Action Languages Based Actual Causality in Decision Making Contexts

Camilo Sarmiento[1] $^{[0000-0003-4581-051X]}$, Gauvain Bourgne[1] $^{[0000-0002-5104-9352]}$, Katsumi Inoue[2] $^{[0000-0002-2717-9122]}$, and Jean-Gabriel Ganascia[1] $^{[0000-0002-8181-1379]}$

[1] Sorbonne Université, CNRS, LIP6, F-75005 Paris, France,
camilo.sarmiento@lip6.fr, gauvain.bourgne@lip6.fr,
jean-gabriel.ganascia@lip6.fr,
[2] National Institute of Informatics, Tokyo, Japan,
inoue@nii.ac.jp

Abstract. Rationally understanding the evolution of the physical world is inherently linked with the idea of causality. It follows that agents based on automated planning have inevitably to deal with causality, especially when considering imputability. However, the many debates around causation in the last decades have shown how complex this notion is and thus, how difficult it is to integrate it with planning. This paper's contribution is to link up two research topics—automated planning and causality—by proposing an actual causation definition suitable for action languages. This definition is a formalisation of Wright's NESS test of causation.

Keywords: Causality, Actual Causality, Regularity Theories of Causation, Action Languages

1 Introduction

Because of its essential role in human reasoning—both in trivial and in complex situations—numerous works in a variety of disciplines have tried—unsuccessfully—to propose a widely agreed upon theory of causation. The purpose of this article is not to enter into the debates that animate the community, but to propose a definition of causality that can be used by agents to enrich their knowledge of the world evolution and thus make better decisions. Since we are in an operational framework given that our focus is on decision making, we can make a couple of assumptions while remaining relevant. Therefore, we place ourselves in a classical planning framework which assumes problems are discrete and deterministic. Unlike *type causality* which seeks to determine general causal relationships, *actual causality* fits our purpose because it is concerned with particular events [18]. Limiting ourselves to a simplified framework and to actual causality does not make causality trivial, many issues remain.

Recent works [4, 8, 21, 24] have attempted to link action languages and causation. However, each work has its own limitations. Thus, our work is a logical continuation of those mentioned above and our aim is to address the main remaining limitations: (i) a definition of actual causality conflating causality and responsibility, and (ii) a framework leaving aside the cases of overdetermination—subject of many crucial debates in the field of causality—by their inability to deal with concurrency of events. Overdetermined causation was defined by Wright [36] as:

> cases in which a factor other than the specified act would have been sufficient to produce the injury in the absence of the specified act, but its effects either (1) were preempted by the more immediately operative effects of the specified act or (2) combined with or duplicated those of the specified act to jointly produce the injury.

We deal with the first problem by imposing ourselves the constraints of factuality and independence of policy choices defended by Wright [36]. Regarding the second issue, the solution lies in the essential choice of the formalism encoding causal knowledge. The advanced state of maturity of PDDL [15, 20], its vocation to facilitate interchangeability, and its use by a large community, are all meaningful arguments in favour of this formalism. However, the semantics of its ADL [33] fragment does not allow concurrency of events. To have a semantics that takes into account the concurrency it is necessary to jump directly to PDDL+ [12] whose semantics is adapted to durative actions, thus inconsistent with our discrete time assumption. We therefore base our approach on an action language whose semantics is an intermediate point between the ADL fragment of PDDL and PDDL+.

This paper is structured as follows. Section 2 discusses what is the appropriate approach to causation for our causal inquiry. In this section we explore two main highly influential theories of causation: *regularity* and *counterfactual*. Section 3 introduces the action language semantics—in which we encode causal knowledge—allowing concurrency of events. Section 4 offers a description of our actual causality definition proposal that we compare to Batusov and Soutchanski's approach [4]—the one we consider to be the most convincing so far. Finally, we conclude and give some perspectives in Section 5.

