4.4 The Identification of Plans

This section, based on Pearl and Robins (1995), concerns the probabilistic evaluation of plans in the presence of unmeasured variables, where each plan consists of several concurrent or sequential actions and each action may be influenced by its predecessors in the plan. We establish a graphical criterion for recognizing when the effects of a given plan can be predicted from passive observations on measured variables only. When the criterion is satisfied, a closed-form expression is provided for the probability that the plan will achieve a specified goal.

4.4.1 Motivation

To motivate the discussion, consider an example discussed in Robins (1993, apx. 2), as depicted in Figure 4.4. The variables X_1 and X_2 stand

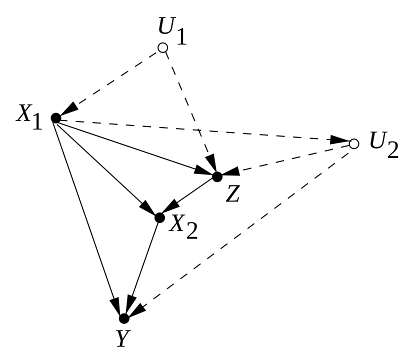

Figure 4.4: The problem of evaluating the effect of the plan $(do(x_1),\ do(x_2))$ on Y, from nonexperimental data taken on $X_1,\ Z,\ X_2$, and Y.

for treatments that physicians prescribe to a patient at two different times, Z represents observations that the second physician consults to determine X_2, and Y represents the patient's survival. The hidden variables U_1 and U_2 represent, respectively, part of the patient's history and the patient's disposition to recover. A simple realization of such structure could be found among AIDS patients, where Z represents episodes of PCP. This is a common opportunistic infection of AIDS patients that (as the diagram shows) does not have a direct effect on survival Y because it can be treated effectively, but it is an indicator of the patient's underlying immune status (U_2), which can cause death. The terms X_1 and X_2 stand for bactrim, a drug that prevents PCP (Z) and may also prevent death by other mechanisms. Doctors used the patient's earlier PCP history (U_1) to prescribe X_1, but its value was not recorded for data analysis.

The problem we face is as follows. Assume we have collected a large amount of data on the behavior of many patients and physicians, which is summarized in the form of (an estimated) joint distribution P of the observed four variables $(X_1,\ Z,\ X_2,\ Y)$. A new patient comes in, and we wish to determine the impact of the (unconditional) plan $(do(x_1),\ do(x_2))$ on survival, where x_1 and x_2 are two predetermined dosages of bactrim to be administered at two prespecified times.

In general, our problem amounts to that of evaluating a new plan by watching the performance of other planners whose decision strategies are indiscernible. Physicians do not provide a description of all inputs that prompted them to prescribe a given treatment; all they communicate to us is that U_1 was consulted in determining X_1 and that Z and X_1 were consulted in determining X_2. But U_1, unfortunately, was not recorded. In epidemiology, the plan evaluation problem is known as "time-varying treatment with time-varying confounders" (Robins 1993). In artificial intelligence applications, the evaluation of such plans enables one agent to learn to act by observing the performance of another agent, even in cases where the actions of the other agent are predicated on factors that are not visible to the learner. If the learner is permitted to act as well as observe, then the task becomes much easier: the topology of the causal diagram could also be inferred (at least partially), and the effects of some previously unidentifiable actions could be determined.

As in the identification of actions (Section 4.3), the main problem in plan identification is the control of "confounders," that is, unobserved factors that trigger actions and simultaneously affect the response. However, unlike the problem treated in Section 4.3, plan identification is further complicated by the fact that some of the confounders (e.g. Z) are affected by control variables. As we remarked in Chapter

3, one of the deadliest sins in the design of statistical experiments (Cox 1958, p. 48) is to adjust for such variables, because the adjustment would simulate holding a variable constant; holding constant a variable that stands between an action and its consequence interferes with the very quantity we wish to estimate—the total effect of that action.

Two other features of Figure 4.4 are worth noting. First, the quantity $P(y|\hat{x}_1, \hat{x}_2)$ cannot be computed if we treat the control variables X_1 and X_2 as a single compound variable X. The graph corresponding to such compounding would depict X as connected to Y by both an arrow and a curved arc (through U) and thus would form a bow pattern (see Figure 3.9), which is indicative of nonidentifiability. Second, the causal effect $P(y|\hat{x}_1)$ in isolation is not identifiable because U_1 creates a bow pattern around the link $X \rightarrow Z$, which lies on a directed path from X to Y (see the discussion in Section 3.5).

