Causes and Strategies in Multiagent Systems

Sylvia S. Kerkhove
Utrecht University
Utrecht, The Netherlands
s.s.kerkhove@uu.nl

Natasha Alechina
Open University
Heerlen, The Netherlands
Utrecht University
Utrecht, The Netherlands
natasha.alechina@ou.nl

Mehdi Dastani
Utrecht University
Utrecht, The Netherlands
m.m.dastani@uu.nl

ABSTRACT

Causality plays an important role in daily processes, human reasoning, and artificial intelligence. There has however not been much research on causality in multi-agent strategic settings. In this work, we introduce a systematic way to build a multi-agent system model, represented as a concurrent game structure, for a given structural causal model. In the obtained so-called causal concurrent game structure, transitions correspond to interventions on agent variables of the given causal model. The Halpern and Pearl framework of causality is used to determine the effects of a certain value for an agent variable on other variables. The causal concurrent game structure allows us to analyse and reason about causal effects of agents' strategic decisions. We formally investigate the relation between causal concurrent game structures and the original structural causal models.

KEYWORDS

Causality, Multi-Agent Systems, Strategic Behaviour

ACM Reference Format:
Sylvia S. Kerkhove, Natasha Alechina, and Mehdi Dastani. 2025. Causes and Strategies in Multiagent Systems. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 9 pages.

1 INTRODUCTION

Causality plays an important role in Artificial Intelligence [16, 21]. A specific type of causality, called 'actual causality', concerns causal relations between concrete events (e.g. throwing a specific rock shatters a specific bottle) [16]. There is still discussion on what the best definition of actual causality is (see [13, 15, 16] and [7] for some of those definitions). However, most approaches like [13] and [7] use Pearl's [21] structural model framework. In this structural model framework, the world is modelled through variables, which are divided in exogenous and endogenous variables. The former are variables whose values are determined by causes outside of the model and the latter are variables whose values are determined by the variables inside the model (both exogenous and endogenous variables). The functional dependencies between variables are formalised through structural equations. There also exists a rule-based approach that uses logical language to capture causal relations (see [8] and [20]), but we focus on the structural model framework due to its prominence in the literature [1, 7, 10, 12].

While causal models can in principle depict multi-agent systems by making a distinction between agent and environment events, they are less appropriate for reasoning about the abilities and strategies of agents. Concurrent game structures (CGS) have been proposed to reason about agent interactions and strategies [2]. These structures are graphs where nodes correspond to states of the world and edges, labelled with agents' actions, correspond to state transitions [5, 14]. In deterministic settings, an agent strategy specifies the actions to take by the agent.

Let us introduce an example of a causal model. Consider a semi-autonomous vehicle controlled jointly by a human driver and an automatic driving assistance system. This driving assistance system is in turn supported by an obstacle detection system that signals to the driving assistant whether there is an obstacle in front of the vehicle. Both the human driver and the driving assistant control the forward movement of the vehicle, though the human driver can always take full control. In a scenario where there is an obstacle in front of the car, the obstacle causes the obstacle detection system to send a signal to the driving assistant. If the human driver is in a distracted state, this signal causes the driving assistant to avoid an accident. This scenario can be described as a causal system, but can also be viewed as a multi-agent system where the obstacle detection system, the driving assistant and the human driver are all seen as agents that make decisions based on their state observations.

The fundamental relationship between structural causal models and multi-agent system models manifests itself in modelling phenomena such as responsibility for realising a certain outcome by a group of agents. In the literature of multi-agent systems, both structural causal models and CGS are used to define the responsibility of a group of agents for an outcome [10, 23]. Agents in a structural causal model are seen as responsible for an outcome if they have caused it [10]. On the other hand, in a CGS a coalition of agents is deemed responsible for an outcome if they had a strategy to prevent it [23]. By establishing the relationship between structural causal models and CGS, different modelling approaches to multi-agent phenomena (e.g., responsibility) can be compared and unified.

In this paper, we aim to establish a formal relationship between structural causal models and concurrent game structures by constructing a CGS for a given structural causal model such that if a group of agents is an actual cause for an outcome in the causal model, then this group had a strategy in the constructed CGS to prevent the outcome, provided the other agents act as prescribed by the causal model. The CGS is built by distinguishing between agent and environment variables. We consider the values of an agent variable as possible actions of the agent and interventions as agents'

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

decisions. We provide several formal results on how strategies in this causal concurrent game structure (causal CGS) relate to the original structural causal model, establishing a formal relationship between structural causal models and CGS. In particular, we show that a choice of actions by a group of agents is a cause of an outcome in a structural causal model (under the Halpern-Pearl definition of an actual cause) if and only if this set of agents has a strategy for the negation of the outcome in the corresponding causal CGS, provided the other agents act according to the causal model. We believe that our framework will be beneficial for supporting causal inference in multi-agent systems, for example, for reasoning and attributing responsibility for certain outcomes to groups of agents.

We will now first give some preliminaries on causality and concurrent game structures. In Section 3, we define the translation from a structural causal model to a causal CGS, after which, in Section 4, we show how causality in the structural causal model relates to agent strategies in the causal CGS.

2 BACKGROUND

In this section, we introduce the structural causal model framework that we will use. We also shortly introduce concurrent game structures and give a formal definition of agent strategies.

Definition 2.1 (Structural Causal Model, Causal Setting [16]). A *structural causal model* $\mathcal{M}$ is a pair $(\mathcal{S}, \mathcal{F})$, where $\mathcal{S}$ is a signature and $\mathcal{F}$ defines a set of structural equations, relating the values of the variables. A *signature* $\mathcal{S}$ is a tuple $(\mathcal{U}, \mathcal{V}, \mathcal{R})$, where $\mathcal{U}$ is a set of exogenous variables, $\mathcal{V}$ is a set of endogenous variables and $\mathcal{R}$ associates with every variable $X \in \mathcal{U} \cup \mathcal{V}$ a non-empty set $\mathcal{R}(X)$ of possible values for X.

A *causal setting* is a tuple $(\mathcal{M}, \mathbf{u})$, where $\mathcal{M}$ is a causal model and $\mathbf{u}$ a setting for the exogenous variables in $\mathcal{U}$.

The *exogenous variables* are variables whose values depend on factors outside of the model, their causes are not explained by the model [16, 22]. On the other hand, the *endogenous variables* are fully determined by the variables in the model. Note that with $\mathbf{u}$, we use the bold-face notation to denote that $\mathbf{u}$ is a tuple. When we use this bold-face notation for capital letters $\mathbf{X}$ and $\mathbf{Y}$, we are slightly abusing notation by treating them both as tuples and as sets. This follows Halpern's use of the vector notation for both concepts [16]. This means that we can write $\mathbf{X} = \mathbf{x}$ to indicate that the first element of $\mathbf{X}$ gets assigned the value of the first element of $\mathbf{x}$ and so on, but that we can also write $\mathbf{X}' \subseteq \mathbf{X}$.

Example 2.2. Consider the semi-autonomous vehicle example we discussed in the introduction. We can model this example with exogenous binary variables U_O, that determines whether there will be an obstacle on the route, and U_{Att}, that determines whether the human driver is paying attention. For the endogenous variables we introduce the binary variables O, indicating that there is an obstacle, Att, indicating that the human driver is paying attention, HD for whether the human driver keeps driving or brakes. Note that we use HD when the human driver keeps driving ($\neg HD$ indicates that they brake). ODS, indicating that the obstacle detection system detects an obstacle, DA, for whether the driving assistant keeps driving or brakes. Note that we use DA when the driving assistant keeps driving ($\neg DA$ indicates that they brake). And Col, indicating a collision. The set $\mathcal{U}$ is hence $\{U_O, U_{Att}\}$ and the set $\mathcal{V}$ is hence $\{O, Att, HD, ODS, DA, Col\}$. We consider all variables to be Boolean, so for any variable $X \in \mathcal{U} \cup \mathcal{V}$, $\mathcal{R}(X) = \{0, 1\}$.

The following structural equations are defined for this model:

$$O := U_O \qquad Att := U_{Att}$$
$$HD := \neg O \vee (O \wedge \neg Att) \qquad ODS := O$$
$$DA := HD \wedge \neg ODS \qquad Col := DA \wedge HD \wedge O.$$

A *causal network* is a directed graph with nodes corresponding to the causal variables in $\mathcal{V}$ (and $\mathcal{U}$) with an edge from the node labelled X to the node labelled Y if and only if the structural equation for Y depends on X. In other words, we put an edge from node X to node Y if and only if X can influence the value of Y [17]. We call Y a *descendant* of X if the graph contains a path from X to Y.