2 Adapted Causal Inquiry

Of the many fields studying causality, our approach is especially close to *tort law* whose interest is about causation in specific situations. Hence, works in this field are a good source of inspiration. In a series of influential papers [36, 37], Wright demonstrates how essential a causal inquiry is in the process of determining *tort liability*. He emphasises the fundamental difference between causation and responsibility—or in the words of Vincent's taxonomy [35], between 'causal responsibility' and 'outcome responsibility'.

Wright argues that a satisfying tort liability analysis—which goal is to determine if a defendant is the 'responsible cause' of an injury—requires a factual and independent of policy choices causal inquiry. In his papers Wright criticises the processes to determine responsibility for an injury in which the causal inquiry is flawed and polluted with subjective aspects—a process where causality and responsibility are conflated. Wright's initial observation is that those two notions are too often conflated. The fact that 'the phrase "*the cause*" is simply an elliptical way of saying "*the responsible cause*"' [36] shows how thin the boundary between those notions is. To clarify this conflation, he describes the process to determine if an individual is legally responsible for an injury. This process has three stages: (i) *tortious-conduct inquiry*, where are identified the defendant's conducts that could potentially imply legal responsibility (intentional, negligent, hazardous, . . .); (ii) *causal inquiry*, where is evaluated if the identified tortious conducts really contributed to cause the harm, i.e. if they can be considered as causes of the injury; (iii) *proximate-cause inquiry*, where other causes of the injury are considered, so as to evaluate if they mitigate or eliminate the defendant's legal responsibility for the injury. Of those three stages, only the second is entirely factual and independent of policy choices. It determines if a conduct was a cause of the injury. The two others are subject to policy considerations that 'determine which causes and consequences will give rise to liability' [36]. Not to yield into the easy confusion between responsibility and causality, our goal is to propose a definition of actual causality suitable for a causal inquiry as presented by Wright, i.e. factual and independent of policy choices.

The actual causation definitions based solely on strong necessity—also known as counterfactual dependence—fail to capture the commonly accepted intuition on overdetermination cases (early preemption, late preemption, and symmetric overdetermination) [17, 28]. The commonly used in law *But-for test* is one of those unsuccessful definitions. This test states that 'an act was a cause of an injury if and only if, but for the act, the injury would not have occurred' [36]. 'In the context of structural equations, this flawed account can be described as equating causation with counterfactual dependence' [6]. Given that overdetermination cases are not just hypothetical and rare cases (cases of pollution, suicide, economic loss, . . .), those strong necessity based approaches are not suitable for our purposes.

The dominant approach of actual causality—HP definition [18]—deals with those cases, but at the cost of the factualness of the causal inquiry. This definition has the same roots than the But-for test, Hume's definition of causation second formulation [22]:

> we may define a cause to be an object followed by another, and where all objects, similar to the first, are followed by objects similar to the second. Or, in other words, where if the first object had not been, the second had never existed.

It is the result of an iterative process that originates in Pearl's formalisation of Lewis' vision [25] in structural equations framework (SEF) [32]. HP approach is more complete than the But-for test in the sense that other elements in addition to counterfactual dependence where included in order to deal with some complex cases. One of those elements is interventionism. This assumption states that an event C causes a second event E if and only if, both events occur, and that, given an intervention allowing to fix the occurrence of a certain set of other events in the context—without being constrained to respect the physical coherence of the world—there is a context where if the first event had not occurred, the second would not have occurred either. This assumption is described by Beckers [6] using SEF notation as:

> **Interventionism** They all share the assumption [HP-style definitions] that the relation between counterfactual dependence and causation takes on the following form: $C = c$ causes $E = e$ iff $E = e$ is counterfactually dependent on $C = c$ given an intervention $\vec{X} \leftarrow x$ that satisfies some conditions P. The divergence between these definitions is to be found in the condition P that should be satisfied.

