Chapter 3: Theoretical Framework

In this section, we present the causal reasoning framework alongside a novel example that show-cases the ability of the framework to reason about the direct and indirect causal influence of *events* over *literals*, originally presented in [48]. We will then present and discuss two counterexamples to our original definition of indirect cause. Following that, we will present an improved definition and show that it is able to solve the counterexamples. Finally, we will discuss some technical and conceptual advantages of the improved definition of indirect cause over the original definition.

Given a consistent set of literals θ representing an outcome of interest, an action description AD describing the effects of events in a dynamic domain, and a path ρ in $\tau(AD)$ corresponding to a scenario of interest, we construct a *problem* $\psi = \langle \theta, \rho, AD \rangle$. The framework can be used to reason over the elements of a problem ψ to identify events that are responsible for causing the literals of θ to hold simultaneously in a state of ρ . We now present the details of the running example.

3.1 Running Example

Let $\psi_E = \langle \theta_E, \rho_E, AD_E \rangle$ be a problem representing the running example. The outcome of interest in our example is $\theta_E = \{A, B, C, D, E, F\}$. The following action description AD_E characterizes elementary events in the example domain:

$$e_1$$
 causes A if $\neg A$ (3.1)

$$e_2$$
 causes A if $\neg A$ (3.2)

$$e_3$$
 causes C if $\neg C$ (3.3)

$$e_4$$
 causes E if $\neg E$ (3.4)

$$e_5$$
 causes F if $\neg F$ (3.5)

$$e_5$$
 causes C if $\neg C$ (3.6)

$$B ext{ if } C ag{3.7}$$

$$D ext{ if } E, F ag{3.8}$$

The dynamic law (3.1) states that e_1 will cause A to hold if it does not already hold when the event occurs. Similarly, laws (3.2) through (3.6) describe the direct effects of events e_2 , e_3 , e_4 , and e_5 . The state constraint (3.7) tells us that B holds whenever C holds, and the state constraint (3.8) tells us that D holds whenever both E and E hold. Note that although there are no executability conditions in this action description, it is straightforward to use them to model the events in AD_E in greater detail.

Table 3.1: Tabular representation of path $\rho_E \in \tau(AD_E)$.

State	Event
$\sigma_1 = \{\neg A, \neg B, \neg C, \neg D, \neg E, \neg F\}$	$\epsilon_1 = \{e_1, e_2\}$
$\sigma_2 = \{A, \neg B, \neg C, \neg D, \neg E, \neg F\}$	$\epsilon_2 = \{e_3\}$
$\sigma_3 = \{A, B, C, \neg D, \neg E, \neg F\}$	$\epsilon_3 = \{e_4, e_5\}$
$\sigma_4 = \{A, B, C, D, E, F\}$	_

The dynamics of the scenario are given by path $\rho \in \tau(AD_E)$. Path ρ consists of three compound events, $\epsilon_1 = \{e_1, e_2\}$, $\epsilon_2 = \{e_3\}$, and $\epsilon_3 = \{e_4, e_5\}$. Table 3.1 shows the evolution of state in ρ_E in response to these events. The first column lists each state of ρ_E , and the second column gives the event α_i that caused a transition to the subsequent state. The outcome θ_E is not satisfied in the first three states of the path, however, the events of ρ_E have somehow caused the outcome to be satisfied in state σ_4 .

By examining the laws of AD_E together with the states and transitions of ρ_E , the reader can reason about which event(s) directly and (or) indirectly caused every literal in θ_E to hold by state σ_4 . In the first transition, for instance, we see that events e_1 and e_2 overdetermining direct causes of A holding in state σ_2 according to laws (3.1) and (3.2), respectively. Next, event e_3 directly causes C to hold in state σ_3 according to law (3.3). In the same transition, e_3 indirectly causes B because of laws (3.3) and (3.7). Here, the direct effect of e_3 occurring in σ_2 was needed to satisfy the state

Literal Compound Event Direct Cause Indirect Cause Laws (3.1) e_1 $\epsilon_1 = \{e_1, e_2\}$ A(3.2) e_2 B(3.3),(3.7) e_3 $\epsilon_2 = \{e_3\}$ \overline{C} (3.3) e_3 D_ (3.4),(3.5),(3.8) e_4, e_5 $\epsilon_3 = \{e_4, e_5\}$ E(3.4) e_4 F(3.5) e_5

Table 3.2: Explanation of how each literal of θ_E was caused in path ρ_E .

constraint (3.7), which in turn caused B to hold. In the final transition, e_4 and e_5 directly cause E and F, respectively, as per (3.4) and (3.5). Finally, the co-occurrence of e_4 and e_5 in this transition indirectly causes D in accordance with laws (3.4), (3.5), and (3.8). This is a case of contributory cause in which the direct effects of e_4 an e_5 were both required to satisfy the preconditions of the state constraint (3.8), which results in D holding in state σ_4 . Moreover, notice that if e_3 had not occurred in σ_2 , then e_5 would have caused C to hold by law (3.6). However, we do not identify e_5 as a cause because it was preempted by e_3 .