A model that has an acyclic causal network is called strongly recursive [16]. In such models, a setting $\mathbf{u}$ of the exogenous variables $\mathcal{U}$ fully determines the values of all other (endogenous) variables. We call a causal model with an acyclic causal network recursive because the exogenous variables determine the values of the endogenous variables in a recursive manner. As Halpern explains, some endogenous variables only depend on exogenous variables, we call them first-level variables [16]. They get their value directly from the causal setting. After that, there are the second-level variables, the endogenous variables that depend on both the first-level variables and possibly on the exogenous variables. Likewise, the third-level variables depend on the second-level variables, and possibly on the exogenous and the first-level variables, and so on for higher levels. We only consider strongly recursive models in this paper.

Example 2.3. The causal network for the causal model as described in Example 2.2 is given in Figure 1 (the exogenous variables are not drawn). The graph makes it easy to see that the causal model is recursive, i.e. the causal network does not contain cycles. We can also see the variable levels. O and Att are first-level variables, they only depend on the exogenous variables. HD and ODS only depend on O and Att and hence are second-level variables. DA is a third-level variable, as it depends on second-level variables, and Col is a fourth-level variable, as it depends on both DA and lower-level variables.

Given a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, a formula of the form $X = x$, for $X \in \mathcal{V}$ and $x \in \mathcal{R}(X)$ is called a *primitive event* [16, 17]. These primitive events can be combined with the Boolean connectives $\wedge$, $\vee$ and $\neg$, to form a *Boolean combinations of primitive events* [16, 17]. We follow Halpern and use $(\mathcal{M}, \mathbf{u}) \models \phi$ to denote that formula ϕ holds

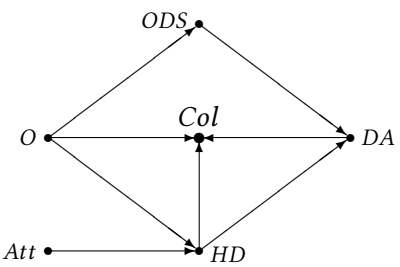

Figure 1: The causal network for the causal model for the semi-autonomous vehicle example described in Example 2.2.

given the values of all variables determined by the causal setting $(\mathcal{M}, \mathbf{u})$ (see [16] for details). A *causal formula* has the form $[Y_1 \leftarrow y_1, ..., Y_k \leftarrow y_k]\varphi$, where φ is a Boolean combination of primitive events, $Y_1, ..., Y_k \in \mathcal{V}$ with $Y_i = Y_j$ if and only if $i = j$, and $y_i \in \mathcal{R}(Y_i)$ for all $1 \leq i \leq k$. Such a formula can be shortened to $[\mathbf{Y} \leftarrow \mathbf{y}]\varphi$, and when $k = 0$ it is written as just φ [17]. $(\mathcal{M}, \mathbf{u}) \models [\mathbf{Y} \leftarrow \mathbf{y}](X = x)$ says that after an intervention that sets all variables of $\mathbf{Y}$ to $\mathbf{y}$, it must be the case that $X = x$ holds in the causal setting $(\mathcal{M}, \mathbf{u})$ (see [16, 17] for more details). We call $\mathbf{y}$ a *setting* for the variables in $\mathbf{Y}$. We now have the necessary background to give the modified HP definition of causality:

Definition 2.4 (modified HP Definition [16]). $\mathbf{X} = \mathbf{x}$ is an *actual cause* of φ in the causal setting $(\mathcal{M}, \mathbf{u})$ if the following 3 conditions hold:

AC1. $(\mathcal{M}, \mathbf{u}) \models \mathbf{X} = \mathbf{x}$ and $(\mathcal{M}, \mathbf{u}) \models \varphi$;
AC2. There is a set $\mathbf{W}$ of variables in $\mathcal{V}$ and a setting $\mathbf{x}'$ of variables in $\mathbf{X}$ s.t. if $(\mathcal{M}, \mathbf{u}) \models \mathbf{W} = \mathbf{w}^*$, then $(\mathcal{M}, \mathbf{u}) \models [\mathbf{X} \leftarrow \mathbf{x}', \mathbf{W} \leftarrow \mathbf{w}^*]\neg\varphi$.
AC3. X is minimal; there is no strict subset $\mathbf{X}'$ of $\mathbf{X}$ s.t. $\mathbf{X}' = \mathbf{x}'$ satisfies AC1 and AC2, where $\mathbf{x}'$ is the restriction of $\mathbf{x}$ to the variables in $\mathbf{X}'$.

If $\mathbf{W} = \emptyset$, we call $\mathbf{X} = \mathbf{x}$ a *but-for cause* of φ.

Example 2.5. Consider our semi-autonomous vehicle example again. Take the causal setting where $\mathbf{u} = (1, 0)$, i.e. $U_O = 1$, there is an obstacle on the route, and $U_{Att} = 0$, the human driver is not paying attention. Following the equations provided in Example 2.2, we have that $(\mathcal{M}, \mathbf{u}) \models O \wedge \neg Att \wedge HD \wedge ODS \wedge \neg DA \wedge \neg Col$. We want to know which agent was the cause of there being no collision. It turns out that both ODS and $\neg DA$ are but-for causes of $\neg Col$, i.e., $(\mathcal{M}, \mathbf{u}) \models [ODS \leftarrow 0]Col$ and $(\mathcal{M}, \mathbf{u}) \models [DA \leftarrow 1]Col$. After all, if we intervene by turning off the object detection system ODS (setting its value to 0 in our model, i.e., replacing equation $ODS = 1$ in our model with $ODS = 0$, which is formally represented as $[ODS \leftarrow 0]$), the driving assistant DA will no longer get a signal that there is an obstacle on the route. This gives $DA = 1$, meaning that the driving assistant will not brake. Because the human driver is distracted in this setting, they will also not brake, and so there will be a collision. Similarly we can also directly intervene on the driving assistant by turning it off (setting its value to 1, not braking, in our model by replacing the equation for DA with $DA := 1$, represented by $[DA \leftarrow 1]$) and there will be a collision as well.

The aim of this work is to connect this concept of structural causal models and causality to concurrent game structures. We use the following definition of concurrent game structures:

Definition 2.6 (Concurrent Game Structures [2]). A *concurrent game structure* (CGS) is a tuple $GS = \langle N, Q, d, \delta, \Pi, \pi \rangle$ with the following components:

- A natural number $N \geq 1$ of agents. We identify the *agents* with the numbers $1, ..., N$.
- A finite set Q of states.
- For each agent $a \in \{1, ..., N\}$ and each state $q \in Q$, a natural number $d_a(q) \geq 1$ of moves available at state q to agent a. We identify the moves of agent a at state q with the numbers $1, ..., d_a(q)$. For each state $q \in Q$, a move vector at q is a tuple $\langle j_1, ..., j_N \rangle$ such that $1 \leq j_a \leq d_a(q)$ for each agent a. Given a state $q \in Q$, we write $D(q)$ for the set $\{1, ..., d_1(q)\} \times \cdots \times \{1, ..., d_N(q)\}$ of *move vectors*. The function D is called *move function*.
- For each state $q \in Q$ and each move vector $\langle j_1, ..., j_N \rangle \in D(q)$, a state $\delta(q, j_1, ..., j_N) \in Q$ that results from state q if every agent $a \in \{1, ..., N\}$ chooses move j_a . The function δ is called *transition function*.
- A finite set Π of *propositions*.
- For each state $q \in Q$, a set $\pi(q) \subseteq \Pi$ of propositions true at q. The function π is the *labelling function*.

When we have a CGS, we can reason about what the optimal actions for a coalition of agents would be in a certain situation. We often use the concept of strategies for this.

Definition 2.7 (Strategy in Concurrent Game Structures [2]). Given a concurrent game structure $S = \langle N, Q, d, \delta, \Pi, \pi \rangle$, a *strategy* for agent $a \in \{1, ..., N\}$ is a function f_a, that maps any (non-empty) finite sequence λ of states in Q to an action the agent can take at the last state of the sequence. I.e. if q is the last state of λ, then $f_a(\lambda) \leq d_a(q)$. We write $F_A = \{f_a \mid a \in A\}$ for a set of strategies of the agents in $A \subseteq \{1, ..., N\}$.