Interventionism—that Beckers' CNESS [6] and Beckers and Vennekens' BV [7] definitions reject—introduces non factual elements to the causal inquiry which appear problematic even for the author [19]:

> if I fix BH [Billy hits] to zero here, I am sort of violating the way the world works. [...] I am contemplating counterfactuals are inconsistent with the equations but I seem to need to do that in order to get things to work out right. Believe me, we tried many other definitions.

In addition to non factual elements, the divergence on which 'conditions P' to apply can be equated with policy choices. These elements make HP-style definitions non adequate for our context.

STIT approaches are also part of this family where strong necessity is central. Usual STIT approaches focus on the relationship between the agent and the states of the world. In order to be closer to the philosophical tradition according to which the actual causal relationship is defined between two events, we find action languages ideal because the events are central elements. Because of their modal approach, STIT works such as [1, 26] easily involve epistemic aspects—outside of the scope of this paper—fundamental when one wishes to go beyond causality by looking at responsibility.

The NESS test which subordinates necessity to sufficiency is an approach that deals with overdetermination cases [5, 36–38] and that satisfies our inquiry needs. Introduced by Wright in response to But-for test flaws, this test states that [36, 37]:

> A particular condition was a cause of a specific consequence if and only if it was a necessary element of a set of antecedent actual conditions that was sufficient for the occurrence of the consequence.

Unlike approaches mentioned above, it belongs to a second high impact approach family, regularity theories of causation [2]. Those theories are also based on Hume's definition of causation, but on the first formulation. Specifically, the NESS test is closer to Mill's interpretation of this formulation which introduced that there are potentially a multiplicity of distinct, but equally sufficient sets of conditions [30]. The NESS test is even closer to Mackie's proposal. Indeed, unlike Mill's vision whereby the cause is the sufficient set, Mackie considers that each element of the set is a cause [27].

The actual causation definition we propose is an action languages suitable formalisation of Wright's NESS test. Even if accepted by influential counterfactual theories of causation authors as embodying our basic intuition of causation—such as Pearl [6]—criticism of the use of logic as formalism has prevented the popularisation of this test. What is argued is the inadequacy of logical sufficiency and logical necessity to formalise these intuitions. Recent works have shown that rejecting the formalism is not a reason to reject the idea behind it by successfully formalising the NESS test in causal calculus [9] and in the structural equations framework [6]. It is conceivable to work on a way of compiling existing action languages problems and translating them into SEF. However, works have shown SEF flaws [10] and that in complex evolving contexts [4, 21] like ours, this translating approach is not necessarily desirable [21]:

> Structural causal models are excellent tools for many types of causality-related questions. Nevertheless, their limited expressivity render them less than ideal for some of the more delicate causal queries, like actual causation. These queries require a language that is suited for dealing with complex, dynamically changing situations.

Our contribution is to link automated planning and causality by continuing this momentum proposing an action languages suitable formalisation of Wright's NESS test.

3 Action Language Semantics

The whole purpose of an action language is to determine the evolution of the world given a set of actions corresponding to deliberate choices of the agent. Those actions might trigger some chain reaction through external events. As a result, we need to keep track of both: the state of the world and the occurrence of events—the term 'event' connoting 'the possibility of agentless actions' [34, chap 12]. This task is the simplest kind of temporal reasoning—temporal projection. Different action languages allowing temporal projection have been proposed such as PDDL [15, 20] and action description languages $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ [14]. However, the semantics of $\mathcal{A}$ [13], $\mathcal{B}$, and PDDL deterministic fragment—corresponding to ADL [33]—do not allow concurrency of events. To have a semantics that takes into account concurrency it is necessary to jump directly to $\mathcal{C}$ [16] or PDDL+ [12] which semantics is adapted respectively to non deterministic actions or durative actions, thus inconsistent with either our deterministic actions assumption or our discrete time assumption. The advanced state of maturity of PDDL [15, 20], its vocation to facilitate interchangeability, and its use by a large community, are all meaningful arguments in favour of this formalism—gradually extended by different fragments. We therefore base our approach on an action language whose semantics is an intermediate point between the deterministic fragment of PDDL and PDDL+. This formalism works on a decomposition of the world into two sets: $\mathbb{F}$ corresponding to variables describing the state of the world, more precisely ground fluents representing time-varying properties; $\mathbb{E}$ representing variables describing transitions, more precisely ground events that modify fluents.

