A Transition System for Causality and Strategic Responsibility

Sylvia S. Kerkhove

Utrecht University, Utrecht

This is an extended abstract of my thesis for the master Artificial Intelligence at the University of Utrecht, supervised by Prof. dr. M.M. (Mehdi) Dastani with Dr. N.A. (Natasha) Alechina as the second examiner [5].

In his book Actual Causality, Halpern claims that determining causality is crucial in the attribution of responsibility for an outcome [4]. Other works use responsibility to define agent strategies [6,2]. Agent strategies are usually discussed in the context of labelled transition systems, or more general, concurrent game structures (CGS), which generalise LTS to multi-agent systems [1]. Just like these CGS, a model for actual causality can also be represented as a graph, called a causal network [4]. The main difference between causal networks and LTS is that nodes in LTS represent states of the environment, with the edges representing events that change this state, while in a causal model the nodes represent variables that can have a certain value in an environment, with the edges representing the causal effect the variables have on one another.

Causality plays an important role in many daily processes, despite this, there has not been a lot of research into causality in (multi-)agent systems. This work aims to address this problem by integrating causal models with CGS as used for (multi-)agent systems. To generate a causal CGS from a structural causal model the causal variables are partitioned in a set of agent variables that are directly controlled by an agent, and a set of environment variables that are not directly controlled by an agent, this is similar to the approach of Gladyshev et al. [3]. Causal variables are also defined to have a rank. A ranking function can be any function that maps the causal variables to the positive integers, representing the rank of the variable, in such a way that any descendant of a variable X gets a higher rank than the rank of X. An agent ranking function then 'compresses' the ranking function to only have as many values as there are agent variables with distinct ranks. This is used to determine which variables will be updated in which states of the causal CGS. The causal CGS starts from a starting state, that is defined to correspond to a causal setting while every agent action will be reminiscent of an intervention on this causal setting. In this starting state agents with the lowest rank get to take actions, this leads to new states, in which all the agents with the next lowest rank get to take actions, and so on until all agents have taken an action. The actions for an agent correspond to the possible values of the agent variable in the structural causal model. This gives a tree-structure for the causal CGS. Not all variables get evaluated in every state, in principle, all variables will only be evaluated once on a path through the causal CGS and

in every state, only the agent variables of the agents that just took an action and all environment variables that depend on those variables and not on higher ranked agent variables will be evaluated. Because of this only the leaf states will fully correspond to interventions on the original causal setting.

See Figure 1 for an illustration of a possible causal CGS that can be generated from the structural causal model for the rock-throwing example. In this example two agents, Billy and Suzy, are throwing rocks at a bottle. Suzy throws faster than Billy, so if they both throw, Suzy's rock is the one to shatter the bottle. The causal model has variables Suzy throws, ST, Billy throws, BT, Suzy's rock hits the bottle, SH, Billy's rock hits, BH, and the bottle shatters, BS. The model is constructed in such a way that if Suzy throws, she hits, and that if Billy then also throws he does not hit. All variables can have value 0 or 1.

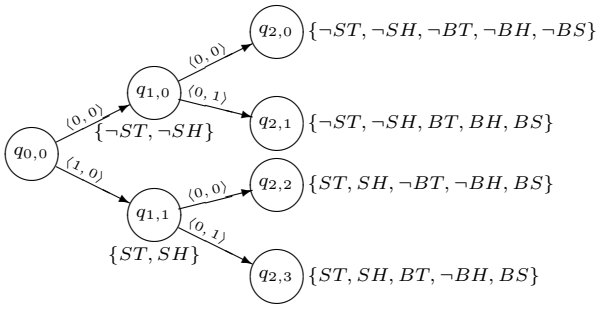

Fig. 1: A possible causal CGS of the rock-throwing example given a ranking function. The initial values in the starting state are not shown. In the middle states only the variables that were changed in that state are shown.

Given a causal model and the generated causal CGS, we have the following result. Let both $\mathbf{X}$ and $\mathbf{W}$ only contain agent variables, $\mathbf{X} = \mathbf{x}$ is a cause of φ with witness $\mathbf{W} = \mathbf{w}^*$ according to the modified HP definition of causality [4], if and only if the set of agents corresponding to the agent variables in $\mathbf{X}$ and $\mathbf{W}$ has a strategy such that $\neg\varphi$ will hold in the leaf-state that results from these agents following this strategy and the other agents following a strategy that gives the actions they would have taken according to the causal model.

A limitation of this approach is that since the agent actions are seen as interventions, not all causal relations are carried over in to the causal CGS. Another limitation is that the causal CGS is generated with respect to a specific causal setting, if the context is uncertain, multiple causal CGS have to be made to evaluate all possible outcomes.

This research could be used in multi-agent systems with a clear causal structure or in the analysis of multi-player games, after all, players could cause other players to make a certain move. In these situations this research could be used to help making decisions, or after something has gone wrong to help attributing responsibility for this.

References

1. Alur, R., Henzinger, T.A., Kupferman, O.: Alternating-time temporal logic. Journal of the ACM (JACM) **49**(5), 672–713 (2002)
2. Baier, C., Funke, F., Majumdar, R.: A game-theoretic account of responsibility allocation. In: Zhou, Z.H. (ed.) Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21. pp. 1773–1779. International Joint Conferences on Artificial Intelligence Organization (8 2021). https://doi.org/10.24963/ijcai.2021/244, https://doi.org/10.24963/ijcai.2021/244, main Track
3. Gladyshev, M., Alechina, N., Dastani, M., Doder, D.: Dynamics of causal dependencies in multi-agent settings. In: Ciortea, A., Dastani, M., Luo, J. (eds.) Engineering Multi-Agent Systems. pp. 95–112. Springer Nature Switzerland, Cham (2023)
4. Halpern, J.Y.: Actual causality. MiT Press (2016)
5. Kerkhove, S.S.: A Transistion System for Causality and Strategic Responsiblity. Master's thesis, Utrecht University (2023)
6. Yazdanpanah, V., Dastani, M., Alechina, N., Logan, B., Jamroga, W.: Strategic responsibility under imperfect information. In: Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems AAMAS 2019. pp. 592–600. IFAAMAS (2019)