We now have all preliminaries ready to move on and combine causality with concurrent game structures.

3 FROM CAUSAL MODEL TO CGS

The goal of this paper is to define a systematic approach to generate a causal CGS based on a strongly recursive structural causal model. The motivation is that we want to compare the strategic ability of coalitions of agents to realise outcomes to causes in the causal model. Similar translations have been attempted by [3, 12] and [18].

Gladyshev et al. make, like us, a distinction between agent and environment variables, and they also construct a CGS that takes the causal structure between agents' decision and environment variables into account [12] . However, they take a 'zoomed out' approach to the causal model by considering every state in the CGS as a causal model. In contrast, in this paper, we are interested in the specific variable values, which we will consider as specific actions in strategic setting. Another difference with our work is that they do not look at the relationship between causality in the original causal model and strategies in the CGS.

A more similar approach to ours was defined by Baier et al. [3], but they use extensive form games rather than CGS, and do not distinguish between agent and environment variables. Furthermore, while they do show a result relating actual causality in the causal model to some type of strategy in their extensive form game, they only do this for but-for causes, where we consider the modified HP definition as well.

Hammond et al. translate the causal model to a multi-agent influence diagram (MAID) that includes utility variables, with the primary goal of studying rational outcomes of the grand coalition [18]. They hence take a game-theoretic approach, where we take a logic-based approach by focusing on strategic abilities of coalitions of agents. Nevertheless, we could also apply a game-theoretic analysis to our model, by extending our CGS to include utility variables. This is however beyond the scope of this paper.

3.1 Defining a Causal CGS

In this section we will propose a systematic approach to generate a causal concurrent game structure based on a strongly recursive structural causal model. We will use the notion of first-level, second-level and higher-level variables as explained in the previous section to determine in which order the agents of the causal model will get to take actions. For this we define the notion of agent rank:

Definition 3.1. An *agent ranking function* of a causal model $\mathcal{M}$ is a function $\rho : \mathcal{V} \to \{0, ..., n\}$, where n is the number of distinct variable levels for agent variables in $\mathcal{M}$, such that for all $A, B \in V_a$, $\rho(A) > \rho(B) > 0$ if and only if the variable level of A is higher than the variable level of B, and $\rho(A) = \rho(B)$ if and only if A and B have the same variable level. For all $X \in V_e$, $\rho(X) = \rho(A) - 1$ if $\exists A \in V_a$ such that the variable level of X is lower or equal to the variable level of A, and there is no $B \in V_a$ that has a variable level between X and A. If such an A does not exist, i.e. if the variable level of X is higher than the variable level of all $A \in V_a$, then $\rho(X) = n$. The *agent rank* of a variable $A \in V_a$ is $\rho(A)$.

Example 3.2. In the semi-automated vehicle example we say that HD, ODS and DA are the agent variables. We have that $n = 2$ as HD and ODS are both second-level variables and DA is a third level variable as Example 2.3 discusses. There are hence 2 distinct variable levels for the agent variables. From this, it follows that $\rho(HD) = \rho(ODS) = 1$ and $\rho(DA) = 2$, as the variable level of DA is higher than that of HD and ODS and their agent rank needs to be higher than 0 and maximally 2. For the environment variables, we have that $\rho(O) = \rho(Att) = \rho(HD) - 1 = 0$, because there are no first-level agent variables, so we need a second-level agent variable like HD. Finally, we have that $\rho(Col) = 2$, since the variable level of Col is 4 which is higher than all agent variable levels, and hence the agent rank of Col will be the maximum of 2.

We will first define several components of the causal CGS separately before putting them all together. From now on, we will assume that all causal models are recursive and have variables which can only attain finitely many values. Moreover we assume that a set of agent variables $V_a \subseteq \mathcal{V}$ is given.

Definition 3.3 (States of a causal CGS). Given a causal setting $(\mathcal{M}, \mathbf{u})$, let $n = \max_{Y \in V_a} \rho(Y)$ be the maximum value of the agent ranks for the agents in V_a and let $m_i = \prod_{\substack{Y \in V_a, \\ \rho(Y) \leq i}} |\mathcal{R}(Y)|$ be the number of possible combinations of action values for agents with an agent rank of no more than i. The set of *states of a causal CGS*, Q, generated based on $(\mathcal{M}, \mathbf{u})$, is given by:

$$Q = \{q_{0,0}\} \cup \{q_{i,j} \mid 1 \leq i \leq n \text{ and } 0 \leq j < m_i\}.$$

We call $q_{0,0}$ the *starting state* of the causal CGS. Later, we will see that the evaluation in a state $q_{i,j}$ follows from the actions of agents whose agent variables have agent rank i or less.

Example 3.4. We will use the causal model for the semi-automated vehicle example to define a causal CGS (see Figure 1). See Example 3.2 for the agent rank of all variables of the causal model. We start with the setting $(\mathcal{M}, \mathbf{u})$ with $\mathbf{u} = (U_O = 1, U_{Att} = 0)$. The set of states is then $Q = \{q_{0,0}, q_{1,0}, q_{1,1}, q_{1,2}, q_{1,3}, q_{2,0}, q_{2,1}, q_{2,2}, q_{2,3}, q_{2,4}, q_{2,5}, q_{2,6}, q_{2,7}\}$. Note that:
$\prod_{Y \in V_a, \rho(Y) \leq 1} |\mathcal{R}(Y)| = \prod_{Y \in \{HD, ODS\}} |\mathcal{R}(Y)| = |\{0,1\} \times \{0,1\}| = 4$ and $\prod_{Y \in V_a, \rho(Y) \leq 2} |\mathcal{R}(Y)| = \prod_{Y \in \{HD, ODS, DA\}} |\mathcal{R}(Y)| = 8$, so for $i = 1$, we have $j \in \{0, \ldots, 3\}$ and for $i = 2$, we have $j \in \{0, \ldots, 7\}$. These are all the states, because the maximum value of the agent rank ρ is 2.

We will now define the agent actions in those states.

Definition 3.5 (Actions in a causal CGS). Given a causal setting $(\mathcal{M}, \mathbf{u})$ and Q the corresponding set of states as defined by Definition 3.3. The possible *actions* for an agent $k \in \{1, ..., N\}$ in a state $q_{i,j} \in Q$ are $d_k(q_{i,j}) = \mathcal{R}(A_k)$, where A_k is the agent variable controlled by agent k, and $\rho(A_k) = i + 1$. Otherwise $d_k(q_{i,j}) = \{0\}$.

The intuition behind this definition is that agent variables that are earlier on a causal path will earlier get to take an action as the agent variables later on a causal path depend on them. The order of agent variables on a causal path can be seen as representing a protocol that determines when each agent has to take its action. We write a_k to denote an action of agent $k \in N$ and $\mathbf{a}_{i,j} = \langle a_1, ..., a_N \rangle$ to denote an action profile taken in a certain state $q_{i,j}$, i.e., all actions taken by all agents in state $q_{i,j}$. It is important to note that for a given index i all states $q_{i,j}$ have the same action profiles that can be taken in them, regardless of the value of j. We denote this set with $\mathbf{A}_i$. Instead of d_k for agent k, we will sometimes write d_{A_k} for the agent variable A_k corresponding to agent k.

Example 3.6. We continue with the situation as in Example 3.4. The available actions for each agent in each state are:

$$\begin{array}{ll} d_{HD}(q_{0,0}) = d_{ODS}(q_{0,0}) = \{0,1\}, & d_{DA}(q_{0,0}) = \emptyset, \\ d_{HD}(q_{1,j}) = d_{ODS}(q_{1,j}) = \emptyset, & d_{DA}(q_{1,j}) = \{0,1\}, \\ \forall j \in \{0, \ldots, 3\}, \text{ and} & \\ d_{HD}(q_{2,j}) = d_{ODS}(q_{2,j}) = \emptyset, & d_{DA}(q_{2,j}) = \emptyset, \\ \forall j \in \{0, \ldots, 7\}. & \end{array}$$

These actions must of course lead to transitions to new states.