The feature that facilitates the identifiability of $P(y|\hat{x}_1, \hat{x}_2)$ is the identifiability of $P(y|x_1, z, \hat{x}_2)$—the causal effect of the action $do(X_2 = x_2)$ alone, conditional on the observations available at the time of this action. This can be verified using the back-door criterion, observing that $\{X_1, Z\}$ blocks all back-door paths between X_2 and Y. Thus, the identifiability of $P(y|\hat{x}_1, \hat{x}_2)$ can be readily proven by writing

$$\begin{aligned} P(y|\hat{x}_1, \hat{x}_2) &= P(y|x_1, \hat{x}_2) && (4.1)\\ &= \sum_z P(y|z, x_1, \hat{x}_2) P(z|x_1) && (4.2)\\ &= \sum_z P(y|z, x_1, x_2) P(z|x_1), && (4.3) \end{aligned}$$

where (4.1) and (4.3) follow from Rule 2, and (4.2) follows from Rule 3. The subgraphs that permit the application of these rules are shown in Figure 4.5 (in Section 4.4.3).

This derivation also highlights how conditional plans can be evaluated. Assume we wish to evaluate the effect of the plan $\{do(X_1 = x_1),\ do(X_2 = g(x_1, z))\}$. Following the analysis of Section 4.2, we write

$$\begin{aligned} P(y|do(X_1 = x_1), do(X_2 = g(x_1, z))) &= P(y|x_1, do(X_2 = g(x_1, z)))\\ &= \sum_z P(y|z, x_1, do(X_2 = g(x_1, z))) P(z|x_1)\\ &= \sum_z P(y|z, x_1, x_2) P(z|x_1)|_{x_2 = g(x_1, z)}. && (4.4) \end{aligned}$$

Again, the identifiability of this conditional plan rests on the identifiability of the expression $P(y|z, x_1, \hat{x}_2)$, which reduces to $P(y|z, x_1, x_2)$ because $\{X_1, Z\}$ blocks all back-door paths between X_2 and Y.

The criterion developed in the next section will enable us to recognize in general, by graphical means, whether a proposed plan can be evaluated from the joint distribution on the observables and, if so, to identify which covariates should be measured and how they should be adjusted.

4.4.2 Plan Identification: Notation and Assumptions

Our starting point is a knowledge specification scheme in the form of a causal diagram, like the one shown in Figure 4.4, that provides a qualitative summary of the analyst's understanding of the relevant data-generating processes.[5]

Notation:

A *control problem* consists of a directed acyclic graph (DAG) G with vertex set V, partitioned into four disjoint sets $V = \{X,\ Z,\ U,\ Y\}$, where

$X =$ the set of control variables (exposures, interventions, treatments, etc.);

$Z =$ the set of observed variables, often called *covariates*;

$U =$ the set of unobserved (latent) variables; and

$Y =$ an outcome variable.

We let the control variables be ordered $X = X_1, X_2, \ldots, X_n$ such that every X_k is a nondescendant of X_{k+j} $(j > 0)$ in G, and we let the outcome Y be a descendant of X_n. Let N_k stand for the set of observed nodes that are nondescendants of any element in the set $\{X_k, X_{k+1}, \ldots, X_n\}$. A *plan* is an ordered sequence $(\hat{x}_1,\ \hat{x}_2, \ldots, \hat{x}_n)$ of value assignments to the control variables, where $\hat{x}_k$

[5]An alternative specification scheme using counterfactual statements was developed by Robins (1986, 1987), as described in Section 3.6.4.

means "X_k is set to x_k." A *conditional plan* is an ordered sequence $(\hat{g}_1(z_1), \hat{g}_2(z_2), \ldots, \hat{g}_n(z_n))$ where each g_k is a function from a set Z_k to X_k and where $\hat{g}_k(z_k)$ stands for the statement "set X_k to $g_k(z_k)$ whenever Z_k attains the value z_k." The support Z_k of each $g_k(z_k)$ function must not contain any variables that are descendants of X_k in G.