A fluent literal is either a fluent $f \in \mathbb{F}$, or its negation $\neg f$. We denote by $Lit_{\mathbb{F}}$ the set of fluent literals in $\mathbb{F}$, where $Lit_{\mathbb{F}} = \mathbb{F} \cup \{\neg f | f \in \mathbb{F}\}$. The complement of a fluent literal l is defined as $\bar{l} = \neg f$ if $l = f$ or $\bar{l} = f$ if $l = \neg f$. By extension, for a set $L \subseteq Lit_{\mathbb{F}}$, we have $\overline{L} = \{\bar{l}, l \in L\}$.

Definition 1 (state S). *The set $L \subseteq Lit_{\mathbb{F}}$ is a state iff it is:*

- *Coherent:* $\forall l \in L, \bar{l} \notin L$.
- *Complete:* $|L| = |\mathbb{F}|$, *i.e.* $\forall f \in \mathbb{F}, f \in L$ *or* $\neg f \in L$.

A complete and coherent set of fluent literals thus determines the value of each of the fluents. An incoherent set cannot describe a reality. However, in the absence of information or for the sake of simplification, we can describe a problem through a coherent but incomplete set. We will call such a set a partial state. We model time linearly and in a discretised way to associate a state S_t to each time point t of a set $\mathbb{T} = \{-1, 0, \ldots, N\}$. Having a bounded past formalisation of a real problem, we gather all states before $t = 0$—time point to which corresponds the state S_0 that we call initial state—in an empty state $S_{-1} = \varnothing$.

We place ourselves in a framework of concurrency where E_t is the set of all events which occur at a time point t. Therefore, E_t is what generates the transition between the states S_t and S_{t+1}. Thus, the states follow one another as events occur, simulating the evolution of the world. E_{-1} is the set that gathers all events which took place before $t = 0$, such that $E_{-1} = \{ini_l, l \in S_0\}$. Events are characterised by two elements: preconditions give the conditions that must be satisfied by the state in order for them to take place; effects indicate the changes to the fluents that are expected to happen if they occur. The preconditions and effects are respectively represented as formulas of the language $\mathcal{P}$ and $\mathcal{E}$ defined as follows:

$$\mathcal{P} ::= l | \psi_1 \wedge \psi_2 | \psi_1 \vee \psi_2 \qquad \mathcal{E} ::= [\psi]l | \varphi_1 \wedge \varphi_2$$

where $l \in Lit_{\mathbb{F}}$, $[\psi]l$ is the notation for the conditional effect indicating that l is an effect if the condition ψ is satisfied—$[\top]l$ is just written l—and the logical connectives $\wedge$, $\vee$ have standard first-order semantics. We can then deduce that if $\varphi \in \mathcal{E},\ \varphi = \bigwedge_{i \in 1,\ldots,m} [\psi_i] l_i$. For the sake of brevity, we adopt a set notation for $\varphi \in \mathcal{E}$ which we will use where relevant, such that $\varphi = \{[\psi_i]l_i,\ i \in 1, \ldots, m\}$. We denote pre and eff the functions which respectively associate preconditions and effects with each event: $pre : \mathbb{E} \mapsto \mathcal{P}$, $eff : \mathbb{E} \mapsto \mathcal{E}$. Given the expression of E_{-1}, the application of eff to each element of the set is $eff(ini_l) = l$ with $l \in S_0$, thus $eff(E_{-1}) = S_0$. Moreover, given a formula $\psi \in \mathcal{P}$ and a partial state L, $L \vDash \psi$ is defined classically: $L \vDash l$ if $l \in L$, $L \vDash \psi_1 \wedge \psi_2$ if $L \vDash \psi_1$ and $L \vDash \psi_2$, and $L \vDash \psi_1 \vee \psi_2$ if $L \vDash \psi_1$ or $L \vDash \psi_2$.