Definition 3.7 (Transitions in a causal CGS). Given a causal setting $(\mathcal{M}, \mathbf{u})$, Q the corresponding set of states as defined by Definition 3.3 and actions as defined by Definition 3.5, the state following from the action profile $\mathbf{a}_{i,j} \in \mathbf{A}_i$, with $i < \max_{X \in V_a} \rho(X)$, is given by the transition function δ, where $\delta(q_{i,j}, \mathbf{a}_{i,j}) = q_{i+1,j'}$ and $|\mathbf{A}_i| \cdot j \leq j' \leq |\mathbf{A}_i| \cdot (j+1) - 1$, under the condition that if $\mathbf{a}_{i,j} \neq \mathbf{a}'_{i,j}$, then $\delta(q_{i,j}, \mathbf{a}_{i,j}) \neq \delta(q_{i,j}, \mathbf{a}'_{i,j})$. If $i = \max_{X \in V_a} \rho(X)$, we define $\delta(q_{i,j}, \mathbf{a}_{i,j}) = q_{i,j}$. In this case, there is only one possible action profile $\mathbf{a}_{i,j}$ consisting of only the 0 action.

This definition simply says that every unique action profile in a state leads to a unique new state. This leads to the causal CGS having a tree structure. It is impossible to return to an earlier state and every node can only branch out

Example 3.8. Continuing with our running example, we will write $\langle 1, 0, 0 \rangle$ for the action profile $\langle HD = 1, ODS = 0, DA = 0 \rangle$. We get that the transitions are:

$$\begin{array}{ll} \delta(q_{0,0}, \langle 0,0,0 \rangle) = q_{1,0}, & \delta(q_{0,0}, \langle 0,1,0 \rangle) = q_{1,1}, \\ \delta(q_{0,0}, \langle 1,0,0 \rangle) = q_{1,2}, & \delta(q_{0,0}, \langle 1,1,0 \rangle) = q_{1,3}, \\ \delta(q_{1,0}, \langle 0,0,0 \rangle) = q_{2,0}, & \delta(q_{1,0}, \langle 0,0,1 \rangle) = q_{2,1}, \\ \delta(q_{1,1}, \langle 0,0,0 \rangle) = q_{2,2}, & \delta(q_{1,1}, \langle 0,0,1 \rangle) = q_{2,3}, \\ \delta(q_{1,2}, \langle 0,0,0 \rangle) = q_{2,4}, & \delta(q_{1,2}, \langle 0,0,1 \rangle) = q_{2,5}, \\ \delta(q_{1,3}, \langle 0,0,0 \rangle) = q_{2,6}, & \delta(q_{1,3}, \langle 0,0,1 \rangle) = q_{2,7}, \\ \delta(q_{2,j}, \langle 0,0,0 \rangle) = q_{2,j} & \forall j \in \{0, \ldots, 7\}. \end{array}$$

Now that we have states, actions and transitions, we just need the evaluations of the states. The evaluation of a state will depend on an initial causal setting and the actions the agents have taken up to this state. The agents fully determine the values of the agent variables, the environment variables follow from these values and the context that was used to define the causal CGS.

Definition 3.9 (Evaluation of states in a causal CGS). Given a causal setting, $(\mathcal{M}, \mathbf{u})$, the set of all possible propositions for the generated causal CGS is $\Pi = \{X = x \mid X \in \mathcal{V}, x \in \mathcal{R}(X)\}$. The valuation of each state $q_{i,j} \in Q$, with Q the set of states of the causal CGS according to Definition 3.3, is defined recursively by the *labelling function* π, as:

$$\begin{aligned} \pi(q_{0,0}) &= \{Y = y \mid (\mathcal{M}, \mathbf{u}) \models Y = y\} \\ \pi(\delta(q_{i,j}, \mathbf{a}_{i,j})) &= \{Y = y \mid (\mathcal{M}^{\mathbf{X}_{i,j} \leftarrow \mathbf{x}_{i,j}, \mathbf{A}_{i,j} \leftarrow \mathbf{a}_{i,j}}, \mathbf{u}) \models Y = y\}, \end{aligned}$$

where $\mathbf{a}_{i,j}$ is an action profile for state $q_{i,j}$, $\mathbf{A}_{i,j} \leftarrow \mathbf{a}_{i,j} := \{A_k \leftarrow a_k \mid A_k \in V_a, \rho(A_k) = i+1 \text{ and } a_k \in \mathbf{a}_{i,j}\}$ is an intervention constructed based on action profile $\mathbf{a}_{i,j}$, and $\mathbf{X}_{i,j} \leftarrow \mathbf{x}_{i,j}$ is recursively defined by: $\mathbf{X}_{i+1,j'} \leftarrow \mathbf{x}_{i+1,j'} := \mathbf{X}_{i,j} \leftarrow \mathbf{x}_{i,j} \cup \mathbf{A}_{i,j} \leftarrow \mathbf{a}_{i,j}$, if $\delta(q_{i,j}, \mathbf{a}_{i,j}) = q_{i+1,j'}$ with $\mathbf{X}_{0,0} \leftarrow \mathbf{x}_{0,0} = \emptyset$.

Definition 3.9 says that an agent action leads to an intervention on the causal setting the causal CGS was based upon. We can see $\mathbf{A}_{i,j} \leftarrow \mathbf{a}_{i,j}$ as the intervention that directly follows from the agent action(s) taken in the state $q_{i,j}$, $\mathbf{X}_{i,j} \leftarrow \mathbf{x}_{i,j}$ stores the previous interventions that were made leading up to the state $q_{i,j}$. We will illustrate this in the following example.

Example 3.10. We continue with the situation as in Example 3.8. We start with the causal setting where $U_O = 1$ and $U_{Att} = 0$, so $\pi(q_{0,0}) = \{O, \neg Att, HD, ODS, \neg DA, \neg Col\}$. To determine $\pi(q_{1,0}) = \pi(\delta(q_{0,0}, \langle 0,0,0 \rangle))$, we need $\mathbf{A}_{0,0} \leftarrow \mathbf{a}_{0,0} = \{HD \leftarrow 0, ODS \leftarrow 0\}$. This gives us that $\pi(q_{1,0}) = \{Y = y \mid (\mathcal{M}^{HD \leftarrow 0, ODS \leftarrow 0}, \mathbf{u}) \models Y = y\}$ $= \{O, \neg Att, \neg HD, \neg ODS, \neg DA, \neg Col\}$. Similarly we can determine that $\pi(q_{1,1}) = \{O, \neg Att, \neg HD, ODS, \neg DA, \neg Col\}$, $\pi(q_{1,2}) = \{O, \neg Att, HD, \neg ODS, DA, Col\}$ and $\pi(q_{1,3}) = \{O, \neg Att, HD, ODS, \neg DA, \neg Col\}$.

Let us now look at $\pi(q_{2,1}) = \pi(\delta(q_{1,0}, \langle 0,0,1 \rangle))$. We need $\mathbf{X}_{1,0} \leftarrow \mathbf{x}_{1,0} = (\mathbf{X}_{0,0} \leftarrow \mathbf{x}_{0,0} \cup \mathbf{A}_{0,0} \leftarrow \mathbf{a}_{0,0}) = \emptyset \cup \{HD \leftarrow 0, ODS \leftarrow 0\}$ as we determined above. The new $\mathbf{A}_{1,0} \leftarrow \mathbf{a}_{1,0} = \{DA \leftarrow 1\}$ and so $\pi(q_{2,1}) = \{Y = y \mid (\mathcal{M}^{HD \leftarrow 0, ODS \leftarrow 0, DA \leftarrow 1}, \mathbf{u}) \models Y = y\}$ $= \{O, \neg Att, \neg HD, \neg ODS, DA, \neg Col\}$. The valuations for the other states are determined similarly (and are shown in Figure 2).

Now that we have these four definitions, we can give the full definition of a causal CGS.

Definition 3.11 (Causal CGS). Given a causal setting, $(\mathcal{M}, \mathbf{u})$, a *causal concurrent game structure* is defined as a tuple $GS = \langle N, Q, d, \delta, \Pi, \pi \rangle$ where $N = |V_a|$, every agent only controls one agent variable, Q is a set of states, as defined by Definition 3.3. For every agent $k \in \{1, ..., N\}$, $d_k(q_{i,j})$ gives the moves available to this agent in state $q_{i,j} \in Q$, as given by Definition 3.5. The transition function δ is defined as in Definition 3.7. The set of possible propositions Π and the valuation function π are given by Definition 3.9.

