1, y_4, r_6(z_2, z_5), r_8(z_7)) \neq 0.$$

Intuitively, Definition 2.7 concludes that  $\mathcal{X}_2 \stackrel{I[\sim A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_9$  if the response of  $\mathcal{X}_9$  differs when the effects of setting  $\mathcal{X}_2$  to the value  $z_2$  as opposed to  $z_2^*$  are transmitted indirectly through all succeeding variables except through  $\mathcal{X}_A$ .

## 2.4 Total Causality in Recursive Settable Systems

In analyzing relations between causality and conditional independence, it turns out to be important to keep track of channels of both indirect and direct causality. Consider the system illustrated in  $G_3$  for example. There, we say that  $\mathcal{X}_0$  causes  $\mathcal{X}_2$  via  $\mathcal{X}_1$  if there exists an admissible intervention  $(z_0, r_1(z_0)) \rightarrow (z_0^*, r_1(z_0^*))$  such that

$$r_2(z_0^*, r_1(z_0^*)) - r_2(z_0, r_1(z_0)) \neq 0.$$

Intuitively, the response of  $\mathcal{X}_2$  differs when the effect of setting  $Z_0$  to the value  $z_0$  as opposed to  $z_0^*$  is transmitted fully, taking into account both direct and indirect effects. Similarly, in the system illustrated in  $G_4$  we say that  $\mathcal{X}_0$  causes  $\mathcal{X}_3$  via  $\mathcal{X}_1$  if there exists an admissible intervention  $(z_0, r_1(z_0), z_2) \rightarrow (z_0^*, r_1(z_0^*), z_2)$  such that

$$r_2(z_0^*, r_1(z_0^*), z_2) - r_2(z_0, r_1(z_0), z_2) \neq 0.$$

We now provide formal definitions of (total) causality via and exclusive of a set of variables.

**Definition 2.8** *A–Causality* Let  $\mathcal{S}$  and  $A$  be as Definition 2.6. Then  $\mathcal{X}_i$  **causes**  $\mathcal{X}_j$  **via**  $\mathcal{X}_A$  (or  $\mathcal{X}_i$  **A–causes**  $\mathcal{X}_j$ ) **in**  $\mathcal{S}$  if there exists an admissible intervention to  $(\mathcal{X}_{[0:b_1](i)}, \mathcal{X}_i, \mathcal{P}_{i;j}^A, \mathcal{X}_{\underline{A}}, \mathcal{X}_A, \mathcal{X}_{\overline{A}}, \mathcal{S}_{i;j}^A)$  with corresponding responses for  $\mathcal{X}_j$  such that

$$\begin{aligned} & r_j(z_{[0:b_1](i)}, z_i^*, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1](i)}, z_i, z_{i:A}, y_A^*)), \\ & r_{A;j}(z_{[0:b_1](i)}, z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1](i)}, z_i, z_{i:A}, y_A^*))) \\ & - r_j(z_{[0:b_1](i)}, z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1](i)}, z_i, z_{i:A}, y_A)), \\ & r_{A;j}(z_{[0:b_1](i)}, z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1](i)}, z_i, z_{i:A}, y_A))) \neq 0; \end{aligned}$$

and we write  $\mathcal{X}_i \stackrel{[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ . Otherwise, we say that  $\mathcal{X}_i$  **does not A–cause**  $\mathcal{X}_j$  **in**  $\mathcal{S}$  and we write  $\mathcal{X}_i \not\stackrel{[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ . When  $A = \text{ind}(\mathcal{I}_{i;j})$  and  $\mathcal{X}_i \stackrel{[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ , we say  $\mathcal{X}_i$  **causes**  $\mathcal{X}_j$  **in**  $\mathcal{S}$  and we write  $\mathcal{X}_i \Rightarrow_{\mathcal{S}} \mathcal{X}_j$ ; when  $A = \text{ind}(\mathcal{I}_{i;j})$  and  $\mathcal{X}_i \not\stackrel{[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ , we say that  $\mathcal{X}_i$  **does not cause**  $\mathcal{X}_j$  **in**  $\mathcal{S}$  and we write  $\mathcal{X}_i \not\Rightarrow_{\mathcal{S}} \mathcal{X}_j$ .

**Definition 2.9**  *$\sim$  A–Causality* Let  $\mathcal{S}$  and  $A$  be as Definition 2.6. Then  $\mathcal{X}_i$  **causes**  $\mathcal{X}_j$  **exclusive of**  $\mathcal{X}_A$  (or  $\mathcal{X}_i \sim$  **A–causes**  $\mathcal{X}_j$ ) **in**  $\mathcal{S}$  if there exists an admissible intervention to  $(\mathcal{X}_{[0:b_1](i)}, \mathcal{X}_i, \mathcal{P}_{i;j}^A, \mathcal{X}_{\underline{A}}, \mathcal{X}_A, \mathcal{X}_{\overline{A}}, \mathcal{S}_{i;j}^A)$  with corresponding responses for  $\mathcal{X}_j$  such that

$$\begin{aligned} & r_j(z_{[0:b_1](i)}, z_i^*, y_{i:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_{[0:b_1](i)}, z_i^*, y_{i:A}^*, z_A)), \\ & r_{A;j}(z_{[0:b_1](i)}, z_i^*, y_{i:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_{[0:b_1](i)}, z_i^*, y_{i:A}^*, z_A))) \\ & - r_j(z_{[0:b_1](i)}, z_i, y_{i:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_{[0:b_1](i)}, z_i, y_{i:A}, z_A)), \\ & r_{A;j}(z_{[0:b_1](i)}, z_i, y_{i:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_{[0:b_1](i)}, z_i, y_{i:A}, z_A))) \neq 0; \end{aligned}$$

and we write  $\mathcal{X}_i \stackrel{\sim A}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ . Otherwise, we say that  $\mathcal{X}_i$  **does not cause**  $\mathcal{X}_j$  **exclusive of**  $\mathcal{X}_A$  (or  $\mathcal{X}_i$  **does not  $\sim$  A–cause**  $\mathcal{X}_j$ ) **in**  $\mathcal{S}$ , and we write  $\mathcal{X}_i \not\stackrel{\sim A}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ .

Thus, Definitions 2.8 and 2.9 are analogous to Definitions 2.6 and 2.7 with the difference that the direct effect of  $\mathcal{X}_i$  on  $\mathcal{X}_j$  is now further taken into account.

## 2.5 Relations among Total, Direct, and Indirect Causality in Recursive Settable Systems

Our first proposition collects together useful basic results on (indirect) causality via or exclusive of  $\mathcal{X}_A$  for the special cases  $A = \emptyset$  or  $A = \text{ind}(\mathcal{I}_{i;j})$ .

**Proposition 2.1** *Let  $\mathcal{S}$ ,  $i$ , and  $j$  be as Definition 2.6. Let  $A = \emptyset$  and  $B = \text{ind}(\mathcal{I}_{i;j})$ . Then*

- (a)  $\mathcal{X}_i \not\stackrel{I[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ ;
- (b)  $\mathcal{X}_i \not\stackrel{I[\sim B]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ ;
- (c)  $\mathcal{X}_i \stackrel{I[\sim A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  if and only if  $\mathcal{X}_i \stackrel{I[B]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ ;
- (d)  $\mathcal{X}_i \stackrel{\sim A}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  if and only if  $\mathcal{X}_i \stackrel{[B]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ ;
- (e)  $\mathcal{X}_i \stackrel{[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  if and only if  $\mathcal{X}_i \stackrel{\sim B}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ ;
- (f)  $\mathcal{X}_i \stackrel{[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  if and only if  $\mathcal{X}_i \stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ .

It follows from (e) and (f) in Proposition 2.1 that  $\emptyset$ -causality and  $\sim A$ -causality with  $A = \text{ind}(\mathcal{I}_{i;j})$  are equivalent to direct causality in recursive systems.

Our second formal result links  $A$ -causality, direct causality, and indirect causality via  $\mathcal{X}_A$ .

**Proposition 2.2** *Let  $\mathcal{S}$  and  $A$  be as Definition 2.6 and suppose that  $\mathcal{X}_i \stackrel{[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ . Then  $\mathcal{X}_i \stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  or  $\mathcal{X}_i \stackrel{I[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  or both.*

An important special case of Proposition 2.2 occurs when  $A = \text{ind}(\mathcal{I}_{i;j})$ .

**Corollary 2.3** *Let  $\mathcal{S}$  and  $A$  be as Definition 2.6 and suppose that  $\mathcal{X}_i \Rightarrow_{\mathcal{S}} \mathcal{X}_j$ . Then  $\mathcal{X}_i \stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  or  $\mathcal{X}_i \stackrel{I}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  or both.*

Corollary 2.3 verifies the plausible claim that if  $\mathcal{X}_i$  causes  $\mathcal{X}_j$ , it does so directly, indirectly, or both. The converse need not hold, however, as direct and indirect causal channels can cancel one another. Proposition 2.2 extends this proposition to  $A$ -causality. A similar result holds for  $\sim A$ -causality:

**Proposition 2.4** *Let  $\mathcal{S}$  and  $A$  be as Definition 2.6, and suppose that  $\mathcal{X}_i \stackrel{\sim A}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ . Then  $\mathcal{X}_i \stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  or  $\mathcal{X}_i \stackrel{I[\sim A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  or both.*

It is of interest to study even more refined notions of (indirect) causality using this framework. For example, for disjoint subsets  $A$  and  $B$  of  $\text{ind}(\mathcal{I}_{i;j})$ , we can study the notions of  $\mathcal{X}_i$  (indirectly) causing  $\mathcal{X}_j$  (a) via  $A$  and via  $B$ ; (b) via  $A$  and exclusive of  $B$ ; (c) via  $A$  or exclusive of  $B$ ; and (d) exclusive of  $A$  or exclusive of  $B$ . To keep the analysis here tractable, we leave a formal treatment of these causal notions to other work.

### 3 Conditional Independence in Recursive Systems

In this section, we study the interrelations between the notions of causality introduced in Section 2 and that of (conditional) independence. We focus on observational situations in which control is not feasible. Specifically, consider a recursive system  $\mathcal{S}$  and denote by  $Z_{[b]}$  a vector of settings corresponding to elements of  $\Pi_b$ . Suppose that there exist admissible setting values  $(z_0, z_{[1]}, \dots, z_{[b]})$  such that  $z_i = r_i(z_0, z_{[1]}, \dots, z_{[b-1]})$  for each  $z_i$  in  $\text{supp}(Z_i)$  for all  $i \in \Pi_b$ ,  $b = 1, \dots, b_1 - 1$ . Such settings  $Z_i$ ,  $i \in \Pi_{[0:b_1-1]}$ , are *canonical*; we also call the responses  $Y_j = r_j(Z_0, Z_{[1]}, \dots, Z_{[b_1-1]})$ ,  $j \in \Pi_{b_1}$ , *canonical*. It follows that, when it exists, a canonical response  $Y_j$ ,  $j \in \Pi_b$ , is determined entirely by  $Z_0$  and  $\{r_i^{\Pi} : i \in \Pi_{[0:b-1]}, r_j^{\Pi}\}$ . We also call the primary response  $Y_0$  canonical.

#### 3.1 The Conditional Reichenbach Principle of Common Cause

We first state a lemma that plays a key role in formalizing our conditional Reichenbach principle of common cause.

**Lemma 3.1** *Let  $\mathcal{S}$  be recursive,  $j \in \Pi_b$ , and  $A$  a subset of  $\text{ind}(\mathcal{I}_{0;j})$ . Suppose that a canonical response  $Y_j$  for  $\mathcal{X}_j$  exists. If  $\mathcal{X}_0 \not\stackrel{\sim A}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ , then there exists a measurable function  $\tilde{r}_j$  such that*

$$y_j := r_j(z_0, y_{0:A}, y_{\underline{A}}, y_A, y_{\overline{A}}, y_{A;j}) = \tilde{r}_j(y_A).$$

Thus, if  $\mathcal{X}_0 \not\stackrel{\sim A}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  then, provided it exists, we can express a canonical response  $Y_j$  as a function of canonical responses  $Y_A$  ( $Y_A$  must exist if  $Y_j$  exists). A special case of Lemma 3.1 occurs when  $A = \emptyset$  :

**Corollary 3.2** *Let  $\mathcal{S}$  and  $Y_j$  be as in Lemma 3.1. If  $\mathcal{X}_0 \not\Rightarrow_{\mathcal{S}} \mathcal{X}_j$  then  $Y_j$  is constant.*

So far, none of our definitions or results have required any probabilistic elements. Our next result, formalizing Reichenbach’s principle of common cause, relates the functionally defined notion of causality adopted here to the stochastic concept of dependence. This requires us to explicitly introduce probability measures  $P$  on  $(\Omega, \mathcal{F})$ .