Our work is a logical continuation of works such as [4, 8, 21, 24], who attempted to link action languages and causation. To the best of our knowledge, [4, 8] are the first to give a definition of actual cause in action languages. However, each work has its own limitations that we try to address. In Batusov and Soutchanski's paper, many working perspectives are mentioned [4]:

> It is clear that a broader definition of actual cause requires more expressive action theories that can model not only sequences of actions, but can also include explicit time and concurrent actions. Only after that one can try to analyze some of the popular examples of actual causation formulated in philosophical literature. Some of those examples sound deceptively simple, but faithful modelling of them requires time, concurrency and natural actions.

At the moment, the proposed action language tackles both concurrency and time—at least discrete time. We will now introduce 'natural actions' that we denote exogenous events. These events are what distinguish our proposal from $\mathcal{A}_c$ [3]—the allowing concurrency version of $\mathcal{A}$. The set $\mathbb{E}$ is divided into two subsets: $\mathbb{A}$, which contains the actions carried out by an agent and thus subjected to a volition; $\mathbb{U}$, which contains the exogenous events—equivalent to `:event` in PDDL+ [12] and triggered axioms in Event Calculus [31]—which are triggered as soon as all the *pre* are fulfilled, therefore without the need for an agent to perform them. Thus, for exogenous events triggering conditions and preconditions are the same. In contrast, the triggering conditions for actions necessarily include preconditions but those are not sufficient. The triggering conditions of an action also include the volition of the agent or some kind of manipulation by another agent. To keep track of these subtleties that could be relevant in the causal inquiry we introduce triggering conditions represented as formulas of the language $\mathcal{P}$. We denote *tri* the function which associates triggering conditions with each event: $tri : \mathbb{E} \mapsto \mathcal{P}$.

The occurrence of events $(e, t) \in \mathbb{E} \times \mathbb{T}$ and $(e', t) \in \mathbb{E} \times \mathbb{T}$ in the state S_t is said to be interfering if the set $\{l, \exists\psi \in \mathcal{P}, S_t \vDash \psi, [\psi]l \in eff(e) \cup eff(e')\}$ is not coherent according to Definition 1.

Definition 2 (context κ). *Given an initial state S_0, the context denoted as κ is the octuple $(\mathbb{E}, \mathbb{F}, pre, tri, eff, S_0, >, \mathbb{T})$, where $>$ is a partial order which represents priorities that ensure the primacy of one event over another when both are interfering.*

As mentioned earlier, effects indicate the changes to the fluents that are expected to happen if an event occurs. Because of the complexity of reality, it may turn out that causally the action has more or less effects than those attributed by $\mathcal{E}$. Let's take the example of an agent who wants to turn on a light by pressing a switch. In a first scenario, it is possible that the agent's action causes an overheating in the electrical circuit and triggers a fire. When formalising the action of switching on the light, besides that it is not intuitive to take into account the overheating and then the fire as intrinsic effects, it affects the generality of the formalisation. In these cases, we will prefer to break down the process by introducing exogenous events. In the above fire example, we will therefore have an exogenous event corresponding to a fire outbreak—an agentless event—which will be triggered when a defective circuit is present and the switch is pressed. We are therefore in the presence of a causal chain. These cases where the action has more effects than those with which it has been formalised are typical cases where causality is necessary. In a second scenario, it may happen that the agent performs the action but the expected effects are not produced simply because the light was already on. This does not prevent the action from having been performed, and we want to keep a trace of the event without having to consider that its effect has taken place. This is especially the case if the action has several effects and only one of them does not actually occur. This second case can be resumed as cases where some of the fluents of the state have already the value attributed by an effect. Since the effects that an event had at the time it occurred is a basic causal information on which we will rely—inextricably linked to imputability—it is important to keep track of them.

