Causation and Accountability in Dynamic Action Logic

Causal reasoning is pervasive in the law. Criminal offenses are typically studied in terms of the defendant's action (*actus reus*) and intentional state at the time the action is performed (*mens rea*). Both elements are understood in respect to causality [4, 7]: as to the *actus reus*, we assess whether the defendant's conduct caused a certain state of affairs; as to the *mens rea*, we assess whether appropriate epistemic states played a causal role in qualifying the action. In addition, we consider whether the action was just a *factual cause* of the result (intuitively, something that, in the specific circumstances, had the result as a consequence), or also a *legal cause* of it (roughly, something that, in the specific circumstances, played a substantial role in bringing it about). We thus want to answer two main questions about the agents interacting in a situation leading to a negative result φ:

Q1. Did agent i causally contribute to φ?

Q2. If so, was i also accountable for φ?

In the last decade, several deontic and action logics have been advanced to clarify the notion of *mens rea* [2, 3, 8, 9], and its connection with legal responsibility. However, less attention has been payed to the notion of *actus reus*, and specifically to the distinction between *causally contributing* to a state of affairs and *being accountable* for it.

This work aims at filling this gap by presenting a logic where a procedure to identify an *actus reus* in multi-agent scenarios is properly modelled. Our contribution is twofold: *from a logical perspective*, we implement a causal theory in a multi-agent action logic; *from a philosophical perspective*, we show how integrating this theory in a dynamic setting, where actions are interpreted as transitions, improves our understanding of the distinction between causation and accountability.

1. Causality in multi-agent action logic

Answering **Q1** requires a proper theory of causality. In the legal literature the limitations of the standard *but-for* criterion are addressed by using the more flexible INUS or NESS accounts [10, 11]. Our first contribution is to implement in a multi-agent action logic a version of the INUS account [1] that integrates suggestions from the structural equation approach [6], thus qualifying as the currently most advanced account of causality consistent with the legal literature.

The basic idea is that agent i causally contributes to φ *via* her action α_i at world w iff there are agents $j_1, \ldots, j_k$ and actions $\beta_{j_1}, \ldots, \beta_{j_k}$ such that:

1. the composite action $\alpha_i\beta_{j_1}...\beta_{j_k}$ has been performed at w;
2. $\alpha_i\beta_{j_1}...\beta_{j_k}$ ensures that φ obtains at w;
3. $\alpha_i\beta_{j_1}...\beta_{j_k}$ minus any of its components does not ensure that φ obtains at w;
4. $\alpha_i\beta_{j_1}...\beta_{j_k}$ is part of a disjunctive necessary condition for φ.

In short, α_i causally contributes to φ at w iff there are agents $j_1, \ldots, j_k$ and actions $\beta_{j_1}...\beta_{j_k}$ such that $\alpha_i\beta_{j_1}...\beta_{j_k}$ is an actual minimal sufficient condition for φ in a context satisfying the *but-for* criterion at w.

2. Accountability in a dynamic setting

Suppose now that agent 1 stabs a man (α_1), agent 2 gives the man inappropriate medical treatment (β_2), and the man dies (φ). Assume that all factors determining φ are taken into account and that appropriate treatment could have healed the man. According to the refined INUS test, α_1 and α_2 are on a par in causally contributing to φ, even if, from a legal perspective, 2's negligence might exonerate 1 from guilt. This brings us from question **Q1** to **Q2**: both agents causally contribute to φ, but who is accountable for φ? Criminal cases of this sort are usually decided on the basis of the principle

P1. Complications occurring in the hospital should be seen as background circumstances, unless medical treatment is so "palpably wrong" that it breaks the causal chain initiated by the defendant (cf. *R v Jordan* (1956)).

Since cases of negligence are structurally similar to many criminal cases creating difficulties for the INUS test, our second contribution will be to introduce an improved test enabling us to treat these cases in a uniform way. To do this, we

generalize **P1** by interpreting it in a *dynamic* multi-agent action logic equipped with a distinction between default and deviant transitions [5].