**Corollary 3.3 *The Reichenbach Principle of Common Cause*** *Let  $\mathcal{S}$  be recursive. For given  $i \in \Pi_{b_1}$  and  $j \in \Pi_{b_2}$ ,  $b_1, b_2 \geq 0$  ( $i \neq j$ ), let  $\mathcal{X}_i$  and  $\mathcal{X}_j$  be settable variables, and suppose that canonical responses  $Y_i$  and  $Y_j$  exist. For every probability measure  $P$  on  $(\Omega, \mathcal{F})$ , if  $Y_i \not\perp Y_j$ , then either:*

- (i)  $i = 0$  and  $\mathcal{X}_i \Rightarrow_{\mathcal{S}} \mathcal{X}_j$ ; or
- (ii)  $j = 0$  and  $\mathcal{X}_j \Rightarrow_{\mathcal{S}} \mathcal{X}_i$ ; or
- (iii)  $i, j \neq 0$  and  $\mathcal{X}_0 \Rightarrow_{\mathcal{S}} \mathcal{X}_i$  and  $\mathcal{X}_0 \Rightarrow_{\mathcal{S}} \mathcal{X}_j$ .

This result provides a fully explicit statement of conditions, both causal and stochastic, under which the Reichenbach principle of common cause holds – that is, under which it is true that when canonical responses for two settable variables are stochastically dependent, either one variable causes the other or there exists an underlying common cause. Note that while the possibility that one variable causes the other is not explicit in (iii) it is nevertheless implicit, as one way in which we may have  $\mathcal{X}_0 \Rightarrow_{\mathcal{S}} \mathcal{X}_j$  is via the indirect channel  $\mathcal{X}_0 \Rightarrow_{\mathcal{S}} \mathcal{X}_i \Rightarrow_{\mathcal{S}} \mathcal{X}_j$ . If this fails in (iii), then there nevertheless must be a common cause,  $\mathcal{X}_0$ .

Although Reichenbach’s principle holds generally for canonical responses, the proof reveals that this is not a particularly deep fact. The reason is that the primary settable variable  $\mathcal{X}_0$  can always serve as a common cause. Moreover, because the primary setting values  $z_0$  are identified with the underlying elements  $\omega$  of the universe  $\Omega$ , *one cannot dispense with this universal common cause without dispensing with the underlying structure supporting probability statements.*

Another way of appreciating the content of the common cause principle is to examine what it tells us about lack of dependence, via the contrapositive of Corollary 3.3. Specifically, the contrapositive says that, for all probability measures, if two settable variables  $\mathcal{X}_i$

and  $\mathcal{X}_j$  do not cause each other (for  $i = 0$  or  $j = 0$ ) or do not share the common cause  $\mathcal{X}_0$  (for  $i, j \neq 0$ ), then their canonical responses are independent. But Corollary 3.2 states that if  $\mathcal{X}_0 \not\Rightarrow_{\mathcal{S}} \mathcal{X}_j$ , then  $Y_j$  is constant. Knowing that if at least one of two random variables is constant, then the two are independent for all probability measures does not afford deep insight into independence.

Nevertheless, deeper insights emerge when we extend Reichenbach's principle to its conditional counterpart.

**Proposition 3.4** *The Conditional Reichenbach Principle of Common Cause (I)*

Let  $\mathcal{S}$ ,  $Y_i$ , and  $Y_j$  be as in Corollary 3.3. Let  $A \subset \Pi \setminus \{i, j\}$ , let  $\mathcal{X}_A$  be the corresponding vector of settable variables, and suppose that canonical responses  $Y_A$  exist. For every probability measure  $P$  on  $(\Omega, \mathcal{F})$ , if  $Y_i \not\perp Y_j \mid Y_A$  then either:

- (i)  $i = 0$  and with  $A_j := A \cap \text{ind}(\mathcal{I}_{0:j})$ ,  $\mathcal{X}_i \stackrel{\sim A_j}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ ; or
- (ii)  $j = 0$  and with  $A_i := A \cap \text{ind}(\mathcal{I}_{0:i})$ ,  $\mathcal{X}_j \stackrel{\sim A_i}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_i$ ; or
- (iii)  $i, j \neq 0$  and  $\mathcal{X}_0 \stackrel{\sim A_j}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  and  $\mathcal{X}_0 \stackrel{\sim A_i}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_i$ .

Proposition 3.4 is an explicit statement of causal and stochastic conditions that extend Reichenbach's principle to its conditional counterpart. This is significant, as it implies that in recursive causal systems, knowledge of conditional dependence relations such as  $Y_i \not\perp Y_j \mid Y_A$  is informative about the possible causal relations that involve settable variables  $\mathcal{X}_i$  and  $\mathcal{X}_j$ . Proposition 3.4 implies that in recursive systems, in order for two canonical responses  $Y_i$  and  $Y_j$  to be conditionally dependent given a vector of canonical responses  $Y_A$ , it must be that the primary variable  $\mathcal{X}_0$  causes at least  $\mathcal{X}_i$  or  $\mathcal{X}_j$  exclusive of the relevant subsets of  $\mathcal{X}_A$ . Otherwise, we can express  $Y_i$  or  $Y_j$  (or both) as a function of the relevant sub-vector of  $Y_A$ . As Proposition 3.4 has  $A \subset \Pi \setminus \{i, j\}$ , it is necessary that  $0 \notin A$ ;  $Y_i \not\perp Y_j \mid Y_A$  can not hold otherwise. Corollary 3.3 obtains as a special case of Proposition 3.4 for  $A = \emptyset$ .

If the conclusion of Proposition 3.4 holds (regardless of whether the stated conditions hold) then the direct causality graph  $G$  associated with a system  $\mathcal{S}$  has the following useful simple property.

**Proposition 3.5** *Let  $\mathcal{S}$ ,  $\mathcal{X}_i$ ,  $\mathcal{X}_j$ , and  $\mathcal{X}_A$  be as in Proposition 3.4 and let  $G$  be the associated direct causality graph. Suppose that the conclusion of Proposition 3.4 holds. Then there*

exist an  $(\mathcal{X}_0, \mathcal{X}_i)$  path (if  $i \neq 0$ ) and an  $(\mathcal{X}_0, \mathcal{X}_j)$  path (if  $j \neq 0$ ) that does not contain elements of  $\mathcal{X}_A$ .

Thus, the contrapositive of Proposition 3.5 gives conditions sufficient for the conclusion of Proposition 3.4 to fail and therefore for  $Y_i \perp Y_j \mid Y_A$ .

To illustrate, we apply Proposition 3.5 to system  $\mathcal{S}_6$ . Throughout, we consider canonical responses that we assume exist. We have  $Y_0 \perp Y_i \mid Y_2$  for  $i = 3, \dots, 8$ , as  $\mathcal{X}_0 \not\stackrel{\sim\{2\}}{\neq}_{\mathcal{S}} \mathcal{X}_i$  for  $i = 3, \dots, 8$ . Similarly,  $Y_0 \perp Y_9 \mid (Y_1, Y_2)$ , as  $\mathcal{X}_0 \not\stackrel{\sim\{1,2\}}{\neq}_{\mathcal{S}} \mathcal{X}_9$ . Also, we have that  $Y_2 \perp Y_5 \mid Y_3$ , since  $\mathcal{X}_0 \not\stackrel{\sim\{3\}}{\neq}_{\mathcal{S}} \mathcal{X}_5$ . In addition, we have that  $Y_3 \perp Y_5 \mid Y_2$ , as  $\mathcal{X}_0 \not\stackrel{\sim\{2\}}{\neq}_{\mathcal{S}} \mathcal{X}_3$  and  $\mathcal{X}_0 \not\stackrel{\sim\{2\}}{\neq}_{\mathcal{S}} \mathcal{X}_5$ .

Our next remarks assume familiarity with the notion of  $d$ -separation; a discussion of its basic content appears at the outset of the next section. As just noted, we have  $Y_3 \perp Y_5 \mid Y_2$ . If we naively attempt to apply the standard graphical criteria for  $d$ -separation to the direct causality graph  $G_6$ , we would conclude that  $\mathcal{X}_3$  and  $\mathcal{X}_5$  are not  $d$ -separated given  $\mathcal{X}_2$ . If, as is common, "faithfulness" or "stability" (Pearl, 2000, pp. 48-49; SGS, pp. 35, 56) are also assumed, the lack of  $d$ -separation implies  $Y_3 \not\perp Y_5 \mid Y_2$ , inconsistent with  $Y_3 \perp Y_5 \mid Y_2$ . Similarly, we have that  $Y_2 \perp Y_5 \mid (Y_3, Y_6)$ , since  $\mathcal{X}_0 \not\stackrel{\sim\{3\}}{\neq}_{\mathcal{S}} \mathcal{X}_5$ , whereas  $\mathcal{X}_2$  and  $\mathcal{X}_5$  in  $G_6$  are not (naively)  $d$ -separated given  $(\mathcal{X}_3, \mathcal{X}_6)$ , due to the "collider"  $\mathcal{X}_2 \rightarrow \mathcal{X}_6 \leftarrow \mathcal{X}_5$ . Nevertheless, there is no paradox here: graphical criteria for  $d$ -separation apply to a certain class of probabilistic DAGs, *not* to direct causality graphs.

As our use of the qualifier "naive" suggests, care is required in the analysis of  $d$ -separation in settable systems and their associated direct causality graphs. Section 4 provides a formal analysis.

The cases we have just considered only require knowledge of direct causality relations. Knowledge of additional features of the response functions may, however, be important. In particular, the conclusion of Proposition 3.4 can fail even in the presence of  $(\mathcal{X}_0, \mathcal{X}_i)$  paths (if  $i \neq 0$ ) and  $(\mathcal{X}_0, \mathcal{X}_j)$  paths (if  $j \neq 0$ ) that do not contain elements of  $\mathcal{X}_A$ . To illustrate, consider determining whether  $Y_2 \perp Y_3$  in  $\mathcal{S}_6$ . Corollary 3.3 gives that either  $\mathcal{X}_0 \not\stackrel{\neq}{\neq}_{\mathcal{S}} \mathcal{X}_2$  or  $\mathcal{X}_0 \not\stackrel{\neq}{\neq}_{\mathcal{S}} \mathcal{X}_3$  (or both) is sufficient for this to hold. We know from  $G_6$  that  $\mathcal{X}_0 \Rightarrow_{\mathcal{S}} \mathcal{X}_2$ , but determining whether  $\mathcal{X}_0 \not\stackrel{\neq}{\neq}_{\mathcal{S}} \mathcal{X}_3$  requires additional information about the functional form of response functions  $r_2$  and  $r_3$ .

Similarly, consider the claim  $Y_2 \perp Y_7 \mid Y_3$  in  $\mathcal{S}_6$ . From the contrapositive of Proposition 3.4 we know that this will hold if  $\mathcal{X}_0 \not\stackrel{\sim\{3\}}{\neq}_{\mathcal{S}} \mathcal{X}_7$ . Nevertheless, determining whether  $\mathcal{X}_0 \not\stackrel{\sim\{3\}}{\neq}_{\mathcal{S}} \mathcal{X}_7$  unavoidably requires additional information beyond that contained in  $G_6$ , specifically, information about the functional forms of the response functions involved. Because the path  $\{\mathcal{X}_2, \mathcal{X}_6, \mathcal{X}_7\}$  does not contain  $\mathcal{X}_3$ , we have that  $\mathcal{X}_2$  and  $\mathcal{X}_7$  are not naively  $d$ -separated given  $\mathcal{X}_3$  in  $G_6$ . To conclude that  $Y_2 \not\perp Y_3$  and  $Y_2 \not\perp Y_7 \mid Y_3$  in such situations, Pearl (2000, p. 48-49) and SGS (pp. 35, 56) introduce the assumptions of “stability” or “faithfulness” of  $P$ . In sharp contrast, Proposition 3.4 does not impose restrictions on  $P$ ; instead the properties of the response functions play the key role.

### 3.2 A Characterization of the Conditional Reichenbach Principle

The principle of (conditional) common cause gives necessary but not sufficient causal conditions for (conditional) dependence. Thus, its contrapositive gives sufficient but not necessary conditions for (conditional) independence. Specifically,  $Y_i \perp Y_j \mid Y_A$  can hold even when the conclusion of Proposition 3.4 holds. Examples of this are ubiquitous.

**Example 3.6** Consider system  $\mathcal{S}_4$  illustrated in  $G_4$  and suppose that  $Y_1$  and  $Y_2$  are jointly normally distributed with mean zero and variance one. Then  $Y_1 \perp Y_2$  are independent if and only if  $\rho$ , the correlation between  $Y_1$  and  $Y_2$ , is zero. In this case,  $Y_1 \perp Y_2$  even though  $\mathcal{X}_1$  and  $\mathcal{X}_2$  share the common cause  $\mathcal{X}_0$ .

It is also easy to construct examples in which independence holds between directly causally related variables.