Our problem is to *evaluate* an unconditional plan[6] by computing $P(y|\hat{x}_1, \ \hat{x}_2, \ldots, \hat{x}_n)$, which represents the impact of the plan $(\hat{x}_1, \ldots, \hat{x}_n)$ on the outcome variable Y. The expression $P(y|\hat{x}_1, \ \hat{x}_2, \ldots, \hat{x}_n)$ is said to be *identifiable* in G if, for every assignment $(\hat{x}_1, \ \hat{x}_2, \ldots, \hat{x}_n)$, the expression can be determined uniquely from the joint distribution of the observables $\{X, Y, Z\}$. A control problem is identifiable whenever $P(y|\hat{x}_1, \ \hat{x}_2, \ldots, \hat{x}_n)$ is identifiable.

Our main identifiability criteria are presented in Theorems 4.4.1 and 4.4.6. These invoke d-separation tests on various subgraphs of G, defined in the same manner as in Section 4.3. We denote by $G_{\overline{X}}$ (and $G_{\underline{X}}$, respectively) the graphs obtained by deleting from G all arrows pointing to (emerging from) nodes in X. To represent the deletion of both incoming and outgoing arrows, we use the notation $G_{\overline{X}\underline{Z}}$. Finally, the expression $P(y|\hat{x}, z) \triangleq P(y, z|\hat{x})/P(z|\hat{x})$ stands for the probability of $Y = y$ given that $Z = z$ is observed and X is held constant at x.

4.4.3 Plan Identification: A General Criterion

Theorem 4.4.1 (Pearl and Robins 1995)
The probability $P(y|\hat{x}_1, \ldots, \hat{x}_n)$ is identifiable if, for every $1 \leq k \leq n$, there exists a set Z_k of covariates satisfying

$$Z_k \subseteq N_k, \tag{4.5}$$

(i.e., Z_k consists of nondescendants of $\{X_k, X_{k+1}, \ldots, X_n\}$) and

$$(Y \perp\!\!\!\perp X_k | X_1, \ldots, X_{k-1}, Z_1, Z_2, \ldots, Z_k)_{G_{\underline{X}_k, \overline{X}_{k+1}, \ldots, \overline{X}_n}}. \tag{4.6}$$

[6] Identification of conditional plans can be obtained from Theorem 4.4.1 using the method described in Section 4.2 and exemplified in Section 4.4.1.

When these conditions are satisfied, the effect of the plan is given by

$$P(y|\hat{x}_1, \ldots, \hat{x}_n) = \sum_{z_1,\ldots,z_n} P(y|z_1, \ldots, z_n, x_1, \ldots, x_n) \prod_{k=1}^{n} P(z_k|z_1, \ldots, z_{k-1}, x_1, \ldots, x_{k-1}). \quad (4.7)$$

Before presenting its proof, let us demonstrate how Theorem 4.4.1 can be used to test the identifiability of the control problem shown in Figure 4.4. First, we will show that $P(y|\hat{x}_1, \hat{x}_2)$ cannot be identified without measuring Z; in other words, that the sequence $Z_1 = \emptyset$, $Z_2 = \emptyset$ would not satisfy conditions (4.5)–(4.6). The two d-separation tests encoded in (4.6) are

$$(Y \perp\!\!\!\perp X_1)_{G_{\underline{X}_1,\overline{X}_2}} \quad \text{and} \quad (Y \perp\!\!\!\perp X_2|X_1)_{G_{\underline{X}_2}}.$$

The two subgraphs associated with these tests are shown in Figure 4.5. We see that $(Y \perp\!\!\!\perp X_1)$ holds in $G_{\underline{X}_1,\overline{X}_2}$ but that $(Y \perp\!\!\!\perp X_2|X_1)$ fails to