Definition 3 (actual effects $actualEff(E, L)$). *Given a context κ, the predicate $actualEff(E, L)$ which associates a set of events $E \in \mathbb{E}$ given a partial state L, to a partial state representing the actual effects of E when L is true, is defined as:*

$$\begin{aligned} actualEff(E, L) &= \bigcup_{e \in E} actualEff(\{e\}, L) \\ &= \{l_i, \exists e \in E, [\psi_i]l_i \in eff(e),\ L \vDash \psi_i,\ \textit{and } l_i \notin L\} \end{aligned}$$

For the sake of conciseness we adopt an update operator giving the resulting state when performing an event at a given state.

Definition 4 (update operator $\triangleright$). *Given a context κ and set of events $E \in \mathbb{E}$, the update operator which we use as follows $S_t \triangleright E$ expresses $S_t \setminus \overline{actualEff(E, S_t)}$ $\cup actualEff(E, S_t)$.*

The information given by $actualEff(E, L)$ and $\triangleright$ can be equated to basic causal information given by the evolution of the world. Besides being causal, this information is directional since it is inconceivable in our semantics to say that the actual effect of the event is the cause of it. Therefore, we can rely on the events that occur and their actual effects to simulate the evolution of the world from the initial state S_0.

Definition 5 (induced state sequence $\mathcal{S}_\kappa$). *Given a context κ and a sequence of events $\epsilon = E_{-1}, E_0, \ldots, E_n$, such that $n \leq |\mathbb{T}|$, the induced state sequence of ϵ is a sequence of complete states: $\mathcal{S}_\kappa(\epsilon) = S_0, S_1, \ldots, S_{n+1}$ such that $\forall t \in \{-1, \ldots, n\}$, $S_{t+1} = S_t \triangleright E_t$.*

Though this can be defined for every ϵ, not all ϵ are possible given (i) the need to satisfy preconditions, (ii) the concurrency of events that must respect priorities, and (iii) the triggering of events that must respect priorities too.

Definition 6. *Let ϵ be a sequence of events $\epsilon = E_{-1}, E_0, \ldots, E_n$, such that $n \leq |\mathbb{T}|$, and let's denote by $\mathcal{S}_\kappa(\epsilon) = S_0, S_1, \ldots, S_{n+1}$ its induced state sequence. We shall say that ϵ is:*

- *Executable in κ: if $\forall t \in \{0, \ldots, n\}$, $S_t \vDash pre(E_t)$.*
- *Concurrent correct with respect to κ: $\neg\exists(e, e') \in E_t^2,\ e > e'$.*
- *Trigger correct with respect to κ: if $\forall t \in \{0, \ldots, n\}$, $\forall e' \in \mathbb{E}$ such that $S_t \vDash tri(e')$, then $e' \in E_t$ or $\exists e \in E_t,\ e > e'$.*
- *Valid in κ: if and only if, executable in κ, concurrent correct with respect to κ, and trigger correct with respect to κ.*

Finally, if we consider only a set of timed actions as an input which we call scenario, we have:

Definition 7 (traces $\tau^e_{\sigma,\kappa}$ and $\tau^s_{\sigma,\kappa}$). *Given a scenario $\sigma \subseteq \mathbb{A} \times \mathbb{T}$ and a context κ, the event trace $\tau^e_{\sigma,\kappa}$ of σ, κ is the sequence of events $\epsilon = E_{-1}, E_0, \ldots, E_n$ valid in κ, such that: $\forall t \in \{0, \ldots, n\}, \forall e \in E_t,\ e \in \mathbb{A} \Leftrightarrow (e, t) \in \sigma$. Its induced state sequence is the state trace $\tau^s_{\sigma,\kappa}$.*

We now have a tool for temporal projection. Since in future work we plan to evaluate the ethical permissibility of actions in a given scenario, this is sufficient. However, given the flexibility of answer set programming in which we have translated our action language, we could move at low cost to a planning tool managing concurrency of events and exogenous events—the latter allowing to handle dynamic environments. For more complex contexts involving multiple agents, this action language only gives a partial solution. Indeed, the actions of other agents can be represented as exogenous events. However, this solution does not capture the full complexity of multiagent contexts. In future work we plan to study this issue, in particular by formalising a causal relationship specific to these contexts, 'enables' [8, 11].