**Example 3.7** Suppose further in Example 3.6 that  $\mathcal{X}_0 \stackrel{D}{\neq}_{\mathcal{S}} \mathcal{X}_3$  and that

$$Z_1 = Y_1, \quad Z_2 = Y_2, \quad \text{and} \quad Y_3 = Z_1 + aZ_2.$$

Then  $Y_3$  and  $Y_2$  are also jointly normally distributed with mean zero. Thus, if  $Y_3$  and  $Y_2$  have zero correlation, then they are independent. But this can be ensured by taking  $a = -\rho$ , where  $\rho$  is the correlation between  $Z_1$  and  $Z_2$ . (Note that  $Y_3$  has non-zero variance as long as  $|a| < 1$ .)

It is thus useful to refine the possibilities for conditional independence to distinguish (a) situations in which causal restrictions among settable variables ensure that their responses are (conditionally) independent for any probability measure and (b) those where (conditional) independence among random variables is due to a particular choice of  $P$ . Direct causality restrictions may be sufficient but are not necessary for (a) to obtain, as discussed above. Also, (b) can obtain due to: (i) a particular choice of  $P$  only; or (ii) both a particular configuration of the response functions and a particular choice of  $P$ . The following definitions are useful for this.

**Definition 3.1** *Conditional Causal Isolation and Conditional  $P$ -Stochastic Isolation* Let  $\mathcal{S}, Y_i, Y_j, Y_A$  be as in Proposition 3.4. Suppose that the conclusion of Proposition 3.4 fails; then  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are **causally isolated given  $\mathcal{X}_A$** . Let  $P$  be a probability measure on  $(\Omega, \mathcal{F})$  and suppose that  $Y_i \perp Y_j \mid Y_A$  when  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are not causally isolated given  $\mathcal{X}_A$ ; then we say that  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are  **$P$ -stochastically isolated given  $\mathcal{X}_A$** .

From Definition 3.1, we have that  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are causally isolated given  $\mathcal{X}_A$  when  $\mathcal{X}_0 \stackrel{\sim A_i}{\not\#}_S \mathcal{X}_i$  or  $\mathcal{X}_0 \stackrel{\sim A_j}{\not\#}_S \mathcal{X}_j$ , where  $A_i$  and  $A_j$  are as in Proposition 3.4. When  $A = \emptyset$ , we say that  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are causally isolated when the conclusion of Proposition 3.3 does not hold, that is, when  $\mathcal{X}_0 \not\#_S \mathcal{X}_i$  or  $\mathcal{X}_0 \not\#_S \mathcal{X}_j$ . Conditional causal isolation arises when, for one or the other of  $\mathcal{X}_i$  and  $\mathcal{X}_j$ , the response functions channel the effects of the fundamental cause  $\mathcal{X}_0$  in just the right way so as to yield canonical responses  $Y_i$  or  $Y_j$  (or both) expressible as a function of the relevant subsets of  $Y_A$  (i.e.,  $Y_{A_i}$  or  $Y_{A_j}$ ).

For variables that are not (conditionally) causally isolated, (conditional) independence (i.e., (conditional)  $P$ -stochastic isolation) can arise either directly from  $P$  alone (as in the case of  $Y_1$  and  $Y_2$  in Example 3.6) or from just the right functional relations between multiple causes (common or direct) and  $P$  (as in the case of  $Y_2$  and  $Y_3$  in Example 3.7).

Conditional  $P$ -stochastic isolation is a nontrivial restriction on  $P$ . Nevertheless, if, given  $Y_A$ , the conditional probabilities of  $Y_i$  and  $Y_j$  are regular (see e.g. Dudley, 2002, p. 341–344), then one can always construct a probability measure  $P^*$  ensuring that  $Y_i$  and  $Y_j$  are conditionally independent given  $Y_A$ , regardless of the causal relations involving  $\mathcal{X}_i$ ,  $\mathcal{X}_j$ , and  $\mathcal{X}_A$  (see proposition III.2.1 of Neveu, 1965, p. 74–75). In particular, if  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are

not causally isolated given  $\mathcal{X}_A$ , then  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are  $P^*$ -stochastically isolated given  $\mathcal{X}_A$ .

We are now ready to state necessary and sufficient conditions for probabilistic conditional dependence among canonical responses  $Y_i$  and  $Y_j$  in recursive systems given any other vector of canonical responses  $Y_A$ .

**Corollary 3.8 *Conditional Reichenbach Principle of Common Cause (II)*** *Suppose the conditions of Proposition 3.4 hold. For given probability measure  $P$  on  $(\Omega, \mathcal{F})$ ,  $Y_i \not\perp Y_j \mid Y_A$  if and only if (a) either:*

- (i)  $i = 0$  and with  $A_j := A \cap \text{ind}(\mathcal{I}_{0:j})$ ,  $\mathcal{X}_i \stackrel{\sim A_j}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$ ; or
- (ii)  $j = 0$  and with  $A_i := A \cap \text{ind}(\mathcal{I}_{0:i})$ ,  $\mathcal{X}_j \stackrel{\sim A_i}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_i$ ; or
- (iii)  $i, j \neq 0$  and  $\mathcal{X}_0 \stackrel{\sim A_j}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  and  $\mathcal{X}_0 \stackrel{\sim A_i}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_i$ ;

and (b)  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are not  $P$ -stochastically isolated given  $\mathcal{X}_A$ .

Stated in the contrapositive, Corollary 3.8 tells us that, for a given probability measure,  $Y_i \perp Y_j \mid Y_A$  if and only if either (a)  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are causally isolated given  $\mathcal{X}_A$  or (b)  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are  $P$ -stochastically isolated given  $\mathcal{X}_A$ . When  $A = \emptyset$ , Corollary 3.8 strengthens Reichenbach's principle of common cause to give necessary and sufficient conditions for stochastic dependence.

### 3.3 The Vector Case

In applications, we are usually interested in conditional independence relations between vectors of variables. Accordingly, we now extend the results of this section to accommodate such vectors.

First, we note that the meaning of the notations  $\mathcal{X}_i \stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  and  $\mathcal{X}_i \not\stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  extends to accommodate disjoint sets of multiple settable variables appearing on the right and left hand sides. For example, if  $A$  and  $B$  are non-empty disjoint collections of indexes, we let  $\mathcal{X}_A$  be a vector of settable variables whose indexes belong to  $A$  and similarly for  $\mathcal{X}_B$ , and we write  $\mathcal{X}_A \stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_B$  if  $\mathcal{X}_i \stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  for some  $i \in A \cap \Pi_{b_1}$  and  $j \in B \cap \Pi_{b_2}$  with  $b_1 < b_2$ . Otherwise, we write  $\mathcal{X}_A \not\stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_B$  indicating that  $\mathcal{X}_i \not\stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  for all  $i \in A \cap \Pi_{b_1}$  and  $j \in B \cap \Pi_{b_2}$  with  $b_1 < b_2$ . Observe that even though  $\mathcal{S}$  is a recursive system, it is possible to have  $\mathcal{X}_A \stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_B$  and  $\mathcal{X}_B \stackrel{D}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_A$ .

Similarly, we can extend the meaning of the notations  $\overset{I[A]}{\Rightarrow}$ ,  $\overset{I[\sim A]}{\Rightarrow}$ ,  $\overset{A}{\Rightarrow}$ , and  $\overset{\sim A}{\Rightarrow}$  and their negations to accommodate disjoint sets of multiple settable variables appearing on the right and left hand sides. To do this requires some further notation. Let  $\mathcal{I}_{A:B} = \cup_{i \in A} \cup_{j \in B} \mathcal{I}_{i:j} \setminus (\mathcal{X}_A \cup \mathcal{X}_B)$  denote the set of  $(\mathcal{X}_A, \mathcal{X}_B)$  intercessors and let  $C \subset \text{ind}(\mathcal{I}_{A:B})$ . For given  $i \in A$  and  $j \in B$ , let  $C_{i:j} = C \cap \text{ind}(\mathcal{I}_{i:j})$ . Then, we say that  $\mathcal{X}_A \overset{C}{\Rightarrow}_S \mathcal{X}_B$  if there exists  $i \in A \cap \Pi_{b_1}$  and  $j \in B \cap \Pi_{b_2}$  with  $b_1 < b_2$ , such that  $\mathcal{X}_i \overset{C_{i:j}}{\Rightarrow}_S \mathcal{X}_j$ . Otherwise, we write  $\mathcal{X}_A \overset{C}{\not\Rightarrow}_S \mathcal{X}_B$ , indicating that  $\mathcal{X}_i \overset{C_{i:j}}{\not\Rightarrow}_S \mathcal{X}_j$  for all  $i \in A \cap \Pi_{b_1}$  and  $j \in B \cap \Pi_{b_2}$  with  $b_1 < b_2$ . The notations other than  $\overset{A}{\Rightarrow}_S$  and  $\overset{A}{\not\Rightarrow}_S$  in the list above are defined analogously for vectors of variables.

The definitions of conditional causal isolation and conditional  $P$ -stochastic isolation generalize to the vector case in the obvious way. Thus, if  $Y_A \perp Y_B \mid Y_C$  when  $\mathcal{X}_A$  and  $\mathcal{X}_B$  are not causally isolated given  $\mathcal{X}_C$  (that is, condition (a) in Theorem 3.9 below holds), then we say that  $\mathcal{X}_A$  and  $\mathcal{X}_B$  are  $P$ -stochastically isolated given  $\mathcal{X}_C$ .

**Theorem 3.9 Conditional Reichenbach Principle of Common Cause (III)** *Let  $\mathcal{S}$  be recursive. Let  $A$  and  $B$  be non-empty disjoint subsets of  $\Pi \cup \Pi_0$  and  $C \subset \Pi \setminus (A \cup B)$ . Let  $\mathcal{X}_A$ ,  $\mathcal{X}_B$ , and  $\mathcal{X}_C$  be the corresponding vectors of settable variables, and suppose that canonical responses  $Y_A$ ,  $Y_B$ , and  $Y_C$  exist. For given probability measure  $P$  on  $(\Omega, \mathcal{F})$ ,  $Y_A \not\perp Y_B \mid Y_C$  if and only if (a) either:*

- (i)  $0 \in A$ , and with  $C_B := C \cap \text{ind}(\mathcal{I}_{\{0\}:B})$ ,  $\mathcal{X}_0 \overset{C_B}{\Rightarrow}_S \mathcal{X}_B$ ; or
- (ii)  $0 \in B$ , and with  $C_A := C \cap \text{ind}(\mathcal{I}_{\{0\}:A})$ ,  $\mathcal{X}_0 \overset{C_A}{\Rightarrow}_S \mathcal{X}_A$ ; or
- (iii)  $0 \notin A \cup B$ , and  $\mathcal{X}_0 \overset{C_A}{\Rightarrow}_S \mathcal{X}_A$  and  $\mathcal{X}_0 \overset{C_B}{\Rightarrow}_S \mathcal{X}_B$ ;

and (b)  $\mathcal{X}_A$  and  $\mathcal{X}_B$  are not  $P$ -stochastically isolated given  $\mathcal{X}_C$ .

Corollary 3.8 obtains as a special case of Theorem 3.9 by taking  $A = \{i\}$  and  $B = \{j\}$ . Corollary 3.3 follows from Theorem 3.9 by further taking  $C = \emptyset$ .

### 3.4 Section Summary

The results of this section serve several purposes. First, they characterize conditional dependence (and thus conditional independence) in terms of functionally defined causal relations, rigorously establishing and refining the Reichenbach principle and extending it to

its conditional counterpart. Second, they demonstrate the indispensable and dramatically simplifying role played by the fundamental variable  $\mathcal{X}_0$  as a universal common cause. Once this role is accepted and understood in the context of settable systems, the content of the unconditional Reichenbach principle is no longer mysterious, nor is it apparently deep. Third, and of particular significance, these results provide a rigorous basis for studying the relations between causal structures and conditional independence. These relations are in turn central to testing hypothesized causal structures and to the identification of causal effects in experimental and non-experimental studies. Fourth, these results provide a rigorous and general framework for understanding the interrelations between notions of causality and graphical separation. We take up this topic in our next section.

## 4 Settable Systems and Graphical Separation

As we discuss in the introduction, implications of  $d$ -separation in probabilistic DAGs have sometimes been ascribed causal intuition (e.g., Pearl, 2000, p. 16-17). Absent other causal relations and expressed in the present notation and nomenclature, these can be stated for canonical responses  $Y_i$  and  $Y_j$  as:

*d.1 Conditioning on canonical responses for variables that fully mediate the effect of  $\mathcal{X}_i$  on  $\mathcal{X}_j$  renders  $Y_i$  and  $Y_j$  conditionally independent;*

*d.2 Conditioning on canonical responses for common causes for  $\mathcal{X}_i$  and  $\mathcal{X}_j$ , or conditioning on canonical responses for variables that fully mediate the effects of these common causes on either (or both)  $\mathcal{X}_i$  or  $\mathcal{X}_j$ , renders  $Y_i$  and  $Y_j$  conditionally independent;*

*d.3 Conditioning on canonical responses of settable variables caused by both  $\mathcal{X}_i$  and  $\mathcal{X}_j$  renders  $Y_i$  and  $Y_j$  conditionally dependent.*

We emphasize that these causal interpretations are problematic in the context of probabilistic DAGs. Nevertheless, as we show in this section, statements of this sort can have validity within the recursive settable systems framework under specific conditions.

