# Chapter 4: Examples

In this chapter, we test the framework on a number of examples from the literature, as well as a novel example concerning a crash scenario in the context of a self-driving car. By testing the framework on the literature examples, we intend to demonstrate that our framework is able to match or exceed the intuition gained using a counterfactual dependence as a condition for cause. In proposing the self-driving car example, we aim to demonstrate, albeit at a high level, that it is possible to reason about independently operating components in a theoretical distributed system.

## 4.1 Scenario Paths

We represent the elements of each example using an *example tuple* $\Psi = \langle \theta, v, AD \rangle$, where $\theta$ is an outcome of interest, $v$ is a sequence of compound events representing the events of the example, and $AD$ is an action description of the example's domain. In order to ensure that our approach is starting with the same ingredients as other approaches to reasoning about actual cause (events representing a scenario instead of a path), we define *scenario paths* to map the sequence of compound events $v$ to one or more paths of the transition diagram $\tau(AD)$. Scenario paths represent a unique unfolding of a scenario's events with respect to a given domain and provide a convenient representation of how the domain changes over time in response to the events of the scenario. We reason over these paths to explain actual causation.

**Definition 11.** *Given an outcome $\theta$, a sequence of events $v = \langle \epsilon_1, \dots, \epsilon_k \rangle$, and an action description $AD$, a* scenario path *is a path $\rho = \langle \sigma_1, \epsilon_1, \sigma_2, \epsilon_2, \dots, \epsilon_k, \sigma_{k+1} \rangle$ in $\tau(AD)$ satisfying the following conditions:*

1. $\theta \neq \sigma_1$
2. $\exists i, 1 < i \leq k + 1, \theta \subseteq \sigma_i$

Condition 1 requires that the set of fluent literals $\theta$ is not satisfied in the initial state of $\rho$, ensuring that the outcome has not already been caused prior to the known events of the story. Condition 2 requires that $\theta$ is satisfied in at least one state after the initial state in $\rho$. Conditions 1 and 2 together ensure that at least one event is responsible for causing $\theta$ to hold in $\rho$. The successor state equation (2.4) tells us some event in the scenario path must have directly or indirectly caused $\theta$ to be satisfied at some point after the initial state. The set of all scenario paths for an example tuple $\Psi$ is given by $S(\Psi) = \{\rho_1, \rho_2, \ldots, \rho_m\}$. A problem $\psi = \langle \theta, \rho, AD \rangle$ where $\rho \in S(\Psi)$ allows us to inquire about causal information for a specific path rather than for a sequence of events as in $\Psi$.

### 4.1.1 Rock Throwing Problem

First, we will explore two versions of the rock-throwing problem [49], demonstrating that our approach can identify causation in the original example of preemption, and a reformulation of the problem exhibiting overdetermination [50].

**Preemption**

The original scenario posed in [49] is as follows:

> *Throwing a rock at a bottle will cause it to break. Suzy throws the rock at a bottle, and Billy throws a second rock at the bottle. Suzy's rock hits first and the bottle is broken. Who is to blame for the bottle's breaking?*

Intuition tells us that Suzy is responsible for breaking the bottle. We can build upon this example to say that Tommy handed Suzy the rock at the start of the scenario. We extend the scenario to capture these dynamics as follows:

> *Handing a rock to Suzy causes it to be in her possession. Throwing a rock at a bottle will cause it to break. Tommy hands a rock to Suzy, she throws the rock at a bottle, and Billy throws a second rock at the bottle. Suzy's rock hits first and the bottle is broken. Who is to blame for the bottle's breaking?*

Mapping this example to a practical setting, say a vandalism trial, we would certainly want to blame Suzy for succeeding in breaking the bottle, and possibly Tommy as well for his action of

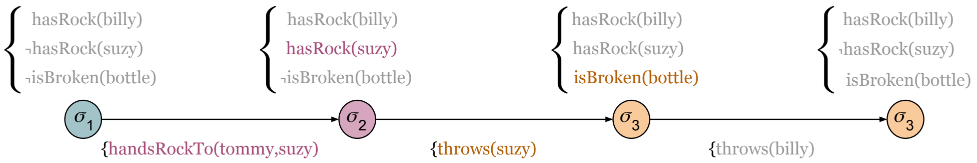

**Figure 4.1:** Path $\rho_B \in S(\Psi_B)$ is a representation of the bottle breaking scenario.

handing the rock to Suzy [1].

The elements of the problem are given by the example tuple $\Psi_Y = \langle \theta_B, v_B, AD_B \rangle$. In this example the outcome of interest is given by $\theta_B = \{isBroken(bottle)\}$. The action description $AD_B$ characterizes the events of the bottle breaking domain:

$$\begin{cases} handsRockTo(tommy, suzy) \text{ \textbf{causes} } hasRock(suzy) & (4.1) \\ throws(suzy) \text{ \textbf{causes} } isBroken(bottle) & (4.2) \\ \qquad \text{\textbf{if} } \neg isBroken(bottle) & \\ throws(billy) \text{ \textbf{causes} } isBroken(bottle) & (4.3) \\ \qquad \text{\textbf{if} } \neg isBroken(bottle) & \\ throws(suzy) \text{ \textbf{causes} } \neg hasRock(suzy) & (4.4) \\ throws(suzy) \text{ \textbf{impossible\_if} } \neg hasRock(suzy) & (4.5) \end{cases}$$