We can now add the results of the previous examples together and give a full causal CGS for the semi-automated vehicle example.

Example 3.12. Using Definition 3.11, we define $N = |V_a| = |\{HD, ODS, DA\}| = 3$. This gives us a full causal CGS, illustrated in Figure 2.

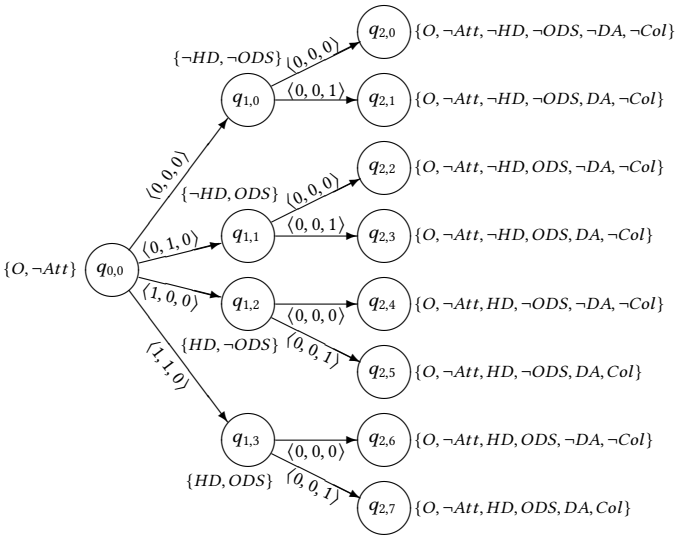

Figure 2: The causal CGS of the semi-automated vehicle example. We only show the initial values of the variables of agent rank 0 in the starting state. In the middle states we only show the variables with agent rank corresponding to that state. We also do not show the transitions to the same state in the leaf-states.

3.2 Properties of Causal Concurrent Game Structures

We already mentioned that a causal CGS has a tree structure. In the rest of this paper, we will call states $q_{i,j}$, with $i = \max_{X \in \mathcal{V}} \rho(X)$, the *leaf-states*. We will call actions in states where an agent does not control a variable, i.e. $a_k = 0$, when $d_k(q_{i,j}) = \{0\}$, with $\rho(X) \neq i+1$, *no-op actions*. It is also useful to define an *action path* for a state $q_{i,j}$, that contains all the non no-op actions that led to the state. In other words, the action path contains only the actions that agents took in a state where they could actually choose an action. We will denote this sequence of actions as $\alpha[q_{i,j}]$. Formally, for $0 \leq k \leq N$, an action a_k is in this set of actions $\alpha[q_{i,j}]$ if and only if $\rho(A_k) \leq i$ and there exists an action profile $\mathbf{a}_{i',j'}$, containing a_k, such that $q_{i',j'} \in \lambda[q_{i,j}, i]$ (the history of $q_{i,j}$) and $\delta(q_{i',j'}, \mathbf{a}_{i',j'}) \in \lambda[q_{i,j}, i]$. In other words, an action is on the action path for a state $q_{i,j}$, if the state $q_{i',j'}$ in which the action is taken lies on the history of $q_{i,j}$, and the successor of $q_{i',j'}$ can be reached when taking this action.

Our first result is on the size of the causal CGS.

Proposition 3.13. *Let $\mathcal{M} = (\mathcal{S}, \mathcal{F})$ be a causal model. The size of the causal CGS generated by $\mathcal{M}$ is linear in the size of the extension of $\mathcal{F}$.*

Proof. Consider a structural causal model $\mathcal{M} = (\mathcal{S}, \mathcal{F})$. Observe that $\mathcal{F}$ specifies the value of each variable for all possible combinations of values of all other variables. Hence $\mathcal{F}$ corresponds to a table of size $|\mathcal{V}| \times \prod_{X \in \mathcal{V}} |\mathcal{R}(X)|$ (the number of cells), which is actually the extension of $\mathcal{F}$. We now show that the number of states in the causal CGS is $O(\prod_{Y \in V_a} |\mathcal{R}(Y)|)$.

By Definition 3.3 we have that the number of states of the causal CGS, is given by $|Q| = 1 + \sum_{i=1}^{n} \prod_{\substack{Y \in V_a, \\ \rho(Y) \leq i}} |\mathcal{R}(Y)|$, where

$n = \max_{Y \in V_a} \rho(Y)$. The number of leaf-states is hence given by $\prod_{Y \in V_a} |\mathcal{R}(Y)| =: |R(V_a)|$. The number of states for $i = n-1$ will be at most half $|R(V_a)|$, as there will be at least one variable of rank n that is hence not included in $\prod_{\substack{Y \in V_a, \\ \rho(Y) \leq n-1}} |\mathcal{R}(Y)|$, and this variable will have at least two possible values. We can continue this argument until $i = 1$, which shows us that $|Q|$ is bounded by $1 + \frac{1}{2^{n-1}}|R(V_a)| + \cdots + \frac{1}{2}|R(V_a)| + |R(V_a)| \leq 2|R(V_a)|$. Hence the number of states in the causal CGS is $O(\prod_{Y \in V_a} |\mathcal{R}(Y)|)$. Since a causal CGS is a tree and each state has at most one predecessor, the number of transitions (the size of δ) is also $O(\prod_{Y \in V_a} |\mathcal{R}(Y)|)$, hence linear in the size of $\mathcal{F}$ in the original model. □

The statement in the following lemma is a direct consequence of the way the valuation of states is determined in a causal CGS. It states that a variable value cannot change in states corresponding to a higher agent rank than the agent rank of the variable itself.

Lemma 3.14. *Let GS be a causal CGS generated by the causal model $\mathcal{M}$. For any endogenous causal variable $X \in \mathcal{V}$ of $\mathcal{M}$, with $\rho(X) = i$, it holds that $(X = x) \in \pi(q_{i,j})$ for some state $q_{i,j}$ of GS, if and only if $(X = x) \in \pi(q_{i',j'})$ for all states $q_{i',j'}$ that are descendants of $q_{i,j}$.*

Proof. Let $(X = x) \in \pi(q_{i,j})$. Variable values can change in a state due to interventions, but the only new interventions done in states descended from $q_{i,j}$ are interventions on variables with an agent rank higher than i. X has agent rank i, so by the definition of agent rank none of those variables can be ancestors of X. They are hence unable to influence the value of X. Therefore $(X = x) \in \pi(q_{i',j'})$ for all states $q_{i',j'}$ descended from $q_{i,j}$.

Now, let $(X = x) \in \pi(q_{i',j'})$ for all states $q_{i',j'}$ that are descended from $q_{i,j}$. The value of X was not changed in any of those states, because the value of X can only change due to an intervention on X or an ancestor variable of X, so only due to variables of agent rank smaller or equal to $\rho(X)$. The only interventions on variables that happen in the descendants of $q_{i,j}$ are on variables of agent rank higher than $\rho(X)$, hence X must have had the same value in $q_{i,j}$, i.e. $(X = x) \in \pi(q_{i,j})$. □

We define the notion of *correspondence* to talk about how states in a causal CGS connect to a causal model.

Definition 3.15 (Correspondence). We say that a state $q_{i,j}$ of a causal CGS *corresponds* to a causal setting $(\mathcal{M}^{\mathbf{Y} \leftarrow \mathbf{y}}, \mathbf{u})$, where $\mathbf{Y} \subseteq \mathcal{V}$, if for all causal variables X of $\mathcal{M}$, $(X = x) \in \pi(q_{i,j})$ if and only if $(\mathcal{M}^{\mathbf{Y} \leftarrow \mathbf{y}}, \mathbf{u}) \models X = x$.[1]

We will sometimes say that a causal setting $(\mathcal{M}^{\mathbf{Y} \leftarrow \mathbf{y}}, \mathbf{u})$ corresponds to a state $q_{i,j}$ of a causal CGS and mean the same thing. Note that the set $\mathbf{Y}$ could also be empty. Hence the causal model $\mathcal{M}^{\mathbf{Y} \leftarrow \mathbf{y}}$ in Definition 3.15 could also be $\mathcal{M}$.

We can show that a leaf-state of a causal CGS corresponds to a causal setting $(\mathcal{M}^{\mathbf{Y} \leftarrow \mathbf{y}}, \mathbf{u})$, where $\mathbf{Y} \leftarrow \mathbf{y}$ depends on the action path that leads to the leaf-state. This connects the definition of causal CGS to the theory of causal models.