First, we consider the validity of statements *d.1* and *d.2* for general recursive settable systems, including systems with structure analogous to non-Markovian PCMs. Second, we discuss special settable systems, analogous to Markovian PCMs, in which the directed local Markov property holds for certain canonical responses. For these special systems,

$d$ -separation identifies exactly the conditional independence statements implied by the directed local Markov property for certain canonical responses. Third, we discuss further special settable systems in which conditional independence statements not necessarily implied by the directed local Markov property may hold for certain canonical responses. For these special systems, the  $D$ -separation criteria discussed in Geiger et. al. (1990) identify exactly the conditional independence statements that hold among certain canonical responses. Finally, we demonstrate that  $d.3.$  can fail without further assumptions, and we provide general conditions under which  $d.3$  does hold.

Notions of  $d$ -separation and  $D$ -separation, as well as their underlying assumptions, are therefore not fundamental to establishing the interrelations between functionally defined causal relations and conditional independence, nor are they a natural starting point or context for this study. Nevertheless, as the results of this section show, they can be helpful for understanding conditional independence relations in special settable systems.

## 4.1 Conditioning on Predecessors

The results of Section 3 demonstrate that  $d.1$  and  $d.2$  hold for an arbitrary choice of  $P$  if  $\mathcal{X}_i$  and  $\mathcal{X}_j$  are conditionally causally isolated. Otherwise,  $d.1$  or  $d.2$  may only hold for specific choices of  $P$  and  $r$ , or for particular canonical responses. For example, in system  $\mathcal{S}_7$  below,  $d.1$  fails except when  $i = 0$ . Similarly, our conditional Reichenbach results demonstrate that one need only condition on  $Y_0$ , or on canonical responses for variables that fully mediate the effects of the primary common causes on either (or both)  $\mathcal{X}_i$  or  $\mathcal{X}_j$ , in order for  $d.2$  to hold.

Thus,  $d.1$  and  $d.2$  are not generally valid in recursive settable systems. Moreover, this is not a weakness or drawback of settable systems, as the results of Section 3 provide the same information (and more) about conditional independence relations as  $d.1$  and  $d.2$  might provide (if true). Nevertheless,  $d.1$  and  $d.2$  do hold in certain special cases, as we see next.

## 4.2 Settable Systems and Markovian PCMs

The PCM (Pearl, 2000, definition 7.1.1) assumes that each "endogenous variable" is determined as a function of its "parents" and "background variables" that are "often unob-

servable" (Pearl, p. 203). For example, Pearl (2000, p. 68) employs a DAG to represent a PCM in which every node is a function of its parents and a non-degenerate unobserved random variable.

In "Markovian models," these "arbitrary distributed random disturbances" are assumed to be "jointly independent" and are not explicitly represented in the DAG (see Pearl, 2000, pp. 68 - 69). Pearl (2000, p. 68) states that "these disturbances represent independent background factors that the investigator chooses not to include in the analysis."

If we interpret  $G_1$  in this way, we assume the existence of jointly independent random variables  $\epsilon_1, \dots, \epsilon_5$  such that  $X_1 = f_1(\epsilon_1)$  and  $X_3 = f_3(X_1, X_2, \epsilon_3)$  for example. The introduction of these disturbances and the assumption of their joint independence are strong assumptions that do not emerge naturally from the system of interest; rather, they seem artificially annexed. In fact, as discussed in WC, the PCM rules out any causal role for the background factors since these are not subject to counterfactual variation (see Pearl 2000, definition 7.1.3). Specifically, the PCM rules out "exogenous" causes. Further, as Dawid (2002, p. 183) notes, "when the additional variables are pure mathematical fictions, introduced merely so as to reproduce the desired probabilistic structure of the domain variables, there seems absolutely no good reason to include them in the model."

Our results in Section 3 do not require the existence of "background variables," jointly independent or not. Further, we do not impose any "Markov" assumptions on the probability measure  $P$ ; instead, the directed local Markov property is a *consequence* of our framework. Specifically, the presence of the fundamental variable  $\mathcal{X}_0$  and Proposition 3.4 imply that this property always holds for canonical responses, provided they exist. For example, in system  $\mathcal{S}_6$  we have that  $Y_3 \perp (Y_0, Y_1) \mid Y_2$ ,  $Y_5 \perp Y_2 \mid Y_3$ , etc. In fact, *additional* conditional independence relations not implied by the directed local Markov property (such as  $Y_3 \perp Y_5 \mid Y_2$  in system  $\mathcal{S}_6$ ) may also hold in settable systems, as we discuss in Sections 3 and 4.3.

Nevertheless, our framework permits special systems analogous to the Markovian and non-Markovian PCM. Our next result, a consequence of Proposition 3.4, establishes conditional independence relations that hold among canonical responses corresponding to variables that need not be causally isolated. Specifically, we consider a settable system with

restrictions on (a) direct causality relations and (b) the probability measure  $P$ , yielding a structure analogous to the Markovian PCM. We then show that the local Markov property holds for certain random variables in this system analogous to endogenous variables in the PCM.

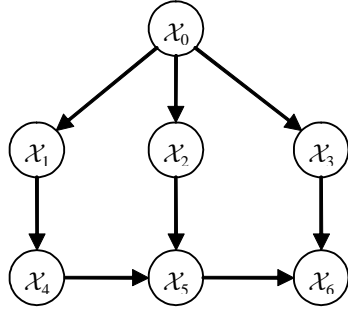
**Proposition 4.1** *Let  $\mathcal{S}$  be recursive. Suppose the elements of  $\Pi_1$  are in one-to-one correspondence with those of  $\Pi_{[2:B]}$ , such that for each  $i \in \Pi_{[2:B]}$ , there exists a unique  $k \in \Pi_1$  such that  $\mathcal{X}_k \stackrel{D}{\Rightarrow} \mathcal{X}_i$ . Suppose further that  $\mathcal{X}_0 \stackrel{D}{\Rightarrow} \mathcal{X}_k$  for all  $k \in \Pi_1$  and  $\mathcal{X}_0 \not\stackrel{D}{\Rightarrow} \mathcal{X}_i$  for all  $i \in \Pi_{[2:B]}$ . For given  $i \in \Pi_b$ ,  $b > 1$ , let  $C := \{l \in \Pi_{[2:b-1]} : \mathcal{X}_l \stackrel{D}{\Rightarrow} \mathcal{X}_i\}$  and let  $A := \{j \in \Pi_{[2:B]} \setminus C : \mathcal{X}_j \text{ does not succeed } \mathcal{X}_i\}$ . Assume that corresponding canonical responses  $Y_i, Y_C$ , and  $Y_A$  exist. Let  $P$  be a probability measure on  $(\Omega, \mathcal{F})$  such that canonical responses  $\{Y_k : k \in \Pi_1\}$  are jointly independent. Then  $Y_i \perp Y_A \mid Y_C$ .*

A special case of Proposition 4.1 obtains for  $C = \emptyset$ , in which case  $Y_i \perp Y_A$  follows.

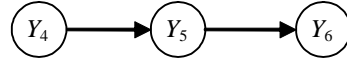
Proposition 4.1 gives restrictions on a settable system sufficient for the directed local Markov property to hold among canonical responses  $\{Y_i : i \in \Pi_{[2:B]}\}$ , provided they exist. It is then easy to construct a probabilistic DAG compatible with the distribution of canonical responses  $\{Y_i : i \in \Pi_{[2:B]}\}$ . This DAG is isomorphic (i.e., has topology identical) to the subgraph of the settable system direct causality graph corresponding to elements of  $\Pi_{[2:B]}$ , substituting canonical responses for settable variables at the nodes. Further, Lauritzen et al. (1990, proposition 3) ensures that for such systems  $d$ -separation or equivalent graphical criteria can identify exactly the conditional independence relations implied by the directed local Markov property.

To demonstrate, consider system  $\mathcal{S}_7$  illustrated in  $G_7$  and its canonical responses, which

we assume exist.



Graph 7 ( $G_7$ )



Graph 8 ( $G_8$ )

With no restrictions on  $P$ ,  $Y_4 \not\perp Y_6 \mid Y_5$  may hold in  $\mathcal{S}_7$ , since  $\mathcal{X}_4$  and  $\mathcal{X}_6$  need be not causally isolated given  $\mathcal{X}_5$ . This is despite the fact that  $\mathcal{X}_5$  fully mediates the effect of  $\mathcal{X}_4$  on  $\mathcal{X}_6$ . Nevertheless, if we assume that  $(Y_1, Y_2, Y_3)$  are jointly independent, then Proposition 4.1 ensures that  $\mathcal{X}_4$  and  $\mathcal{X}_6$  are  $P$ -stochastically isolated given  $\mathcal{X}_5$  and hence that  $Y_4 \perp Y_6 \mid Y_5$ . This is illustrated in the probabilistic DAG  $G_8$  associated with this "Markovian" structure. There,  $Y_5$   $d$ -separates  $Y_4$  and  $Y_6$ .

### 4.3 Deterministic and Chance Nodes

Geiger et. al. (1990) study DAGs that distinguish between "deterministic" and "chance" nodes. A deterministic node corresponds to a random variable that is conditionally independent of all other random variables given its "parents" in the DAG. For example, a deterministic node may represent a variable that is determined as a given function of its parents in the DAG. A chance node corresponds to a random variable that is conditionally independent of its "non-descendants" given its "parents" in the DAG.

Geiger et. al. (1990) call the collection of these conditional independence statements corresponding to deterministic and chance nodes an "enhanced basis" and provide an analogue to  $d$ -separation for these DAGs called " $D$ -separation" that can identify exactly the conditional independence relations implied by an enhanced basis under the graphoid axioms. Similar to probabilistic DAGs, these DAGs do not contain any necessary causal content. Since none of the nodes in Markovian PCM graphs (such as  $G_8$ ) are fully determined by its parents, it follows that  $d$ -separation and  $D$ -separation coincide in such

DAGs.

In settable systems, the introduction of the primary settable variable permits us to dispense with the distinctions between "deterministic" and "chance" nodes, as these are no longer relevant. This eliminates the difficulties for causal discourse that these distinctions create. For example, consider the DAG  $G^*$  isomorphic to the direct causality graph  $G$  for a recursive system  $\mathcal{S}$  that substitutes canonical responses (assumed to exist) for settable variables. Theorem 3.9 gives that if  $C \subset \text{ind}(\mathcal{I}_{0:i})$  is such that  $\mathcal{X}_0 \stackrel{C}{\cong}_{\mathcal{S}} \mathcal{X}_i$  and  $A = \Pi \cup \Pi_0 \setminus \{i\} \cup C$ , then  $Y_i \perp Y_A \mid Y_C$ . In particular,  $\mathcal{X}_0 \stackrel{C}{\cong}_{\mathcal{S}} \mathcal{X}_i$  holds when the set  $C$  corresponds to all direct causes of  $\mathcal{X}_i$ . Thus, whereas  $Y_0$  is a "chance" node in  $G^*$ , the nodes  $Y_i$ ,  $i \neq 0$ , in  $G^*$  are "deterministic." It is easy to verify that for disjoint sets  $D, E$ , and  $F$  in  $\Pi \cup \Pi_0$ ,  $Y_D$  and  $Y_E$  are not  $D$ -separated given  $Y_F$  in  $G^*$  if and only if: (a)(i)  $0 \in D$ , and (ii) for some  $j \in E$ , there exists an  $(\mathcal{X}_0, \mathcal{X}_j)$  path in  $G$  that does not contain elements of  $\mathcal{X}_F$ ; or (b)(i)  $0 \in E$ , and (ii) for some  $i \in D$ , there exists an  $(\mathcal{X}_0, \mathcal{X}_i)$  path in  $G$  that does not contain elements of  $\mathcal{X}_F$ ; or (c)(i)  $0 \notin D \cup E$ , and (ii) (a.ii) and (b.ii) hold.

The graphical  $D$ -separation criteria are sufficient for  $Y_D \perp Y_E \mid Y_F$  but not necessary. The failure of condition (a) in Theorem 3.9 is a more general sufficient condition for  $Y_D \perp Y_E \mid Y_F$ , as it is implied by, but does not imply,  $D$ -separation in  $G^*$ .

For completeness, we next describe a settable system similar to that in Proposition 4.1 that generates random variables forming an "enhanced basis."