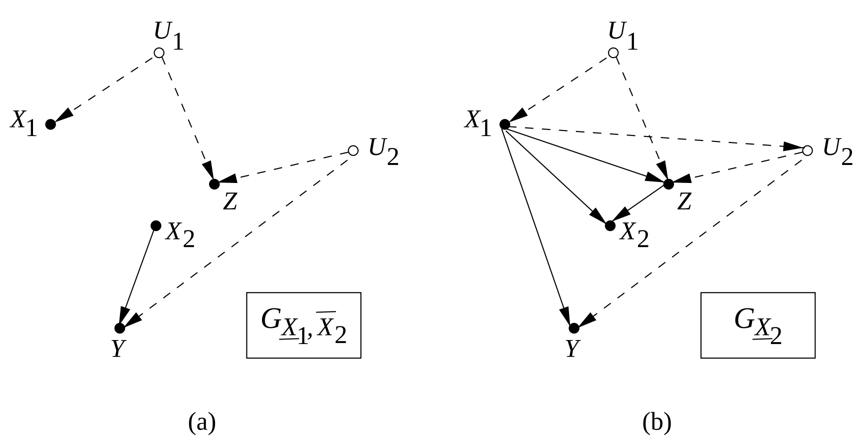

Figure 4.5: The two subgraphs of G used in testing the identifiability of the plan $(\hat{x}_1, \hat{x}_2)$ in Figure 4.4.

hold in $G_{\underline{X}_2}$. Thus, in order to pass the test, we must have either $Z_1 = \{Z\}$ or $Z_2 = \{Z\}$; since Z is a descendant of X_1, only the second alternative satisfies (4.5). The tests applicable to the sequence $Z_1 = \emptyset$, $Z_2 = \{Z\}$ are $(Y \perp\!\!\!\perp X_1)_{G_{\underline{X}_1, \overline{X}_2}}$ and $(Y \perp\!\!\!\perp X_2 | X_1, Z)_{G_{\underline{X}_2}}$. Figure 4.5 shows that both tests are now satisfied, because $\{X_1, Z\}$ d-separates Y from X_2 in $G_{\underline{X}_2}$. Having satisfied conditions (4.5)–(4.6), equation (4.7) provides a formula for the effect of plan $(\hat{x}_1, \hat{x}_2)$ on Y:

$$P(y|\hat{x}_1, \hat{x}_2) = \sum_z P(y|z, x_1, x_2) P(z|x_1), \tag{4.8}$$

which coincides with (4.3).

The question naturally arises of whether the sequence $Z_1 = \emptyset$, $Z_2 = \{Z\}$ can be identified without exhaustive search. This question will be answered in Corollary 4.4.5 and Theorem 4.4.6.

Proof of Theorem 4.4.1: The proof given here is based on the inference rules of *do* calculus (Theorem 3.4.1), which facilitate the reduction of causal effect formulas to hat-free expressions. An alternative proof, using latent variable elimination, is given in Pearl and Robins 1995).

Step 1: The condition $Z_k \subseteq N_k$ implies $Z_k \subseteq N_j$ for all $j \geq k$. Therefore, we have

$$\begin{aligned} & P(z_k | z_1, \ldots, z_{k-1}, x_1, \ldots, x_{k-1}, \hat{x}_k, \hat{x}_{k+1}, \ldots, \hat{x}_n) \\ & \quad = P(z_k | z_1, \ldots, z_{k-1}, x_1, \ldots, x_{k-1}). \end{aligned}$$

This is so because no node in $\{Z_1, \ldots, Z_k, X_1, \ldots, X_{k-1}\}$ can be a descendant of any node in $\{X_k, \ldots, X_n\}$. Hence, Rule 3 allows us to delete the hat variables from the expression.

Step 2: The condition in (4.5) permits us to invoke Rule 2 and write:

$$\begin{aligned} & P(y | z_1, \ldots, z_k, x_1, \ldots, x_{k-1}, \hat{x}_k, \hat{x}_{k+1}, \ldots, \hat{x}_n) \\ & \quad = P(y | z_1, \ldots, z_k, x_1, \ldots, x_{k-1}, x_k, \hat{x}_{k+1}, \ldots, \hat{x}_n). \end{aligned}$$