[1] So the causal variable X has value x in the causal setting $(\mathcal{M}^{\mathbf{Y} \leftarrow \mathbf{y}}, \mathbf{u})$.

Proposition 3.16. *Let GS be a causal CGS generated by a causal setting $(\mathcal{M}, \mathbf{u})$. If $q_{n,m}$ is a leaf-state of GS, then $q_{n,m}$ corresponds to the causal setting $(\mathcal{M}^{\mathbf{Y} \leftarrow \mathbf{y}}, \mathbf{u})$, where $\mathbf{Y} \leftarrow \mathbf{y} = \{A_k \leftarrow a_k \mid A_k \in V_a \text{ and } a_k \in \alpha[q_{n,m}]\}$, with $\alpha[q_{n,m}]$ the action path for $q_{n,m}$.*

Proof. By Definition 3.9, $(X = x) \in \pi(q_{n,m})$ if and only if $(\mathcal{M}^{\mathbf{X}_{i,j} \leftarrow \mathbf{x}_{i,j}, \mathbf{A} \leftarrow \mathbf{a}}, \mathbf{u}) \models X = x$, where $\mathbf{A} \leftarrow \mathbf{a}$ are the actions taken in the state before $q_{n,m}$, and $\mathbf{X}_{i,j} \leftarrow \mathbf{x}_{i,j}$ are all previously taken actions. Hence $\mathbf{Y} \leftarrow \mathbf{y} = (\mathbf{A} \leftarrow \mathbf{a}) \cup \mathbf{X}_{i,j} \leftarrow \mathbf{x}_{i,j}$ and the proposition is proven. □

This gives us a solid grasp on how a causal CGS relates to the causal model that generates it. We will use this in the next section when we talk about the connection between agent strategies in a causal CGS and causality in this structural causal model.

4 CAUSALITY IN CAUSAL CGS

Now that we have defined causal concurrent game structures and shown what their states represent, it is time to look at how we can use them. In this section, we will show some relations between causal CGS and the modified HP definition of actual causality, but we first introduce the notion of a causal strategy profile.

From now on, we will denote the set of all agents in a model by Σ. Specifically, for a causal CGS, $\Sigma = \{k \mid X_k \in V_a\}$. This set will also be called the *grand coalition* at times. We will use the notation $F_{X_k = x}$ to denote the strategy for agent k where it takes action x as its non no-op action. Formally,

$$F_{X_k=x}(q_{i,j}) = \begin{cases} x & \text{if} \quad \rho(X_k) = i+1 \\ 0 & \text{else} \end{cases}$$

For a set of agents $\mathbf{X}$, we write $F_{\mathbf{X}=\mathbf{x}}$ to indicate the set of strategies $\{F_{X_k=x} \mid X_k \in \mathbf{X}, x \in \mathbf{x}\}$. Let F_A be a strategy for a set of agents A, and F_B a strategy for a set of agents B. Following notation in [9], we will write $F_A \circ F_B$ to denote a strategy profile for the agents in $A \cup B$ that follows strategy F_A for agents in A and strategy F_B for agents in $B \backslash A$.

We define the causal strategy profile as a way to capture the 'normal' behaviour of agents when they would follow the causal model.

Definition 4.1 (Causal Strategy Profile). Given a causal setting $(\mathcal{M}, \mathbf{u})$ and the causal CGS generated by this setting. Define the *causal strategy profile* $F_{\mathcal{M}}$ as $F_{\mathcal{M}} = \{F_{X_k} \mid k \in \Sigma\}$, where $F_{X_k}(q_{i,j}) = 0$ if $\rho(X_k) \neq i+1$, and $F_{X_k}(q_{i,j}) = x_k$ otherwise, where x_k is such that $(\mathcal{M}, \mathbf{u}) \models [\mathbf{X} \leftarrow \mathbf{x}]X_k = x_k$, with $\mathbf{X} = \{X_{k'} \mid \rho(X_{k'}) < \rho(X_k)\}$ and $\mathbf{x} = \{x_{k'} \mid x_{k'} \in \alpha[q_{i,j}]\}$.

Recall that $\alpha[q_{i,j}]$ is the action path up to state $q_{i,j}$. If we want an agent k to follow a strategy F_k and the rest of the agents to follow the causal strategy profile, we denote this as $F_k \circ F_{\mathcal{M}}$. If a set of agents follows the causal strategy profile, that means that in every state, the agents take the actions that assign those values to the agent variables that they would also have gotten in the causal setting on which the causal CGS is based, given the actions of the other agents.

Example 4.2. In the semi-automated vehicle example, given the setting where $U_O = 1$ and $U_{Att} = 0$, the causal strategy profile $F_{\mathcal{M}}$ is such that the human driver does not brake, but the obstacle

detection system detects the obstacle. The driving assistant brakes in this case, but whenever one of the *HD* or *ODS* performs another action, *DA* does not brake. The causal strategy profile for a causal CGS generated by this causal setting is given in Figure 3.

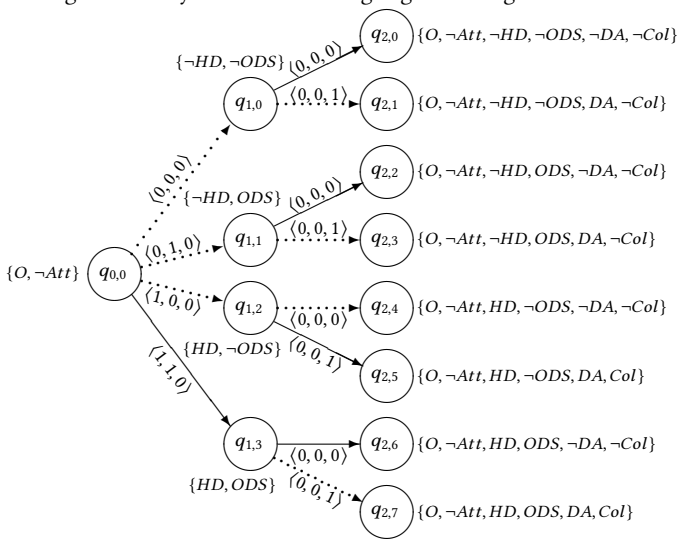

Figure 3: The causal CGS of the semi-automated vehicle example. The dotted lines indicate actions that are not following the causal strategy profile.

In the following lemma, we relate deviations from the causal strategy profile to interventions in the structural causal model that generated the causal CGS. This can be used to relate agent strategies in the causal CGS to causality in the causal model.

Lemma 4.3. *Let GS be a causal CGS based on a causal setting* $(\mathcal{M}, \mathbf{u})$. *If* $q_{n,m}$ *is the leaf-state of GS that results from the strategy profile* $F_{\mathbf{X}=\mathbf{x}} \circ F_{\mathcal{M}}$, *then* $q_{n,m}$ *corresponds to* $(\mathcal{M}^{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{u})$.

Sketch of Proof. The whole proof can be found in the full version of the paper on arXiv [19], here we just give a sketch of the approach. This lemma can be proven by induction on the agent rank of X. For the base step, if $\rho(X) = 0$, it means that X is an environment variable and does not depend on any other endogenous variables. We use Lemma 3.14 and this fact to show that $(\mathcal{M}^{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{u}) \models X = x$. The induction hypothesis (IH) will suppose that for all $X \in \mathcal{V}$ s.t. $\rho(X) \leq i$, $(X = x) \in \pi(q_{n,m})$ if and only if $(\mathcal{M}^{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{u}) \models X = x$. The inductive step will then consider the cases where $X \in V_a$ and $X \in V_e$ separately. In the first case, X can be in $\mathbf{X}$, which means that it gets its value x directly from $\mathbf{x}$. Otherwise, it gets its value from $F_{\mathcal{M}}$, we use the definition of $F_{\mathcal{M}}$ and the IH to show that $(X = x) \in \pi(q_{n,m}) \Leftrightarrow (\mathcal{M}^{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{u}) \models X = x$. If $X \in V_e$ we use what we have just shown and the IH to show the same thing. □

The following corollary follows directly from this lemma, it shows that there is a leaf-state in a causal CGS that corresponds to the original causal setting.