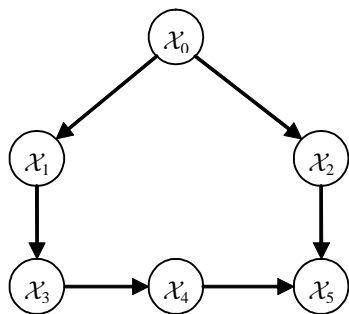
**Proposition 4.2** *Let  $\mathcal{S}$  be recursive. Suppose that for each  $k \in \Pi_1$ , there corresponds a unique  $i \in \Pi_{[1:B]}$  such that  $\mathcal{X}_k \stackrel{D}{\cong} \mathcal{X}_i$ . Suppose further that  $\mathcal{X}_0 \stackrel{D}{\cong} \mathcal{X}_k$  for all  $k \in \Pi_1$  and  $\mathcal{X}_0 \not\stackrel{D}{\cong} \mathcal{X}_i$  for all  $i \in \Pi_{[2:B]}$ . For given  $i \in \Pi_b$ ,  $b > 1$ , let  $C := \{l \in \Pi_{[2:b-1]} : \mathcal{X}_l \stackrel{D}{\cong} \mathcal{X}_i\}$ . Let  $A_1 := \{l \in \Pi_{[2:B]} \setminus C \cup \{i\}\}$  and  $A_2 := \{j \in \Pi_{[2:B]} \setminus C \cup \{i\} : \mathcal{X}_j \text{ does not succeed } \mathcal{X}_i\}$ . Let  $P$  be a probability measure on  $(\Omega, \mathcal{F})$ .*

(i) *Suppose that (a)  $\mathcal{X}_k \not\stackrel{D}{\cong} \mathcal{X}_i$  for all  $k \in \Pi_1$  and (b) canonical settings  $Y_i, Y_C$ , and  $Y_{A_1}$  exist. Then  $Y_i \perp Y_{A_1} \mid Y_C$ .*

(ii) *Suppose that (a)  $\mathcal{X}_k \stackrel{D}{\cong} \mathcal{X}_i$  for some  $k \in \Pi_1$ , (b) canonical settings  $Y_i, Y_C$ , and  $Y_{A_2}$  exist, and (c)  $P$  is such that canonical responses  $\{Y_k : k \in \Pi_1\}$  are jointly independent. Then  $Y_i \perp Y_{A_2} \mid Y_C$ .*

Thus, if  $\{Y_k : k \in \Pi_1\}$  are jointly independent, the settable system described in Proposition 4.2 generates an enhanced basis involving canonical responses  $\{Y_i, i \in \Pi_{[2:B]}\}$  provided they exist. This is represented in the DAG  $G^\dagger$  isomorphic to the subgraph for  $i \in \Pi_{[2:B]}$  of the causal graph  $G$ , substituting canonical responses for settable variables at the nodes. If  $\mathcal{X}_k \stackrel{D}{\not\approx} \mathcal{X}_i$  for all  $k \in \Pi_1$ , then  $Y_i$  is represented by a (dashed) deterministic node in  $G^\dagger$ . Otherwise,  $Y_i$  is represented by a (solid) chance node. Applying the  $D$ -separation criteria to  $G^\dagger$  identifies exactly the conditional independence relations implied by this enhanced basis under the graphoid axioms.

To illustrate, consider system  $\mathcal{S}_9$  illustrated in  $G_9$  and its canonical responses, which we assume exist.



Graph 9 ( $G_9$ )



Graph 10 ( $G_{10}$ )

Since  $\mathcal{X}_0 \stackrel{\sim 3}{\not\approx}_{\mathcal{S}} \mathcal{X}_4$  in  $\mathcal{S}_9$ , Lemma 3.1 ensures that  $Y_4$  is determined as a function of  $Y_3$ . Thus,  $Y_4$  is represented by a "deterministic" (dashed) node in  $G_{10}$ . On the other hand,  $Y_3$  and  $Y_5$  are represented by "chance" (solid) nodes in  $G_{10}$ . Proposition 4.2 gives that  $Y_4 \perp Y_5 \mid Y_3$  for any  $P$ . If  $Y_1$  and  $Y_2$  are independent then Proposition 4.2 ensures that  $\mathcal{X}_3$  and  $\mathcal{X}_5$  are  $P$ -stochastically isolated given  $\mathcal{X}_4$  and hence that  $Y_3 \perp Y_5 \mid Y_4$ .

#### 4.4 Conditioning on Successors

Unlike properties  $d.1$  and  $d.2$ , which concern the conditional independence involving successors conditioning on predecessors,  $d.3$  is a statement about the (lack of) conditional independence of predecessors, conditioning on successors. Thus, although  $d.3$  may indeed be a valid property of recursive systems under given conditions, it will not follow from the conditional Reichenbach principle of common cause.

First, we observe that *d.3* can easily fail in recursive settable systems. For example, for canonical responses in system  $\mathcal{S}_6$ , we have that  $Y_2 \perp Y_5 \mid (Y_3, Y_6)$  since  $\mathcal{X}_0 \stackrel{\sim\{3\}}{\not\sim}_{\mathcal{S}} \mathcal{X}_5$  even though  $\mathcal{X}_2 \stackrel{D}{\Rightarrow} \mathcal{X}_6$  and  $\mathcal{X}_5 \stackrel{D}{\Rightarrow} \mathcal{X}_6$ . It is also easy to construct examples in which *d.3* fails when conditioning only on "common responses."

**Example 4.3** Consider Example 3.7 ( $a \neq 0$ ). Then  $(Y_1, Y_2, Y_3)' \sim N(0, \Sigma)$ , where

$$\Sigma := \begin{bmatrix} 1 & \rho & 1 + a\rho \\ \rho & 1 & \rho + a \\ 1 + a\rho & \rho + a & 1 + 2a\rho + a^2 \end{bmatrix}.$$

Further,  $Y_1 \perp Y_2 \mid Y_3$  if and only if  $|\rho| = 1$ .

In Example 4.3, when  $|\rho| = 1$  it follows that  $Y_1 \perp Y_2 \mid Y_3$  even though  $\mathcal{X}_1 \stackrel{D}{\Rightarrow} \mathcal{X}_3$  and  $\mathcal{X}_2 \stackrel{D}{\Rightarrow} \mathcal{X}_3$ . Pearl (2000, p. 48-49) and SGS (pp. 35, 56) introduce the assumptions of "stability" or "faithfulness" of  $P$  to rule out such situations, which they view as unlikely. Without doubt, the case  $|\rho| = 1$  is a special one.

Our next result gives general conditions under which an extension of *d.3* holds. The extension permits conditioning on both successors and non-successors.

**Theorem 4.4** Let  $\mathcal{S}$  be recursive. Let  $A, B, C$ , and  $D$  be disjoint subsets of  $\Pi$  such that for all  $i \in D$  and  $j \in A \cup B \cup C$ ,  $\mathcal{X}_i$  does not precede  $\mathcal{X}_j$ . Suppose that there exist canonical settings  $Y_A, Y_B, Y_C$ , and  $Y_D$  taking values in supports  $\mathbb{S}_A, \mathbb{S}_B, \mathbb{S}_C$ , and  $\mathbb{S}_D$  respectively, such that  $Y_D = f(Y_A, Y_B, Y_C)$ . Suppose further that there exists sets  $S_A \subseteq \mathbb{S}_A$ ,  $S_B \subseteq \mathbb{S}_B$ , and  $S_{C,D} \subseteq \mathbb{S}_C \times \mathbb{S}_D$  and  $0 \leq \alpha, \beta \leq 1$  such that (i)

$$\begin{aligned} P[(Y_C, Y_D) \in S_{C,D}] &> 0, \\ P[Y_A \in S_A, (Y_C, Y_D) \in S_{C,D}] &= \alpha, \\ P[Y_B \in S_B, (Y_C, Y_D) \in S_{C,D}] &= \beta; \end{aligned}$$

and (ii)  $P[Y_A \in S_A, Y_B \in S_B, (Y_C, Y_D) \in S_{C,D}] \neq \alpha\beta$ . Then  $Y_A \not\perp Y_B \mid (Y_C, Y_D)$ .

A straightforward way to ensure conditions (i) and (ii) is to choose  $S_A, S_B$ , and  $S_{C,D}$  such that  $\alpha\beta > 0$ , and  $Y_A \in S_A$  and  $Y_B \in S_B$  imply  $(Y_C, Y_D) \notin S_{C,D}$ .

This result generalizes *d.3*, as it permits conditioning on both successors and non-successors. For the successors only case ( $C = \emptyset$ ), Theorem 4.4 gives conditions under which *d.3* holds, involving  $f$  and the distributions of  $Y_A, Y_B$ , and  $Y_D$ .

We illustrate this with the following example.

**Example 4.5** *Let the conditions of Example 3.7 hold, but suppose that  $Z_1$  and  $Z_2$  are continuously distributed with standard uniform marginal distributions,  $U[0, 1]$ , such that  $P[Z_1 \neq Z_2] = 1$  and that  $a \neq 0$ . Then  $Y_1 \not\perp Y_2 \mid Y_3$ .*

In Example 4.5, *d.3* holds. We assume the continuity of  $Z_1$  and  $Z_2$  for concreteness; discrete  $Z_1$  and  $Z_2$  can be treated similarly. We impose  $P[Z_1 \neq Z_2] = 1$  for convenience;  $P[Z_1 \neq Z_2] > 0$  suffices, but the stronger condition enables a simpler demonstration. Key aspects of this example are that both  $Z_1$  and  $Z_2$  exhibit non-trivial random variation, neither determines the other, and  $r_3$  ( $r_3(z_1, z_2) = z_1 + az_2$ ) is not a constant function of either of its arguments.

## 5 Conclusion

We study the interrelations between independence or conditional independence and causal relations that hold among variables of interest within the *settable system* framework. We provide rigorous functional definitions of *direct* and *indirect causality* as well as notions of causality *via* a set of variables and *exclusive of* a set of variables. These definitions add to and extend the definitions provided in Robins and Greenland (1992), SGS, Pearl (2000, 2001), Robins (2003), Avin, Shpitser, and Pearl (2005), Didelez, Dawid, and Geneletti (2006), and Geneletti (2007). We provide a proof for the *Reichenbach principle of common cause*, and we introduce and prove its conditional counterpart, the *conditional Reichenbach principle of common cause*. We distinguish between situations in which causal restrictions among settable variables ensure that their responses are (conditionally) independent for any probability measure and those where conditional independence among random variables is due to a particular choice of the probability measure. We introduce concepts of (*conditional*) *causal* and *stochastic isolation* to support these results. We then state necessary and

sufficient conditions for (conditional) dependence among certain random vectors in settable systems.

We relate our results to the (Markovian) PCM,  $d$ -separation, and  $D$ -separation in the artificial intelligence literature. In particular, we study the validity of causal intuitions attributed to  $d$ -separation in recursive settable systems, demonstrate that they can fail, and provide conditions under which they are valid. Taken together, our results show that recursive settable systems constitute an appropriate fundamental framework for studying the interrelations between functionally defined causal relations and conditional independence, and that "background variables", the Markov properties, "enhanced bases," "chance" and "deterministic" nodes, and the assumption of "faithfulness" (SGS, 1993) or "stability" (Pearl, 2000) are not fundamental to establishing these interrelations. Nevertheless, we demonstrate that these notions may be helpful for understanding conditional independence relations in special settable systems.

We focus attention here primarily on recursive systems. An interesting direction for further research is to extend our concepts and results to non-recursive systems (see, e.g., Lauritzen and Richardson, 2002; White and Chalak, 2008). The results of the present work also have direct implications for the structural identification of causal effects in observational studies with use of conditioning instruments and predictive proxies discussed in Chalak and White (2007a, 2007b). Our framework also constitutes an appropriate foundation for studying the identification of direct, indirect, and "path-specific" causal effects (See Avin, Shpitser, and Pearl, 2005; Didelez, Dawid, and Geneletti, 2006; Geneletti, 2007). Finally, our results have direct implications for suggesting and testing for causal models as well as for notions of Granger and Sims causality (Granger, 1969; Sims, 1972). We leave these studies to other work.

## 6 Mathematical Appendix

### Proof of Proposition 2.1

(a) We have  $\text{ind}(\mathcal{I}_{i;j}) = \underline{A}$ , and thus  $r_j(z_{[0:b_2-1]}) = r_j(z_{[0:b_1](i)}, z_i, z_{\underline{A}})$ . It follows from Definition 2.6 that  $\mathcal{X}_i \stackrel{I[\underline{A}]}{\not\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  since  $r_j(z_{[0:b_1](i)}, z_i, z_{\underline{A}}) - r_j(z_{[0:b_1](i)}, z_i, z_{\underline{A}}) = 0$  for all function arguments.

(b) We have  $B = \text{ind}(\mathcal{I}_{i;j})$ , and thus  $r_j(z_{[0:b_2-1]}) = r_j(z_{[0:b_1](i)}, z_i, z_B)$ . It follows from Definition 2.7 that  $\mathcal{X}_i \stackrel{I[\sim B]}{\not\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  since  $r_j(z_{[0:b_1](i)}, z_i, z_B) - r_j(z_{[0:b_1](i)}, z_i, z_B) = 0$  for all function arguments.