Table 3.2 summarizes the results of our reasoning over the problem. Each row of the table (or sets of rows in cases of overdetermination) characterizes the causation of each literal $l \in \theta_E$ by row. The first column lists the literal l in θ_E that is being explained. The second, third, and fourth columns tell us which event(s) caused l to hold, either directly or indirectly. As a reference for the reader, the final column specifies the laws of AD_E that are relevant to the causation of each l.

In this example, we needed only to reason over the laws of AD_E and the path ρ_E , to produce a fine-grained causal explanation about the direct and indirect causation of the literals of θ_E . We were also able to accurately identify causation in cases of overdetermination, contributory cause, and preemption. The goal of the theoretical framework is to mathematically characterize the type of reasoning process that we used to mentally solve this running example.

3.2 Framework Definitions

Here we present the definitions of the framework, and use them to characterize direct and indirect causation of every literal in θ_E in path ρ_E for our example, as in Table 3.2.

3.2.1 Transition States and Causing Compound Events

The first step in explaining how an outcome θ came to be in path ρ is to identify a *transition state* of the outcome in ρ . A transition state tells us when the outcome of interest *appears* in the path.

Definition 4. Given a problem $\psi = \langle \theta, \rho, AD \rangle$, a state σ_j in ρ is a transition state of θ if $\theta \not\subseteq \sigma_{j-1}$ and $\theta \subseteq \sigma_j$.

The state σ_j is a transition state of θ if the outcome is satisfied in σ_j but not in the immediately previous state σ_{j-1} . It is easy to see that if 4 is satisfied for some σ_j in ρ , then by the successor state equation 2.4 it must be the case that one or more elementary events in ϵ_{j-1} has caused at least one of θ 's literals to hold by σ_j . Note that there may be multiple transition states for an outcome θ in a given path ρ .

In our running example, σ_4 is the only transition state of θ_E because it is only state of ρ_E in which the literals A, B, C, D, E and F hold simultaneously. It is easy to verify that $\theta_E \not\subseteq \sigma_3$ and $\theta_E \subseteq \sigma_4$ in ρ_E using Table 3.1.

Given a transition state σ_j and a literal l in outcome θ , we can identify the most recent compound event to σ_j in ρ to result in l. In other words, we want to find a *causing compound event* ϵ_i that resulted in the most recent transition state of the singleton $\{l\}$ with respect to θ 's transition state σ_j . We first provide a preliminary definition for a *possibly causing compound* event.

Definition 5. Given a problem $\psi = \langle \theta, \rho, AD \rangle$, a transition state σ_j of θ in ρ , and a literal $l \in \theta$, ϵ_i is a possibly causing compound event of l for σ_j if the state σ_{i+1} in ρ is a transition state of $\{l\}$ in ρ and i < j.

Now we may define causing compound events.

Definition 6. Given a problem $\psi = \langle \theta, \rho, AD \rangle$, a transition state σ_j of θ in ρ , a literal $l \in \theta$, and a possibly causing compound event ϵ_i of l for σ_j , ϵ_i is a causing compound event of l for σ_j if there is no other possibly causing compound event $\epsilon_{i'}$ for l in σ_j such that i < i'.

It is easy to see that if there is no causing compound event of $l \in \theta$ for a transition state σ_j , then l must have held in the initial state of ρ and was never changed by a subsequent event prior to σ_j .

It is straightforward to verify using Table 3.1 that ϵ_1 is a causing compound event of A, ϵ_2 is a causing compound event of B and C, and ϵ_3 is a causing compound event of D, E, and F. In all cases, Definition 6 is satisfied because σ_2 is a transition state of A, σ_3 is a transition state of B and C, and σ_4 is a transition state of D, E, and F and there are no other transition states for any literal in $l \in \theta_E$. Therefore, there cannot be a more recent causing compound event of for any such l holding in transition state σ_4 of θ_E .

3.2.2 Direct Cause

Once we know that ϵ_i is a causing compound event for l in σ_i , we can "look inside" of ϵ_i to identify direct and/or indirect causes of l.

Definition 7. Given a problem $\psi = \langle \theta, \rho, AD \rangle$, a transition state σ_j of θ in ρ , a literal $l \in \theta$, and a causing compound event ϵ_i of l, the elementary event $e \in \epsilon_i$ is a direct cause of l for σ_j if $l \in E(e, \sigma_i)$.

If l is in the set of direct effects of e occurring in state σ_j , then e's occurrence was sufficient to directly cause l. Note that direct cause is defined in such a way that multiple events can be direct causes simultaneously as long as l is in the corresponding sets of direct effects. For example, we already know that ϵ_1 is a causing compound event of A holding in σ_4 in the example. Definition 7 tells us that $e_1 \in \epsilon_1$ is a direct cause of A for σ_j . This is because A is in the set $E(e_1, \sigma_1)$ due to the dynamic law (3.1) of AD_E . It can be similarly verified that $e_2 \in \epsilon_1$ is also a direct cause of A because the literal is in $E(e_2, \sigma_1)$. We can also verify that $e_3 \in \epsilon_2$ is an direct cause of C, and finally $e_4 \in \epsilon_3$ and $e_5 \in \epsilon_3$ are direct causes of E and E, respectively.