Corollary 4.4. *Let GS be a causal CGS based on a causal setting* $(\mathcal{M}, \mathbf{u})$. *If* $q_{n,m}$ *is the leaf-state resulting from all agents following the causal strategy profile* $F_{\mathcal{M}}$, *then* $q_{n,m}$ *corresponds to* $(\mathcal{M}, \mathbf{u})$.

Proof. This is a special case of Lemma 4.3, where $\mathbf{X} = \emptyset$. □

We can check whether this result holds in our semi-automated vehicle example. We see in Figure 3 that if all agents follow the causal strategy profile, they end up in state $q_{2,6}$ with $\pi(q_{2,6}) = \{O, \neg Att, HD, ODS, \neg DA, \neg Col\}$. The causal CGS was based on the causal setting where there is an obstacle on the road and the driver is not paying attention, in this case we have $(\mathcal{M}, \mathbf{u}) \models O, \neg Att, HD, ODS, \neg DA, \neg Col$ which does correspond to state $q_{2,6}$, as Corollary 4.4 predicted.

With Lemma 4.3 we can show that if a set of agents $\mathbf{X}$ causes φ according to the modified HP definition, with a given witness, then in the causal CGS generated by the causal setting that holds this witness fixed, these agents have a strategy to guarantee $\neg\varphi$ in a leaf-state, provided that all other agents follow the causal strategy profile and vice versa.

Proposition 4.5. *Let* $\Gamma = \{k \mid X_k \in \mathbf{X}\}$ *be a set of agents,* $\mathbf{x}$ *a setting for the variables in* $\mathbf{X}$, *and let* $(\mathcal{M}, \mathbf{u})$ *be a causal setting with* $(\mathcal{M}, \mathbf{u}) \models \varphi$. $\mathbf{X} = \mathbf{x}$ *is, according to the modified HP definition, a cause of causal formula* φ *in this causal setting* $(\mathcal{M}, \mathbf{u})$, *with witness* $\mathbf{W} = \mathbf{w}^*$ *if and only if in the causal CGS generated by the causal setting,* $(\mathcal{M}^{\mathbf{W}\leftarrow\mathbf{w}^*}, \mathbf{u})$, Γ *has a strategy* F_Γ *such that,* $\neg\varphi$ *will hold in the leaf-state* $q_{n,m}$ *resulting from the strategy profile* $F_\Gamma \circ F_{\mathcal{M}}$.

Proof. We first prove the cause to strategy direction. In this case, $\mathbf{X} = \mathbf{x}$ is a cause of φ, with witness $\mathbf{W} = \mathbf{w}^*$ so there exists an alternative value for $\mathbf{X}$, $\mathbf{x}'$ such that $(\mathcal{M}, \mathbf{u}) \models [\mathbf{X} \leftarrow \mathbf{x}', \mathbf{W} \leftarrow \mathbf{w}^*]\neg\varphi$. Let $F_\Gamma = \{F_{X_k=x} \mid x \in \mathbf{x}' \text{ if } X_k \in \mathbf{X}\}$. By Lemma 4.3, the leaf-state $q_{n,m}$ corresponds to $(\mathcal{M}^{\mathbf{X}\leftarrow\mathbf{x}',\mathbf{W}\leftarrow\mathbf{w}^*}, \mathbf{u})$, and we have that $(\mathcal{M}^{\mathbf{X}\leftarrow\mathbf{x}',\mathbf{W}\leftarrow\mathbf{w}^*}, \mathbf{u}) \models \neg\varphi$ and hence $\neg\varphi$ holds in $q_{n,m}$.

Now for the other direction, let F_Γ be the strategy such that $\neg\varphi$ will hold in the leaf-state $q_{n,m}$ that results from the strategy profile $F_\Gamma \circ F_{\mathcal{M}}$ in the causal CGS generated by the causal setting $(\mathcal{M}^{\mathbf{W}\leftarrow\mathbf{w}^*}, \mathbf{u})$. Let $\mathbf{x}$ be such that $(\mathcal{M}, \mathbf{u}) \models \mathbf{X} = \mathbf{x}$ and let $\mathbf{x}'$ be such that $\mathbf{X} = \mathbf{x}' \subseteq \pi(q_{n,m})$. By Lemma 4.3, $q_{n,m}$ must correspond to $(\mathcal{M}^{\mathbf{X}\leftarrow\mathbf{x}',\mathbf{W}\leftarrow\mathbf{w}^*}, \mathbf{u})$. Hence $(\mathcal{M}, \mathbf{u}) \models [\mathbf{X} \leftarrow \mathbf{x}', \mathbf{W} \leftarrow \mathbf{w}^*]\neg\varphi$ and by definition we have that $(\mathcal{M}, \mathbf{u}) \models \mathbf{X} = \mathbf{x} \wedge \varphi$. Moreover, $\mathbf{x} \neq \mathbf{x}'$, because if they were the same it would not be the case that setting $\mathbf{X}$ to $\mathbf{x}'$ would give a different result than the original causal setting (the determinism axiom of the causal reasoning axioms [16]). Hence $\mathbf{X} = \mathbf{x}$ is a cause of φ according to the modified HP definition, with witness $\mathbf{W} = \mathbf{w}^*$. □

This result cannot be used to find causes in a causal CGS, because one would already need to know the witness. However, we have another result for the causal setting where the witness was not held fixed, provided the witness consists of only agent variables. The following proposition states that in that case, the set of agents consisting of both the cause and the witness variables has a strategy to guarantee $\neg\varphi$ in a leaf-state, provided all other agents follow the causal strategy profile and vice versa.

Proposition 4.6. *Let* $\Gamma = \{k \mid X_k \in \mathbf{X} \cup \mathbf{W},$ *and* $\mathbf{X} \cup \mathbf{W} \subseteq V_a\}$ *be a set of agents,* $\mathbf{x}, \mathbf{w}^*$ *are settings for the variables in* $\mathbf{X}, \mathbf{W}$ *respectively, and let* $(\mathcal{M}, \mathbf{u})$ *be a causal setting with* $(\mathcal{M}, \mathbf{u}) \models \varphi$. $\mathbf{X} = \mathbf{x}$ *is, according to the modified HP definition, a cause of causal formula* φ *in this causal setting* $(\mathcal{M}, \mathbf{u})$, *with witness* $\mathbf{W} = \mathbf{w}^*$ *if and only if in the causal CGS generated by this causal setting,* Γ *has a strategy*

F_Γ such that, $\neg\varphi$ will hold in the leaf-state $q_{n,m}$ resulting from the strategy profile $F_\Gamma \circ F_\mathcal{M}$.

Sketch of Proof. The proof for this proposition is very similar to the proof for Proposition 4.5. The main difference is that the strategy for the witness variables also needs to be defined. The full proof can be found in the full version of the paper [19]. □

As but-for causes have no witness, they give a stronger result.

Corollary 4.7. *Let $\Gamma = \{k \mid X_k \in \mathbf{X}\}$ be a set of agents, $\mathbf{x}$ a setting for the variables in $\mathbf{X}$, and let $(\mathcal{M}, \mathbf{u})$ be a causal setting with $(\mathcal{M}, \mathbf{u}) \models \varphi$. $\mathbf{X} = \mathbf{x}$ is a but-for cause of causal formula φ in this causal setting $(\mathcal{M}, \mathbf{u})$ if and only if in the causal CGS generated by the causal setting, $(\mathcal{M}, \mathbf{u})$, Γ has a strategy F_Γ such that, $\neg\varphi$ will hold in the leaf-state $q_{n,m}$ resulting from the strategy profile $F_\Gamma \circ F_\mathcal{M}$*

Proof. A but-for cause is a special case of the modified HP definition where $\mathbf{W} = \emptyset$. This statement is hence a special case of propositions 4.5 and 4.6. □

Example 4.8. In our running semi-automated vehicle example, both *ODS* and $\neg DA$ are but-for causes of $\neg Col$, there being no collision (in the causal setting that there is an obstacle and the human driver is not paying attention). In the case of *ODS* we can define F_{ODS} to be the strategy where the obstacle detection system will not pass on a signal to the driving assistant. If all other agents follow the causal strategy profile, they will reach state $q_{2,5}$. Indeed $Col \in \pi(q_{2,5})$. Similarly, in the case of $\neg DA$, we can define F_{DA} to be the strategy where the driving assistant does not brake. When the other agents follow $F_\mathcal{M}$, they will end up in $q_{2,7}$. In that state it is indeed true that $Col \in \pi(q_{2,7})$.