(c) Definition 2.7 gives that  $\mathcal{X}_i \stackrel{I[\sim A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  if there exists an admissible intervention  $(z_{[0:b_1](i)}, z_i, y_{\underline{A}}) \rightarrow (z_{[0:b_1](i)}, z_i, y_{\underline{A}}^*)$  such that

$$r_j(z_{[0:b_1](i)}, z_i, y_{\underline{A}}^*) - r_j(z_{[0:b_1](i)}, z_i, y_{\underline{A}}) \neq 0.$$

Also, Definition 2.6 gives that  $\mathcal{X}_i \stackrel{I[B]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  if there exists an admissible intervention  $(z_{[0:b_1](i)}, z_i, y_B) \rightarrow (z_{[0:b_1](i)}, z_i, y_B^*)$  such that

$$r_j(z_{[0:b_1](i)}, z_i, y_B^*) - r_j(z_{[0:b_1](i)}, z_i, y_B) \neq 0.$$

But we have  $\underline{A} = \text{ind}(\mathcal{I}_{i;j}) = B$ . The claim is verified, as the two definitions coincide.

(d) Definition 2.9 gives that  $\mathcal{X}_i \stackrel{\sim A}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  if there exists an admissible intervention  $(z_{[0:b_1](i)}, z_i, y_{\underline{A}}) \rightarrow (z_{[0:b_1](i)}, z_i^*, y_{\underline{A}}^*)$  such that

$$r_j(z_{[0:b_1](i)}, z_i^*, y_{\underline{A}}^*) - r_j(z_{[0:b_1](i)}, z_i, y_{\underline{A}}) \neq 0.$$

Also, Definition 2.8 gives that  $\mathcal{X}_i \stackrel{[B]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  if there exists an admissible intervention  $(z_{[0:b_1](i)}, z_i, y_B) \rightarrow (z_{[0:b_1](i)}, z_i^*, y_B^*)$  such that

$$r_j(z_{[0:b_1](i)}, z_i^*, y_B^*) - r_j(z_{[0:b_1](i)}, z_i, y_B) \neq 0.$$

But we have  $\underline{A} = \text{ind}(\mathcal{I}_{i;j}) = B$ . The claim is verified, as the two definitions coincide.

(e) Definition 2.8 gives that  $\mathcal{X}_i \stackrel{[A]}{\Rightarrow}_{\mathcal{S}} \mathcal{X}_j$  if there exists an admissible intervention  $(z_{[0:b_1](i)}, z_i, z_{\underline{A}}) \rightarrow (z_{[0:b_1](i)}, z_i^*, z_{\underline{A}})$  such that

$$r_j(z_{[0:b_1](i)}, z_i^*, z_{\underline{A}}) - r_j(z_{[0:b_1](i)}, z_i, z_{\underline{A}}) \neq 0.$$

Also, Definition 2.9 gives that  $\mathcal{X}_i \xrightarrow{B} \mathcal{X}_j$  if there exists an admissible intervention  $(z_{[0:b_1]}(i), z_i, z_B) \rightarrow (z_{[0:b_1]}(i), z_i^*, z_B)$  such that

$$r_j(z_{[0:b_1]}(i), z_i^*, z_B) - r_j(z_{[0:b_1]}(i), z_i, z_B) \neq 0.$$

But we have  $\underline{A} = \text{ind}(\mathcal{I}_{i:j}) = B$ . The claim is verified, as the two definitions coincide.

(f) Definition 2.8 gives that  $\mathcal{X}_i \xrightarrow{[A]} \mathcal{X}_j$  if there exists an admissible intervention  $(z_{[0:b_1]}(i), z_i, z_{\underline{A}}) \rightarrow (z_{[0:b_1]}(i), z_i^*, z_{\underline{A}})$  such that

$$r_j(z_{[0:b_1]}(i), z_i^*, z_{\underline{A}}) - r_j(z_{[0:b_1]}(i), z_i, z_{\underline{A}}) \neq 0.$$

Also, Definition 2.3 gives that  $\mathcal{X}_i \xrightarrow{D} \mathcal{X}_j$  if there exists an admissible intervention  $(z_{[0:b_1]}(i), z_i, z_{[b_1+1:b_2-1]}) \rightarrow (z_{[0:b_1]}(i), z_i^*, z_{[b_1+1:b_2-1]})$  such that

$$r_j(z_{[0:b_1]}(i), z_i^*, z_{[b_1+1:b_2-1]}) - r_j(z_{[0:b_1]}(i), z_i, z_{[b_1+1:b_2-1]}) \neq 0.$$

But we have  $\underline{A} = \text{ind}(\mathcal{I}_{i:j}) = \Pi_{[b_1+1:b_2-1]}$ . The claim is verified, as the two definitions coincide. ■

**Proof of Proposition 2.2** We prove the contrapositive. We have:

$$\begin{aligned} & r_j(z_{[0:b_1]}(i), z_i^*, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*))) \\ & - r_j(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A))) \\ = & r_j(z_{[0:b_1]}(i), z_i^*, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*))) \\ & - r_j(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*))) \\ & + r_j(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*))) \\ & - r_j(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A))). \end{aligned}$$

Suppose  $\mathcal{X}_i \not\stackrel{D}{\rightarrow}_S \mathcal{X}_j$ . Then by Definition 2.3, for all admissible interventions with the following corresponding responses for  $\mathcal{X}_j$ , we have

$$\begin{aligned} & r_j(z_{[0:b_1]}(i), z_i^*, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*))) \\ & - r_j(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*))) = 0. \end{aligned}$$

Also, suppose  $\mathcal{X}_i \not\stackrel{I[A]}{\rightarrow}_S \mathcal{X}_j$ . Then by Definition 2.6, for all admissible interventions with the following corresponding responses for  $\mathcal{X}_j$ , we have

$$\begin{aligned} & r_j(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*))) \\ & - r_j(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A))) = 0. \end{aligned}$$

Since the space of jointly admissible setting values of the form

$$\begin{aligned} & (z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*))) \end{aligned}$$

includes the space of jointly admissible setting values of the form

$$\begin{aligned} & (z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A))) \end{aligned}$$

it follows that for all admissible interventions with the following corresponding responses for  $\mathcal{X}_j$ , we have

$$\begin{aligned} & r_j(z_{[0:b_1]}(i), z_i^*, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A^*, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A^*))) \\ & - r_j(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A), \\ & \quad r_{A:j}(z_{[0:b_1]}(i), z_i, z_{i:A}, z_{\underline{A}}, y_A, r_{\overline{A}}(z_0, z_{[1:b_1]}(i), z_i, z_{i:A}, y_A))) = 0, \end{aligned}$$

that is,  $\mathcal{X}_i \stackrel{[A]}{\not\cong}_S \mathcal{X}_j$ . This verifies the contrapositive, so the claimed result follows. ■

**Proof of Corollary 2.3** Apply Proposition 2.2 with  $A = \text{ind}(\mathcal{I}_{i:j})$ . ■

**Proof of Proposition 2.4** We prove the contrapositive. We have:

$$\begin{aligned}
& r_j(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, \underline{y}_A^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, \underline{y}_A^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A))) \\
& - r_j(z_{[0:b_1]}(i), z_i, y_{i:A}, \underline{y}_A, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i, y_{i:A}, \underline{y}_A, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A))) \\
= & r_j(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, \underline{y}_A^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, \underline{y}_A^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A))) \\
& - r_j(z_{[0:b_1]}(i), z_i, y_{i:A}, \underline{y}_A, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i, y_{i:A}, \underline{y}_A, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A))) \\
& + r_j(z_{[0:b_1]}(i), z_i, y_{i:A}, \underline{y}_A, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, \underline{y}_A^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A))) \\
& - r_j(z_{[0:b_1]}(i), z_i, y_{i:A}, \underline{y}_A, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i, y_{i:A}, \underline{y}_A, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A))).
\end{aligned}$$

Suppose  $\mathcal{X}_i \stackrel{D}{\not\cong}_S \mathcal{X}_j$ . Then by Definition 2.3, for all admissible interventions with the following corresponding responses for  $\mathcal{X}_j$ , we have

$$\begin{aligned}
& r_j(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, \underline{y}_A^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, \underline{y}_A^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A))) \\
& - r_j(z_{[0:b_1]}(i), z_i, y_{i:A}, \underline{y}_A, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, \underline{y}_A^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A))) = 0.
\end{aligned}$$

Also, suppose  $\mathcal{X}_i \stackrel{I[\sim A]}{\not\cong}_S \mathcal{X}_j$ . Then by Definition 2.7, for all admissible interventions with

the following corresponding responses for  $\mathcal{X}_j$ , we have

$$\begin{aligned}
& r_j(z_{[0:b_1]}(i), z_i, y_{i:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A))) \\
& - r_j(z_{[0:b_1]}(i), z_i, y_{i:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i, y_{i:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A))) = 0.
\end{aligned}$$

Since the space of jointly admissible setting values of the form

$$\begin{aligned}
& (z_{[0:b_1]}(i), z_i, y_{i:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A)))
\end{aligned}$$

includes the space of jointly admissible setting values of the form

$$\begin{aligned}
& (z_{[0:b_1]}(i), z_i, y_{i:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i, y_{i:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A)))
\end{aligned}$$

it follows that for all admissible interventions with the following corresponding responses for  $\mathcal{X}_j$ , we have

$$\begin{aligned}
& r_j(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i^*, y_{i:A}^*, z_A))) \\
& - r_j(z_{[0:b_1]}(i), z_i, y_{i:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A), \\
& \quad r_{A:j}(z_{[0:b_1]}(i), z_i, y_{i:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_{[0:b_1]}(i), z_i, y_{i:A}, z_A))) = 0,
\end{aligned}$$

that is,  $\mathcal{X}_i \not\stackrel{\sim A}{\mathcal{F}}_S \mathcal{X}_j$ . This verifies the contrapositive, so the claimed result follows.  $\blacksquare$

**Proof of Lemma 3.1** Denote by  $\mathbb{S}^*$  the space of jointly admissible settings for  $(\mathcal{X}_0, \mathcal{P}_{0:j}^A, \mathcal{X}_{\underline{A}}, \mathcal{X}_A, \mathcal{X}_{\overline{A}}, \mathcal{S}_{0:j}^A)$  of the form  $(z_0^*, y_{0:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_0^*, y_{0:A}^*, z_A), r_{A:j}(z_0^*, y_{0:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_0^*, y_{0:A}^*, z_A)))$ . Since  $Y_j$  exists,  $\mathbb{S}^*$  is not empty.

First, suppose that  $\mathbb{S}^*$  is a singleton. Then there does not exist an admissible intervention to  $(\mathcal{X}_0, \mathcal{P}_{0:j}^A, \mathcal{X}_{\underline{A}}, \mathcal{X}_A, \mathcal{X}_{\overline{A}}, \mathcal{S}_{0:j}^A)$  of the specified form and thus  $\mathcal{X}_0 \not\stackrel{\sim A}{\mathcal{F}}_S \mathcal{X}_j$ . It follows trivially that there exists a measurable function  $\tilde{r}_j$  such that

$$y_j = r_j(z_0, y_{0:A}, y_{\underline{A}}, y_A, y_{\overline{A}}, y_{A:j}) = \tilde{r}_j(y_A).$$

Second, suppose that  $\mathbb{S}^*$  is a multi-element set and that  $\mathcal{X}_0 \not\stackrel{\sim A}{\neq}_S \mathcal{X}_j$ . Then by Definition 2.9 for all admissible interventions to  $(\mathcal{X}_0, \mathcal{P}_{0:j}^A, \mathcal{X}_{\underline{A}}, \mathcal{X}_A, \mathcal{X}_{\overline{A}}, \mathcal{S}_{0:j}^A)$  with the following corresponding responses for  $\mathcal{X}_j$  we have

$$\begin{aligned} & r_j(z_0^*, y_{0:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_0^*, y_{0:A}^*, z_A), r_{A:j}(z_0^*, y_{0:A}^*, y_{\underline{A}}^*, z_A, r_{\overline{A}}(z_0^*, y_{0:A}^*, z_A))) \\ & - r_j(z_0, y_{0:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_0, y_{0:A}, z_A), r_{A:j}(z_0, y_{0:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_0, y_{0:A}, z_A))) = 0. \end{aligned}$$

Therefore there exists a measurable function  $z_A \rightarrow \tilde{r}_j(z_A)$  such that for all elements of  $\mathbb{S}^*$

$$r_j(z_0, y_{0:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_0, y_{0:A}, z_A), r_{A:j}(z_0, y_{0:A}, y_{\underline{A}}, z_A, r_{\overline{A}}(z_0, y_{0:A}, z_A))) = \tilde{r}_j(z_A).$$

In particular, for all  $(z_0, y_{0:A}, y_{\underline{A}}, y_A, y_{\overline{A}}, y_{A:j}) \in \mathbb{S}^*$  we have

$$r_j(z_0, y_{0:A}, y_{\underline{A}}, y_A, y_{\overline{A}}, y_{A:j}) = \tilde{r}_j(y_A). \blacksquare$$

**Proof of Corollary 3.2** Let  $A = \emptyset$ . Proposition 2.1(d) gives that  $\mathcal{X}_0 \not\stackrel{\sim A}{\neq}_S \mathcal{X}_j$  if and only if  $\mathcal{X}_0 \not\neq_S \mathcal{X}_j$ . Since  $\tilde{r}_j(z_A)$  must be constant, the result follows from Lemma 3.1.  $\blacksquare$

**Proof of Corollary 3.3** Apply Proposition 3.4 with  $A = \emptyset$ . The result follows from Corollary 3.2.  $\blacksquare$

**Proof of Proposition 3.4** Apply Theorem 3.9 with  $A = \{i\}$  and  $B = \{j\}$ .  $\blacksquare$

**Proof of Proposition 3.5** We prove the contrapositive.

(i) Suppose that  $i = 0$  and that there does not exist an  $(\mathcal{X}_0, \mathcal{X}_j)$  path that does not contain elements of  $\mathcal{X}_A$ . Let  $A_j = A \cap \text{ind}(\mathcal{I}_{0:j})$ . Denote by  $\mathbb{S}^*$  the space of jointly admissible settings to  $(\mathcal{X}_0, \mathcal{P}_{0:j}^{A_j}, \mathcal{X}_{\underline{A}_j}, \mathcal{X}_{A_j}, \mathcal{X}_{\overline{A}_j}, \mathcal{S}_{0:j}^{A_j})$  of the form

$$(z_0^*, y_{0:A_j}^*, y_{\underline{A}_j}^*, z_{A_j}, r_{\overline{A}_j}(z_0^*, y_{0:A_j}^*, z_{A_j}), r_{A_j:j}(z_0^*, y_{0:A_j}^*, y_{\underline{A}_j}^*, z_{A_j}, r_{\overline{A}_j}(z_0^*, y_{0:A_j}^*, z_{A_j}))).$$

Since  $Y_j$  exists,  $\mathbb{S}^*$  is not empty.