Here's the idea. Suppose that 1 executes α_1 in the initial state w_1; 2 intervenes with medical treatment in the next state w_2; φ happens in the final state w_3. Assume that two conducts are available to 2 at w_2: a *default conduct* of intervening with regular treatment (β_2^+) and a *deviant conduct* of intervening negligently (β_2^-). Finally, assume that we can figure out what would happen if some agents was removed from the scene (performing the $skip$ action). To answer **Q2**, we then run the following test.

Step 1. Check that α_1 causally contributes to φ in the scenario where 2 performs $skip$ at w_2.

Step 2. Determine whether α_1 is followed by a default or a deviant conduct of 2.

If 2 does β_2^+ at w_2, the test ends: only 1 is accountable for φ: subsequent default conducts only contribute to φ as background conditions.

Otherwise, continue.

Step 3. Compare the possible outcomes of doing β_2^- and β_2^+ at w_2.

Case 1. By doing β_2^- rather than β_2^+, 2 either (a) eliminates the only possibility to avoid φ or (b) does not eliminate a risk she could have eliminated. *Result*: 2 is accountable, but 1 is not discharged: 1 either (a) makes it impossible for 2 to prevent φ or (b) creates the risk in the first place.

Case 2. By doing β_2^- rather than β_2^+, 2 ensures φ rather than preventing it. *Result*: 2 is accountable and 1 is discharged: by ensuring a result a default conduct would have surely prevented, β_2^- supersedes α_1 at w_2.

Intuitively, step 1 identifies the operating causes at the initial state w_1; step 2 checks whether some of the operating causes at w_2 stand out from the background; if so, step 3 determines whether these add up to or replace the previous operating causes. While operating causes are identified by using the INUS test, background conditions are determined in terms of the distinction between default and deviant conducts. Finally, the notion of "palpably wrong", discriminating between additional and superseding causes, is analyzed in terms of a comparison between transitions instantiating default conducts and transitions instantiating deviant conducts.

By exploiting the sketched procedure, a precise notion of actus reus can be defined in a dynamic framework equipped with a notion of causal contribution and a distinction between default and deviant conducts. In this work, we will present

such framework and investigate to what extent the procedure constitutes a general method to identify guilty acts in multi-agent scenarios.

Word Count: 998

References

[1] Baumgartner, M. (2013). A Regularity Theoretic Approach to Actual Causation. *Erkenntnis*, 78: 85-109.

[2] Broersen, J. (2011). Deontic Epistemic Stit Logic Distinguishing Modes of Mens Rea. *Journal of Applied Logic*, 9(2): 137–152.

[3] Broersen, J.. (2011). Making a Start with the Stit Logic Analysis of Intentional Action. *Journal of Philosophical Logic*, 40: 499–530.

[4] P. Fletcher, G.P. (1998). *Basic Concepts of Criminal Law*. Oxford University Press.

[5] Halpern, J.Y. & Hitchcock, C. 2015. Graded Causation and Defaults. *The British Journal for the Philosophy of Science*, 66 (2): 413-457.

[6] Halpern, J.Y. & Pearl, J. (2005). Causes and Explanations: A Structural-Model Approach. Part I: Causes. *The British Journal for the Philosophy of Science*, 56 (4): 843-887.

[7] Lagnado, D. A. & Gerstenberg, T. (2017). Causation in Legal and Moral Reasoning. In M. R. Waldmann (ed.)*, Oxford Handbook of Legal Reasoning*, Oxford University Press: 565-602.

[8] De Lima, T., Royakkers, L. , & Dignum, F. (2010). A Logic for Reasoning about Responsibility. *Logic Journal of IGPL*, 18: 99–117.

[9] Lorini, E., Longin, D., & Mayor, E. (2013) A Logical Analysis of Responsibility Attribution: Emotions, Individuals and Collectives. *Journal of Logic and Computation*, 24(6): 1313-1339.

[10] Mackie, J.L. (1974). *The Cement of the Universe: A Study of Causation*. Oxford University Press.

[11] Wright, R.W. (2011). The NESS account of Natural Causation: a Response to Criticisms. In R. Goldberg (Ed.), *Perspectives on Causation*, Oxford: Hart.