In this section we have shown how agent strategies in a causal CGS relate to the causal relations in the causal setting the causal CGS was based on. In order to do this, we have introduced the notion of a causal strategy profile, a strategy for the grand coalition that makes sure the agents do exactly those actions they would do if all relations in the causal model would be followed.

5 CONCLUSION AND DISCUSSION

This paper investigates the relation between two formalisms that can be used to model multi-agent systems: structural causal models as introduced by Pearl [21] and concurrent game structures. This is done by proposing a systematic way to translate structural causal models to the co-called causal CGS. In such a causal CGS, agents will get to take their actions at a point corresponding to their position in the structural causal model. The causal CGS is defined in such a way that the leaf-states correspond to interventions on the original structural causal model.

In this paper, we have used the variable levels as defined by Halpern to determine the position of the variables in the causal model [16]. However, we can use any function that maps the endogenous variables to the positive integers as long as the function assigns a lower rank to a variable than to its descendants. In general, there are multiple of these functions possible for a given structural causal model. The formal results of this paper will hold for all such functions, though the structure of the resulting causal CGS may change due to the specific function used.

We can also relax the assumption that each agent controls exactly one agent variable. We assumed this to simplify the presentation of the causal CGS, but it is not a strict requirement. In principle an agent could control several variables and perform multiple actions, at several time steps, in the causal CGS.

A limitation of our approach is that in general, we are only able to give a result for actual causes if we already know the witness. For but-for causes, we are able to use the agents' abilities in the causal CGS to determine the but-for causes, but in general, this is not possible. Another limitation is that the causal CGS is generated with respect to a specific causal setting, hence the results only apply to a single context. This means that if the context is uncertain, multiple causal CGS have to be made to evaluate all possible outcomes. However, it is possible that this problem can be solved by using a version of an epistemic CGS. This can be researched in the future.

So far, we have only looked at deterministic and recursive causal models to define the causal CGS. However, causal relations are often probabilistic and cyclic in many practical use cases. Modelling such cases requires probabilistic and non-recursive causal models to, for example, capture the mutual dependencies between agents. In order to deal with probabilities, we will have to either employ probabilistic CGS, or use another type of model (e.g. Markov games). Moreover, allowing cyclic dependencies would make the evaluation of the states difficult, as the variable values would depend on each other. We think that this could possibly be dealt with by adding a temporal component to the model, but this needs more research.

Another direction of future work would be to use this framework to compare different approaches to defining responsibility in multi-agent settings. Some existing works define responsibility based on causal relations between agents and an outcome (like [1, 10, 11] and [6]), while other work is based on whether agents had a strategy to avoid the outcome (like [4] and [23]). The definition of causal CGS might help to combine both directions of research. Moreover, we can also look at how our approach compares to rule-based approaches to causality. Since Lorini's [20] work shows a correspondence between his rule-based framework for causal reasoning and the structural equations framework, it seems possible that his framework can also be shown to have a connection to our causal CGS.

This research could be used in multi-agent systems with a clear causal structure. Examples of this are traffic control environments, like planes that cannot land when another is departing, trains that cannot travel over the same track at the same time, or traffic lights on a junction that cannot all turn to green at the same time. Other applications could be in the analysis of multi-player games, after all, players could cause other players to make a certain move, or even energy management systems, where supply and demand of electricity influence each other. In these situations this research could be used to help making decisions, or after something has gone wrong to help attributing responsibility for this.

ACKNOWLEDGMENTS

This publication is part of the CAUSES project (KIVI.2019.004) of the research programme Responsible Use of Artificial Intelligence which is financed by the Dutch Research Council (NWO) and ProRail.

REFERENCES

[1] Natasha Alechina, Joseph Y Halpern, and Brian Logan. 2017. Causality, Responsibility and Blame in Team Plans. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. ACM, 1091–1099.

[2] Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman. 2002. Alternating-time temporal logic. *J. ACM* 49, 5 (Sept. 2002), 672–713. https://doi.org/10.1145/585265.585270

[3] Christel Baier, Florian Funke, and Rupak Majumdar. 2021. A Game-Theoretic Account of Responsibility Allocation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, Zhi-Hua Zhou (Ed.). International Joint Conferences on Artificial Intelligence Organization, test, 1773–1779. https://doi.org/10.24963/ijcai.2021/244 Main Track.

[4] Christel Baier, Florian Funke, and Rupak Majumdar. 2021. *A Game-Theoretic Account of Responsibility Allocation*. Technical Report arXiv:2105.09129.

[5] Christel Baier and Joost-Pieter Katoen. 2008. *Principles of model checking*. MIT press.

[6] Sander Beckers. 2023. Moral Responsibility for AI Systems. In *Advances in Neural Information Processing Systems*, A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Eds.), Vol. 36. Curran Associates, Inc., 4295–4308.

[7] Sander Beckers and Joost Vennekens. 2018. A principled approach to defining actual causation. *Synthese* 195, 2 (Feb. 2018), 835–862. https://doi.org/10.1007/s11229-016-1247-1

[8] Alexander Bochman. 2018. Actual Causality in a Logical Setting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 1730–1736. https://doi.org/10.24963/ijcai.2018/239

[9] Thomas Brihaye, Arnaud Da Costa, François Laroussinie, and Nicolas Markey. 2009. ATL with Strategy Contexts and Bounded Memory. In *Logical Foundations of Computer Science (Lecture Notes in Computer Science, Vol. 5407)*, Sergei Artemov and Anil Nerode (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 92–106.

[10] Hana Chockler and Joseph Y Halpern. 2004. Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research* 22 (2004), 93–115.

[11] Meir Friedenberg and Joseph Y. Halpern. 2019. Blameworthiness in Multi-Agent Settings. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (July 2019), 525–532. https://doi.org/10.1609/aaai.v33i01.3301525

[12] Maksim Gladyshev, Natasha Alechina, Mehdi Dastani, and Dragan Doder. 2023. Dynamics of Causal Dependencies in Multi-agent Settings. In *Engineering Multi-Agent Systems*, Andrei Ciortea, Mehdi Dastani, and Jieting Luo (Eds.). Springer Nature Switzerland, Cham, 95–112.

[13] Maksim Gladyshev, Natasha Alechina, Mehdi Dastani, Dragan Doder, and Brian Logan. 2023. Dynamic Causality. In *Proceedings of the 26th European Conference on Artificial Intelligence*. 867–874. https://doi.org/10.3233/FAIA230355

[14] Roberto Gorrieri. 2017. *Process algebras for Petri nets: the alphabetization of distributed systems*. Springer.

[15] N. Hall. 2007. Structural equations and causation. *Philosophical Studies* 132, 1 (Jan. 2007), 109–136. https://doi.org/10.1007/s11098-006-9057-9

[16] Joseph Y Halpern. 2016. *Actual causality*. MIT Press.

[17] Joseph Y. Halpern and Judea Pearl. 2005. Causes and Explanations: A Structural-Model Approach. Part I: Causes. *The British Journal for the Philosophy of Science* 56, 4 (2005), 843–887. https://doi.org/10.1093/bjps/axi147

[18] Lewis Hammond, James Fox, Tom Everitt, Ryan Carey, Alessandro Abate, and Michael Wooldridge. 2023. Reasoning about Causality in Games. *Artificial Intelligence* 320 (July 2023), 103919. https://doi.org/10.1016/j.artint.2023.103919

[19] Sylvia S. Kerkhove, Natasha Alechina, and Mehdi Dastani. 2025. Causes and Strategies in Multiagent Systems. arXiv:2502.13701 [cs.AI] https://arxiv.org/abs/2502.13701

[20] Emiliano Lorini. 2023. A Rule-Based Modal View of Causal Reasoning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, Edith Elkind (Ed.). International Joint Conferences on Artificial Intelligence Organization, 3286–3295. https://doi.org/10.24963/ijcai.2023/366 Main Track.

[21] Judea Pearl. 1995. Causal diagrams for empirical research. *Biometrika* 82, 4 (1995), 669–688.

[22] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. 2016. *Causal inference in statistics: A primer*. John Wiley & Sons.

[23] Vahid Yazdanpanah, Mehdi Dastani, Natasha Alechina, Brian Logan, and Wojciech Jamroga. 2019. Strategic responsibility under imperfect information. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems AAMAS 2019*. IFAAMAS, 592–600.