First, suppose that  $\mathbb{S}^*$  is a singleton. Then there does not exist an admissible intervention to  $(\mathcal{X}_0, \mathcal{P}_{0:j}^{A_j}, \mathcal{X}_{\underline{A}_j}, \mathcal{X}_{A_j}, \mathcal{X}_{\overline{A}_j}, \mathcal{S}_{0:j}^{A_j})$  of the specified form and thus  $\mathcal{X}_0 \not\stackrel{\sim A_j}{\neq}_S \mathcal{X}_j$ , a contradiction.

Second, suppose that  $\mathbb{S}^*$  is a multi-element set. By construction we have that for all admissible interventions to  $(\mathcal{X}_0, \mathcal{P}_{0:j}^{A_j}, \mathcal{X}_{\underline{A}_j}, \mathcal{X}_{A_j}, \mathcal{X}_{\overline{A}_j}, \mathcal{S}_{0:j}^{A_j})$  with the following corresponding

responses for  $\mathcal{X}_j$  we have

$$\begin{aligned} & r_j(z_0^*, y_{0:A_j}^*, \underline{y}_{A_j}^*, z_{A_j}, r_{\overline{A_j}}(z_0^*, y_{0:A_j}^*, z_{A_j}), r_{A_j:j}(z_0^*, y_{0:A_j}^*, \underline{y}_{A_j}^*, z_{A_j}, r_{\overline{A_j}}(z_0^*, y_{0:A_j}^*, z_{A_j}))) \\ & - r_j(z_0, y_{0:A_j}, \underline{y}_{A_j}, z_{A_j}, r_{\overline{A_j}}(z_0, y_{0:A_j}, z_{A_j}), r_{A_j:j}(z_0, y_{0:A_j}, \underline{y}_{A_j}, z_{A_j}, r_{\overline{A_j}}(z_0, y_{0:A_j}, z_{A_j}))) = 0. \end{aligned}$$

Otherwise, it follows from Definition 2.3 of direct causality that there must exist an  $(\mathcal{X}_0, \mathcal{X}_j)$  path that does not contain elements of  $\mathcal{X}_{A_j}$  and therefore of  $\mathcal{X}_A$  by definition of  $A_j$ . It follows from Definition 2.9 that  $\mathcal{X}_0 \not\stackrel{\sim A_j}{\Rightarrow}_S \mathcal{X}_j$ , a contradiction.

(ii) Suppose that  $j = 0$  and that there does not exist an  $(\mathcal{X}_0, \mathcal{X}_i)$  path that does not contain elements of  $\mathcal{X}_A$ . Then an argument parallel to (i) leads to  $\mathcal{X}_0 \stackrel{\sim A_i}{\Rightarrow}_S \mathcal{X}_i$  (with  $A_i := A \cap \text{ind}(\mathcal{I}_{0:i})$ ), a contradiction.

(iii) Suppose that  $i, j \neq 0$  and that there does not exist (a) an  $(\mathcal{X}_0, \mathcal{X}_i)$  path that does not contain elements of  $\mathcal{X}_A$  or (b) an  $(\mathcal{X}_0, \mathcal{X}_j)$  path that does not contain elements of  $\mathcal{X}_A$  (or both). Then arguments parallel to (i) or (ii) (or both) imply that  $\mathcal{X}_0 \stackrel{\sim A_j}{\Rightarrow}_S \mathcal{X}_j$  or  $\mathcal{X}_0 \stackrel{\sim A_i}{\Rightarrow}_S \mathcal{X}_i$  (or both), a contradiction. ■

**Proof of Corollary 3.8** The result is immediate from Proposition 3.4 and the contrapositive of the definition of conditional stochastic isolation. ■

**Proof of Theorem 3.9** Let  $P$  be any probability measure. First, we prove that if  $Y_A \not\perp Y_B | Y_C$  then  $\mathcal{X}_A$  and  $\mathcal{X}_B$  are not causally isolated given  $\mathcal{X}_C$ .

(i) Suppose that  $0 \in A$  and  $\mathcal{X}_0 \not\stackrel{\sim C_B}{\Rightarrow}_S \mathcal{X}_B$ . Then  $\mathcal{X}_0 \not\stackrel{\sim C_{0:j}}{\Rightarrow}_S \mathcal{X}_j$  for all  $j \in B$ . For given  $j \in B$ , let  $\mathcal{X}_{C_{0:j}}$  be a vector of settable variables and let  $Y_{C_{0:j}}$  denote the corresponding canonical responses. By Lemma 3.1, it follows that  $Y_j = \tilde{r}_j(Y_{C_{0:j}})$  for all  $j \in B$ . Let  $\mathcal{X}_{C_B}$  be a vector of settable variables and let  $Y_{C_B}$  denote the corresponding canonical response; then we have  $Y_B = \tilde{r}_B(Y_{C_B})$ . Let  $C_B^c = C \setminus C_B$ , let  $\mathcal{X}_{C_B^c}$  be a vector of settable variables, and let  $Y_{C_B^c}$  denote the corresponding canonical responses; then  $Y_C = (Y_{C_B}, Y_{C_B^c})$ . Since  $Y_B = \tilde{r}_B(Y_{C_B})$ , we have that  $(Y_A, Y_{C_B^c}) \perp Y_B | Y_{C_B}$ . We then have that  $Y_A \perp Y_B | (Y_{C_B}, Y_{C_B^c})$  (see, for example, Dawid, 1979, section 4; Dohler, 1980, lemma 3; Smith 1989, property 3; and Florens, Mouchart, and Rolin 1990, theorem 2.2.10), that is,  $Y_A \perp Y_B | Y_C$ , a contradiction. (Note that when  $C_B = C$  the result is immediate. Also, when  $C_B = \emptyset$ ,  $Y_B$  is constant and the result is trivial.)

(ii) Suppose  $0 \in B$ , and that  $\mathcal{X}_0 \stackrel{\sim C_A}{\not\approx}_S \mathcal{X}_A$ . The result is symmetric to (i) yielding that  $Y_A \perp Y_B \mid Y_C$ , a contradiction.

(iii) Suppose that  $0 \notin A \cup B$ , and that  $\mathcal{X}_0 \stackrel{\sim C_A}{\not\approx}_S \mathcal{X}_A$  or  $\mathcal{X}_0 \stackrel{\sim C_B}{\not\approx}_S \mathcal{X}_B$ . Suppose that  $\mathcal{X}_0 \stackrel{\sim C_A}{\not\approx}_S \mathcal{X}_A$ ; then an argument similar to (i) gives that  $Y_A \perp Y_B \mid Y_C$ , a contradiction. Alternatively, suppose that  $\mathcal{X}_0 \stackrel{\sim C_B}{\not\approx}_S \mathcal{X}_B$ . Then by a parallel argument, we obtain that  $Y_A \perp Y_B \mid Y_C$ , a contradiction.

That  $\mathcal{X}_A$  and  $\mathcal{X}_B$  are not stochastically isolated given  $\mathcal{X}_C$  follows by the definition of conditional stochastic isolation. The rest of the proof follows from (the contrapositive of) the definition of conditional stochastic isolation. ■

**Proof of Proposition 4.1** Let  $k \in \Pi_1$  such that  $\mathcal{X}_k \stackrel{D}{\cong} \mathcal{X}_i$ . By construction we have that  $\mathcal{X}_0 \stackrel{\sim C \cup \{k\}}{\not\approx}_S \mathcal{X}_i$ . Theorem 3.9 gives that  $Y_i \perp Y_A \mid (Y_C, Y_k)$ . Further, since elements of  $\mathcal{X}_C$  and  $\mathcal{X}_A$  do not succeed  $\mathcal{X}_i$ , there exists a set  $D \subset \Pi_1 \setminus \{k\}$  such that  $\mathcal{X}_0 \stackrel{\sim D}{\not\approx}_S (\mathcal{X}_C, \mathcal{X}_A)$ . It follows from Lemma 3.1 that there exists a measurable function  $\tilde{r}_{C,A}$  such that  $(y_C, y_A) = \tilde{r}_{C,A}(y_D)$ . Since  $\{Y_k : k \in \Pi_1\}$  are jointly independent we have that  $Y_k \perp Y_D$ . It follows from Dawid (1979, lemma 4.2(i)) that  $Y_k \perp (Y_C, Y_A)$ . Also, Dawid, 1979, section 4 (see also Dohler, 1980, lemma 3; Smith 1989, property 3; and Florens, Mouchart, and Rolin 1990, theorem 2.2.10) gives that  $Y_k \perp Y_A \mid Y_C$ . Given that  $Y_i \perp Y_A \mid (Y_C, Y_k)$ , Dawid (1979, lemma 4.3) gives that  $Y_i \perp Y_A \mid Y_C$ . ■

**Proof of Proposition 4.2** (i) By construction we have that  $\mathcal{X}_0 \stackrel{\sim C}{\not\approx}_S \mathcal{X}_i$ . It follows from Theorem 3.9 that  $Y_i \perp Y_{A_1} \mid Y_C$ . (ii) An argument similar to Proposition 4.1 gives that  $Y_i \perp Y_{A_2} \mid Y_C$ . ■

**Proof of Example 4.3** The joint normality of  $(Y_1, Y_2, Y_3)'$  holds as a standard property of the normal distribution. The elements of the covariance matrix follow by elementary computations. Let

$$\Sigma_0 := \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

then

$$\Sigma := \begin{bmatrix} 1 & \rho & 1 + a\rho \\ \rho & 1 & \rho + a \\ 1 + a\rho & \rho + a & 1 + 2a\rho + a^2 \end{bmatrix} = \begin{bmatrix} \Sigma_0 & \Sigma_{0,1} \\ \Sigma_{1,0} & \Sigma_1 \end{bmatrix}$$

where  $\Sigma'_{0,1} = \Sigma_{1,0} := (1 + a\rho, \rho + a)$  and  $\Sigma_1 := (1 + 2a\rho + a^2)$ . It is also a standard property of the normal distribution (e.g., Hamilton, 1994, p.100) that

$$(Y_1, Y_2)' | Y_3 \sim N(0, \Sigma_0 - \Sigma_{0,1}\Sigma_1^{-1}\Sigma_{1,0}).$$

For the normal distribution,  $Y_1 \perp Y_2 | Y_3$  if and only if  $E(Y_1Y_2 | Y_3) = 0$ . Computing the off-diagonal element of  $\Sigma_0 - \Sigma_{0,1}\Sigma_1^{-1}\Sigma_{1,0}$  corresponding to  $E(Y_1Y_2 | Y_3)$ , we obtain

$$E(Y_1Y_2 | Y_3) = a(\rho^2 - 1).$$

As  $a \neq 0$ , this equals zero if and only if  $|\rho| = 1$ . ■

**Proof of Theorem 4.4** Since  $P[(Y_C, Y_D) \in S_{C,D}] > 0$ , it follows that

$$P[Y_A \in S_A | (Y_C, Y_D) \in S_{C,D}] = \frac{P[Y_A \in S_A, (Y_C, Y_D) \in S_{C,D}]}{P[(Y_C, Y_D) \in S_{C,D}]} = \frac{\alpha}{P[(Y_C, Y_D) \in S_{C,D}]} \text{ and}$$

$$P[Y_B \in S_B | (Y_C, Y_D) \in S_{C,D}] = \frac{P[Y_B \in S_B, (Y_C, Y_D) \in S_{C,D}]}{P[(Y_C, Y_D) \in S_{C,D}]} = \frac{\beta}{P[(Y_C, Y_D) \in S_{C,D}]},$$

so

$$P[Y_A \in S_A | (Y_C, Y_D) \in S_{C,D}] \times P[Y_B \in S_B | (Y_C, Y_D) \in S_{C,D}] = \frac{\alpha\beta}{P[(Y_C, Y_D) \in S_{C,D}]}.$$