Thus, we have

$$\begin{aligned} & P(y|\hat{x}_1, \ldots, \hat{x}_n) \\ & \quad = \sum_{z_1} P(y|z_1, \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) P(z_1|\hat{x}_1, \ldots, \hat{x}_n) \\ & \quad = \sum_{z_1} P(y|z_1, x_1, \hat{x}_2, \ldots, \hat{x}_n) P(z_1) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{z_2}\sum_{z_1} P(y|z_1, z_2, x_1, \hat{x}_2, \ldots, \hat{x}_n) P(z_1) \\
&\quad P(z_2|z_1, x_1, \hat{x}_2, \ldots, \hat{x}_n) \\
&= \sum_{z_2}\sum_{z_1} P(y|z_1, z_2, x_1, x_2, \hat{x}_3, \ldots, \hat{x}_n) \\
&\quad P(z_1) P(z_2|z_1, x_1) \\
&\quad \vdots \\
&= \sum_{z_n} \cdots \sum_{z_2}\sum_{z_1} P(y|z_1, \ldots, z_n, x_1, \ldots, x_n) \\
&\quad \times P(z_1) P(z_2|z_1, x_1) \ldots P(z_n|z_1, x_1, z_2, x_2, \ldots, z_{n-1}, x_{n-1}) \\
&= \sum_{z_1,\ldots,z_n} P(y|z_1, \ldots, z_n, x_1, \ldots, x_n) \prod_{k=1}^{n} P(z_k|z_1, \ldots, z_{k-1}, x_1, \ldots, x_{k-1}).
\end{aligned}
$$

□

Definition 4.4.2 *Any sequence $Z_1, \ldots, Z_n$ of covariates satisfying the conditions in* (4.5)–(4.6) *will be called* admissible*, and any expression $P(y|\hat{x}_1,\ \hat{x}_2, \ldots, \hat{x}_n)$ that is identifiable by the criterion of Theorem 4.4.1 will be called G*-identifiable.[7]

The following corollary is immediate.

Corollary 4.4.3 *A control problem is G-identifiable if and only if it has an admissible sequence.*

G-identifiability is sufficient but not necessary for general plan identifiability as defined in Section 4.4.2. The reasons are twofold. First, the completeness of the three inference rules of *do* calculus is still a pending conjecture. Second, the kth step in the reduction of (4.7) refrains from conditioning on variables Z_k that are descendants of X_k—namely, variables that may be affected by the action $do(X_k = x_k)$. In certain causal structures, the identifiability of causal effects requires that we condition on such variables, as demonstrated by the front-door criterion (Theorem 3.3.4).

[7]The term "G-admissibility" was used in Pearl and Robins (1995) to evoke two associations: (1) Robins's G-*estimation* formula (equation (3.65)), which coincides with (4.7) when G is complete and contains no unobserved confounders; and (2) the *graphical* nature of the conditions in (4.5)–(4.6).

4.4.4 Plan Identification: A Procedure

Theorem 4.4.1 provides a declarative condition for plan identifiability. It can be used to ratify that a proposed formula is valid for a given plan, but it does not provide an effective procedure for deriving such formulas because the choice of each Z_k is not spelled out procedurally. The possibility exists that some unfortunate choice of Z_k satisfying (4.5) and (4.6) might prevent us from continuing the reduction process even through another reduction sequence is feasible.

This is illustrated in Figure 4.6. Here W is an admissible choice for Z_1, but if we make this choice then we will not be able to complete the reduction, since no set Z_2 can be found that satisfies condition (4.6): $(Y \perp\!\!\!\perp X_2 | X_1, W, Z_2)_{G_{\underline{X}_2}}$. In this example it would be wiser to choose $Z_1 = Z_2 = \emptyset$, which satisfies both $(Y \perp\!\!\!\perp X_1 |\ \emptyset)_{G_{\underline{X}_1, \overline{X}_2}}$ and $(Y \perp\!\!\!\perp X_2 | X_1, \emptyset)_{G_{\underline{X}_2}}$.

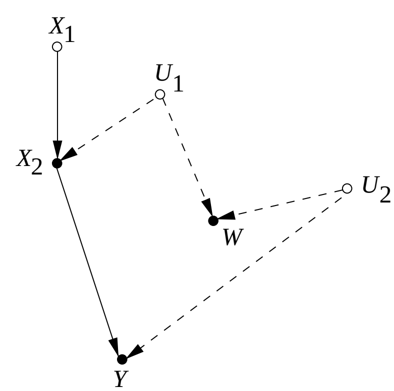

Figure 4.6: An admissible choice $Z_1 = W$ that rules out any admissible choice for Z_2.

The obvious way to avoid bad choices of covariates, like the one illustrated in Figure 4.6, is to insist on always choosing a "minimal" Z_k, namely, a set of covariates satisfying (4.6) that has no proper subset satisfying (4.6). However, since there are usually many such minimal sets (see Figure 4.7), the question remains of whether every choice of a minimal Z_k is "safe:" Can we be sure that no choice of a minimal subsequence $Z_1, \ldots, Z_k$ will ever prevent us from finding an admissible Z_{k+1}, when some admissible sequence $Z_1^*, \ldots, Z_n^*$ exists?