Law (4.1) tells us that if Tommy hands a rock to Suzy, then Suzy has the rock. Laws (4.2) and (4.3) represent the knowledge that if someone throws a rock at a bottle, it will break. Law (4.4) tells us that if Suzy throws the rock, then she no longer is in possession of the rock. Finally, Law (4.5) states that it is impossible for Suzy to throw a rock if it is not in her possession. We assume that there exists a corresponding law for Billy for each of (4.1), (4.4), and (4.5), however we omit them for clarity of the presentation. Consider a path $\rho_B \in S(\Psi_B)$ corresponding to the scenario's events,

[1] The matter of Billy's intention to commit the crime is a matter of judging levels of blame (see e.g. [51]), which is outside of the scope of this work.

represented in Figure 4.1. In the initial state of $\rho_B$, Suzy does not have the rock and the bottle is not broken. After the occurrence of $\epsilon_1 = \{handsRock(tommy, suzy)\}$, $\epsilon_2 = \{throws(suzy)\}$, and $\epsilon_3 = \{throws(billy)\}$, the bottle is broken in state $\sigma_4$. It is straightforward to verify for problem $\psi_B = \langle \theta_B, \rho_B, AD_B \rangle$ that $\sigma_3$ is the only transition state of $\theta_B$ in $\rho_B$ and that $\epsilon_2$ is the causing compound event $isBroken(bottle) \in \theta_B$ holding in $\sigma_3$. The elementary event $throws(suzy)$ is a direct cause of $isBroken(bottle)$ because it is in the set $E(throws(suzy), \sigma_1)$ as per rule (4.2).

There is more causal information to uncover in this problem. We already know that Suzy throwing the rock at the bottle caused it to break, but we want to know if any events supported Suzy's ability to cause the outcome. Rule (4.1) in $AD_B$ tells us that Suzy cannot throw the rock if it is not in her possession. In this case, we can formulate a new problem $\psi'_Y = \langle hasRock(suzy), \rho_B, AD_B \rangle$ and use the framework to determine that $handsRock(tommy, suzy)$ directly caused $hasRock(suzy)$. Table 4.1 summarizes the causal explanations for each of the outcomes we considered. Each row of the first column gives the outcomes of interest for the problems $\psi_Y$ and $\psi'_Y$. The second column identifies the transition state for each outcome in $\rho_B$. The third column gives the direct causes for both problems. It is easy to see that there are no indirect causes for this example.

We have shown that modeling the scenario as a sequence of events occurring over time enables the framework to correctly identify the intuitive cause of the bottle breaking in a classic example of preemption from the literature. We have also shown that it is straightforward to reformulate the problem with a new outcome in order to learn more about the causal mechanism at will.

**Overdetermination**

The scenario can be modified so that Suzy and Billy throw their rocks at the same time, both hitting the bottle simultaneously, upon which it shatters as in [50]. Either rock by itself would have sufficed to shatter the bottle, so we want to identify both throws as direct causes. Omitting the extension in which Tommy hands the rock to either Suzy or Billy, we substitute $v_B$ in $\Psi_B$ with the event sequence $v'_B = \langle \epsilon_1 \rangle$ where $\epsilon_1 = \{throws(suzy), throws(billy)\}$. This yields a new scenario path $\rho'_B$ and the resulting problem for the framework is $\psi'_B = \langle \theta_B, \rho'_B, AD_B \rangle$. It is straightforward to verify that both $throws(suzy)$ and $throws(billy)$ are direct causes for $isBroken(bottle)$ in $\rho'_B$

**Table 4.1:** Overview of explanations of $\{hasRock(suzy)\}$ and $\{isBroken(bottle)\}$ in transition states $\sigma_2$ and $\sigma_3$, respectively.

| *Outcome of Interest* | *State* | *Direct Causes* |
|---|---|---|
| $\{isBroken(bottle)\}$ | $\sigma_4$ | $throws(suzy) \in \epsilon_2$ |
| $\{hasRock(suzy)\}$ | $\sigma_2$ | $handsRockTo(tommy, suzy) \in \epsilon_1,$ |

because $isBroken(bottle)$ is a member of both $E(throws(suzy), \sigma_1)$ and $E(throws(billy)), \sigma_1)$. As before, there are no indirect causes for outcome in the new problem. We have demonstrated that the framework can identify direct causes that match our intuition in this example of overdetermination.

## 4.2 Yale Shooting Problem

### 4.2.1 Direct

Here we use the framework defined above to solve a variant of the well-known Yale shooting problem (YSP) from [52]. The scenario is as follows:

> *Shooting a turkey with a loaded gun will kill it. Suzy shoots the turkey. What is the cause of the turkey's death?*

The YSP example tuple is formalized by $\Psi_Y = \langle \theta_Y, \upsilon_Y, AD_Y \rangle$. The outcome of interest is $\theta_Y = \{\neg isAlive(turkey)\}$. The sequence of events is $\upsilon_Y = \{\epsilon_1, \epsilon_2\}$, where $\epsilon_1 = \{loads(suzy, gun)\}$ and $\epsilon_2 = \{shoots(suzy, turkey)\}$. The action description $AD_Y$ characterizes the events of the YSP domain:

$$\begin{cases} shoots(X, turkey) \textbf{ causes } \neg isAlive(turkey) \textbf{ if } isAlive(turkey) & (4.6) \\ shoots(X, turkey) \textbf{ impossible\_if } \neg isLoaded(gun) & (4.7) \\ \quad loads(X, gun) \textbf{ causes } isLoaded(gun) \textbf{ if } \neg isLoaded(gun) & (4.8) \end{cases}$$

Laws (4.6) and (4.8) are straightforward dynamic laws describing the effects of the events in the YSP domain. Law (4.7) states that the turkey cannot be shot if the gun is not loaded. Consider the path $\rho_Y$, represented in Figure 4.2. In the initial state of $\rho_Y$, the turkey is alive, and the turkey