Now  $P[Y_A \in S_A, Y_B \in S_B, (Y_C, Y_D) \in S_{C,D}] \neq \alpha\beta$  and  $P[(Y_C, Y_D) \in S_{C,D}] > 0$  imply

$$P[Y_A \in S_A, Y_B \in S_B | (Y_C, Y_D) \in S_{C,D}] = \frac{P[Y_A \in S_A, Y_B \in S_B, (Y_C, Y_D) \in S_{C,D}]}{P[(Y_C, Y_D) \in S_{C,D}]}$$

$$\neq \frac{\alpha\beta}{P[(Y_C, Y_D) \in S_{C,D}]}.$$

It follows that

$$P[Y_A \in S_A | (Y_C, Y_D) \in S_{C,D}] \times P[Y_B \in S_B | (Y_C, Y_D) \in S_{C,D}]$$

$$\neq P[Y_A \in S_A, Y_B \in S_B | (Y_C, Y_D) \in S_{C,D}],$$

which implies that  $Y_A \not\perp Y_B | (Y_C, Y_D)$ . ■

**Proof of Example 4.5** We prove the result for  $a > 0$ . The proof for  $a < 0$  is analogous. We pick values  $y_1$  and  $y_2$  such that the conditions of Theorem 4.4 hold with  $\alpha\beta > 0$  for sets  $S_1 = [0, y_1)$  and  $S_2 = [0, y_2)$  and for a suitably chosen set  $S_3$  (possibly depending on

$y_1$  and  $y_2$ ). Write  $Y_3 = Y_1 + aY_2$ . We first ensure that  $P[Y_1 \in S_1, Y_2 \in S_2, Y_3 \in S_3] = 0$ . We have

$$\begin{aligned} P[Y_1 \in S_1, Y_2 \in S_2, Y_3 \in S_3] &= P[Y_1 < y_1, Y_2 < y_2, Y_3 \in S_3] \\ &= \int P[Y_1 < y_1, Y_2 < y_2 \mid Y_3 = y_3] 1\{y_3 \in S_3\} dF_3(y_3), \end{aligned}$$

where  $dF_3$  is the density of  $Y_3$ . For any  $y_3$  we have

$$\begin{aligned} P[Y_1 < y_1, Y_2 < y_2 \mid Y_3 = y_3] &= P[Y_1 < y_1, Y_2 < y_2 \mid Y_1 + aY_2 = y_3] \\ &= P[Y_1 < y_1, Y_2 < y_2 \mid Y_2 = \frac{y_3 - Y_1}{a}] \\ &= P[Y_1 < y_1, \frac{y_3 - Y_1}{a} < y_2] \\ &= P[y_3 - ay_2 < Y_1 < y_1]. \end{aligned}$$

Condition (ii) holds as required if for all  $y_3 \in S_3$ , we have  $y_3 \geq y_1 + ay_2$ , as this ensures  $P[y_3 - ay_2 < Y_1 < y_1] = 0$  and therefore  $P[Y_1 \in S_1, Y_2 \in S_2, Y_3 \in S_3] = 0$ . Similarly, we have

$$\begin{aligned} P[Y_1 < y_1, Y_3 \in S_3] &= \int P[Y_1 < y_1 \mid Y_3 = y_3] 1\{y_3 \in S_3\} dF_3(y_3) \\ &= \int P[Y_2 > \frac{y_3 - y_1}{a}] 1\{y_3 \in S_3\} dF_3(y_3) \\ &= \int (1 - \frac{y_3 - y_1}{a}) 1\{y_3 \in S_3\} dF_3(y_3), \end{aligned}$$

and

$$\begin{aligned} P[Y_2 < y_2, Y_3 \in S_3] &= \int P[Y_2 < y_2 \mid Y_3 = y_3] 1\{y_3 \in S_3\} dF_3(y_3) \\ &= \int P[Y_1 > y_3 - ay_2] 1\{y_3 \in S_3\} dF_3(y_3) \\ &= \int (1 - (y_3 - ay_2)) 1\{y_3 \in S_3\} dF_3(y_3). \end{aligned}$$

We now choose  $y_1$  and  $y_2$  such that in addition to  $y_3 \geq y_1 + ay_2$ , we also have that for some  $0 < \delta < 1$ ,  $1 - (y_3 - y_1)/a \geq \delta$  and  $1 - (y_3 - ay_2) \geq \delta$ . Choose any admissible  $\delta$ . For the former,  $y_3 \leq y_1 + a(1 - \delta)$  suffices, and for the latter,  $y_3 \leq 1 - \delta + ay_2$  suffices. For

convenience and without loss of generality, put  $y_1 = y_2$ . For all  $0 \leq y_1 < 1 - \delta$ , we have that

$$y_1 + ay_1 < \min[1 - \delta + ay_1, y_1 + a(1 - \delta)],$$

so given any  $0 \leq y_1 < 1 - \delta$ , any choice of  $y_3 \in S_{3,1} := [y_1 + ay_1, \min[1 - \delta + ay_1, y_1 + a(1 - \delta)]]$  will satisfy  $y_3 \geq y_1 + ay_1, y_3 \leq y_1 + a(1 - \delta)$  (thus  $1 - (y_3 - y_1)/a \geq \delta$ ) and  $y_3 \leq 1 - \delta + ay_1$  (thus  $1 - (y_3 - ay_1) \geq \delta$ ).

We now choose  $y_1$  to ensure that the conditions of (i) hold. First, we ensure that  $P[y_3 \in S_{3,1}] > 0$ . Now the continuity of  $Y_1$  and  $Y_2$  and the requirement that  $P[Y_1 \neq Y_2] = 1$  ensure that  $Y_3$  has a continuous distribution. Thus,

$$P[y_3 \in S_{3,1}] = F_3(\min[1 - \delta + ay_1, y_1 + a(1 - \delta)]) - F_3((1 + a)y_1) > 0,$$

provided  $y_1 + ay_1 < \min[1 - \delta + ay_1, y_1 + a(1 - \delta)]$ , which holds for all  $0 \leq y_1 < 1 - \delta$ , as shown above. Next, for any  $0 < y_1 < 1 - \delta$ ,

$$\begin{aligned} P[Y_1 < y_1, Y_3 \in S_{3,1}] &= \int (1 - \frac{y_3 - y_1}{a}) 1_{\{y_3 \in S_{3,1}\}} dF_3(y_3) \\ &\geq \delta \int 1_{\{y_3 \in S_{3,1}\}} dF_3(y_3) \\ &= \delta P[y_3 \in S_{3,1}] > 0, \end{aligned}$$

and

$$\begin{aligned} P[Y_2 < y_2, Y_3 \in S_{3,1}] &= \int (1 - (y_3 - ay_2)) 1_{\{y_3 \in S_{3,1}\}} dF_3(y_3) \\ &\geq \delta \int 1_{\{y_3 \in S_{3,1}\}} dF_3(y_3) \\ &= \delta P[y_3 \in S_{3,1}] > 0, \end{aligned}$$

recalling that  $y_1 = y_2$ . Thus, any choice  $0 < y_1 < 1 - \delta$ ,  $y_1 = y_2$  verifies the conditions of Theorem 4.4 under our given conditions. ■

## References

Avin, C., I. Shpitser, and J. Pearl (2005), “Identifiability of Path-Specific Effects,” In *Proceedings of International Joint Conference on Artificial Intelligence*, Edinburgh, Scotland, 357-363.

Bang-Jensen, J. and G. Gutin (2001). *Digraphs: Theory, Algorithms and Applications*. London: Springer-Verlag.

Cartwright, N. (2000). *Measuring Causes: Invariance, Modularity and the Causal Markov Condition*. Monograph, London: Centre for Philosophy of Natural and Social Science.

Chalak, K., and H. White (2007a), "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," UCSD Department of Economics Discussion Paper.

Chalak, K., and H. White (2007b), "Identification with Conditioning Instruments in Causal Systems," UCSD Department of Economics Discussion Paper.

Dawid, A.P. (1979), "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society, Series B*, 41, 1-31 (with discussion).

Dawid, A.P. (1980), "Conditional Independence for Statistical Operation," *The Annals of Statistics*, 8, 598-617.

Dawid, A.P. (2000), "Causal Inference without Counterfactuals," *Journal of the American Statistical Association*, 95, 407-448 (with discussion).

Dawid, A.P. (2002), "Influence Diagrams for Causal Modeling and Inference," *International Statistical Review*, 70, 161-189.

Didelez, V., A. P. Dawid, and S. Geneletti (2006), "Direct and Indirect Effects of Sequential Treatments," In R. Dechter, T.S. Richardson (eds.), *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, pp. 138-146.

Dudley, R.M. (2002). *Real Analysis and Probability*. New York: Cambridge University Press.

Florens, J.-P., M. Mouchart, and J.-M. Rolin (1990). *Elements of Bayesian Statistics*. New York: Marcel Dekker.

Geiger, D., T. S. Verma, and J. Pearl (1990), "Identifying Independence in Bayesian Networks," *Networks*, 20, 507-534.

Geiger, D. and J. Pearl (1993), "Logical and Algorithmic Properties of Conditional Independence and Graphical Models," *The Annals of Statistics*, 21, 2001-2021.

Golubitsky, M. and I. Stewart (2006), "Nonlinear Dynamics of Networks: the Groupoid Formalism," *Bulletin of the American Mathematical Society*, 43, 305-364.

Granger, C. W. J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424-438.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hausman, D. and J. Woodward (1999), "Independence, Invariance and the Causal Markov Condition," *British Journal for the Philosophy of Science*, 50, 521-583.

Heckman, J. (2005), "The Scientific Model of Causality," *Sociological Methodology*, 35, 1-97.

Holland, P.W. (1986), "Statistics and Causal inference," *Journal of the American Statistical Association*, 81, 945-970 (with discussion).

Lauritzen, S. L., and D. J. Spiegelhalter (1988), "Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems," *Journal of the Royal Statistical Society, Series B*, 50, 157-224 (with discussion).

Lauritzen, S. L., A. P. Dawid, B. N. Larsen, and H.-G. Leimer (1990), "Independence Properties of Directed Markov Fields," *Networks*, 20, 491-505.

Lauritzen, S. L., and T. S. Richardson (2002), "Chain Graph Models and their Causal Interpretation," *Journal of the Royal Statistical Society, Series B*, 64, 321-361 (with discussion).

Neveu, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Translated by Amiel Feinstein. San Francisco: Holden-Day.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufman.

Pearl, J. (1993), "Aspects of Graphical Methods Connected with Causality," in *Proceedings of the 49th Session of the International Statistical Institute*, pp. 391-401.

Pearl, J. (1995), "Causal Diagrams for Empirical Research," *Biometrika*, 82, 669-710 (with Discussion).

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Pearl, J. (2001), "Direct and Indirect Effects," In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, 411-420.

- Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.
- Robins, J. (2003), "Semantics of Causal Models and the Identification of Direct and Indirect Effects," In P. Green, N.L. Hjort, S. Richardson (eds.), *Highly Structured Stochastic Systems*, NY: Oxford University Press, pp. 70-81.
- Robins, J. and S. Greenland (1992), "Identifiability and Exchangeability for Direct and Indirect Effects," *Epidemiology*, 3, 143-155.
- Rosenbaum, P. R. (2002). *Observational Studies*. 2nd ed., Berlin: Springer-Verlag.
- Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D. (2004), "Direct and Indirect Causal Effects via Potential Outcomes," *Scandinavian Journal of Statistics*, 31, 161-170.
- Sims, C. (1972), "Money, Income, and Causality," *American Economic Review*, 62, 540-52.
- Smith, J. Q. (1989), "Influence Diagrams for Statistical Modeling," *The Annals of Statistics*, 17, 654-672.
- Spirtes, P., C. Glymour, and R. Scheines (1993). *Causation, Prediction and Search*. Berlin: Springer-Verlag.
- Studeny, M. (1993), "Formal Properties of Conditional Independence in Different Calculi of AI." In M. Clarke, R. Kruse, S. Moral (eds.), *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Berlin: Springer-Verlag, pp. 341-348.
- Spohn, W. (1980), "Stochastic Independence, Causal Independence, and Shieldability," *Journal of Philosophical Logic*, 9, 73-99.
- Verma, T. and J. Pearl (1988), "Causal Networks: Semantics and Expressiveness," in *Proceedings, 4th Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, MN, Mountain View, CA, pp. 352-359.
- White, H. and K. Chalak (2008), "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning," UCSD Department of Economics Discussion Paper.