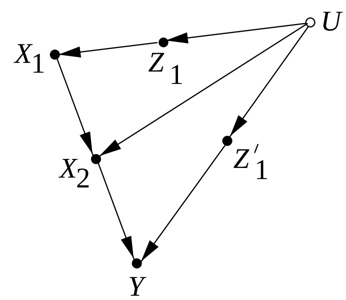

Figure 4.7: Illustrating non-uniqueness of minimal admissible sets: Z_1 and Z_1' are each minimal and admissible.

The next result guarantees the safety of every minimal subsequence $Z_1, \ldots, Z_k$ and hence provides an effective test for G-identifiability.

Theorem 4.4.4 *If there exists an admissible sequence $Z_1^*, \ldots, Z_n^*$ then, for every minimally admissible subsequence $Z_1, \ldots, Z_{k-1}$ of covariates, there is an admissible set Z_k.*

A proof is given in [Pearl and Robins, 1995]Pearl and Robins (1995).

Theorem 4.4.4 now yields an effective decision procedure for testing G-identifiability as follows.

Corollary 4.4.5 *A control problem is G-identifiable if and only if the following algorithm exits with success:*

1. *Set $k = 1$.*
2. *Choose any minimal $Z_k \subseteq N_k$ satisfying* (4.6).
3. *If no such Z_k exists then exit with failure; else set $k = k + 1$.*
4. *If $k = n + 1$ then exit with success, else return to step 2.*

A further variant of Theorem 4.4.4 can be stated that avoids the search for minimal sets Z_k. This follows from the realization that, if an admissible sequence exists, we can rewrite Theorem 4.4.1 in terms of an explicit sequence of covariates $W_1, W_2, \ldots, W_n$ that can easily be identified in G.

Theorem 4.4.6 *The probability $P(y|\hat{x}_1, \ldots, \hat{x}_n)$ is G-identifiable if and only if the following condition holds for every $1 \leq k \leq n$:1*

$$(Y \perp\!\!\!\perp X_k | X_1, \ldots, X_{k-1}, W_1, W_2, \ldots, W_k)_{G_{\underline{X}_k, \overline{X}_{k+1}, \ldots, \overline{X}_n}},$$

where W_k is the set of all covariates in G that are both nondescendants of $\{X_k, X_{k+1}, \ldots, X_n\}$ and have either Y or X_k as descendant in $G_{\underline{X}_k, \overline{X}_{k+1}, \ldots, \overline{X}_n}$. Moreover, if this condition is satisfied then the plan evaluates as

$$P(y|\hat{x}_1, \ldots, \hat{x}_n) = \sum_{w_1, \ldots, w_n} P(y|w_1, \ldots, w_n, x_1, \ldots, x_n) \prod_{k=1}^{n} P(w_k|w_1, \ldots, w_{k-1}, x_1, \ldots, x_{k-1}). \quad (4.9)$$

A proof of Theorem 4.4.6, together with several generalizations can be found in Pearl and Robins (1995). Extensions to G-identifiability are reported in Kuroki and Miyakawa (1999).

The reader should note that, although Corollary 4.4.5 and Theorem 4.4.6 are procedural in the sense of offering systematic tests for plan identifiability, they are still *order-dependent*. It is quite possible that an admissible sequence exists in one ordering of the control variables and not in another when both orderings are consistent with the arrows in G. The graph G in Figure 4.8 illustrates such a case. It is obtained from Figure 4.4 by deleting the arrows $X_1 \rightarrow X_2$ and $X_1 \rightarrow Z$, so that the two control variables (X_1 and X_2) can be ordered arbitrarily. The ordering (X_1, X_2) would still admit the admissible sequence $(\emptyset, Z)$ as before, but no admissible sequence can be found for the ordering (X_2, X_1). This can be seen immediately from the graph $G_{\underline{X}_1}$, in which (according to (4.6) with $k = 1$) we need to find a set Z such that $\{X_2, Z\}$ d-separates Y from X_1. No such set exists.

The implication of this order sensitivity is that, whenever G permits several orderings of the control variables, all orderings need be examined before we can be sure that a plan is not G-identifiable. Whether an effective search exists through the space of such orderings remains an open question.