4 Actual Causality

In the context of action languages, we consider that a first event is an actual cause of a second event if and only if the occurrence of the first is a NESS-cause of the triggering of the second. As commonly accepted by philosophers, the relation of causality we aim to define links two events. However, 'events are not the only things that can cause or be caused' [25]. Action languages represent the evolution of the world as a succession of states produced by the occurrence of events, thus introducing states between events. Therefore, we need to define causal relations where causes are occurrence of events and effects are formulas of the language $\mathcal{P}$ truthfulness. This section will introduce definitions which establish such a relation based on Wright's NESS test of causation.

Definition 8 (causal setting χ). *The action language causal setting denoted χ is the couple (σ, κ) with σ a scenario and κ a context.*

From now on, when reference is made to events and states, they will be those from $\tau^e_{\sigma,\kappa}$ and $\tau^s_{\sigma,\kappa}$ respectively. Thus, the set of all events which actually occurred at time point t is $E^\chi(t) = \tau^e_{\sigma,\kappa}(t)$. Following the same reasoning, the actual state at time point t is $S^\chi(t) = \tau^s_{\sigma,\kappa}(t)$. For the sake of brevity, when a set of occurrences of events $C = \{(e,t),\ e \in E^\chi(t),\ t \in \mathbb{T}\}$ will be used in the context of the update operator $\triangleright$ or the predicate $actualEff(E,L)$, it will actually only refer to the events of the couples in this set.

Definition 9 (Direct NESS-causes). *Given a causal setting χ, the occurrence of events set $C = \{(e,t),\ e \in E^\chi(t),\ t \in \mathbb{T}\}$ is a sufficient set of direct NESS-causes of the truthfulness of the formula ψ at t_ψ, denoted $C \underset{W}{\rightsquigarrow} (\psi, t_\psi)$, iff there exists a partial state $W \subseteq Lit_{\mathbb{F}}$ that we call backing such that:*

- *Causal sufficiency and minimality of W: $W \vDash \psi$ and $\forall W' \subset W,\ W' \nvDash \psi$. There is a decreasing sequence $t_1, \ldots, t_k$ and a partition $W_1, \ldots, W_k$ of W such that $\forall i \in \{1, \ldots, k\}$, given $C(t_i) = C \cap E^\chi(t_i)$:*
  - *Weak necessity and minimality of C at t_i: $S^\chi(t_i) \triangleright C(t_i) \vDash W_i$ and $\forall C' \subset C(t_i),\ S^\chi(t_i) \triangleright C' \nvDash W_i$.*
  - *Persistency of necessity: $\forall t,\ t_i < t \leq t_\psi,\ S^\chi(t) \vDash W_i$.*
- *Minimality of C: $C = \bigcup_{i \in \{1,\ldots,k\}} C(t_i)$.*

(e,t) is a direct NESS-cause of (ψ, t_ψ) iff $\exists C \subseteq \mathbb{E} \times \mathbb{T}$ such that $(e,t) \in C$, and $C \underset{W}{\rightsquigarrow} (\psi, t_\psi)$.

Wright's NESS test is based on three main principles which are formalised in Definition 9: (i) sufficiency of a set, (ii) weak necessity of the conditions in that set, and (iii) actuality of the conditions. (i) In this definition, the sufficient set is the partial state W. More precisely, given the directionality embedded in Section 3 semantics, we have causal sufficiency that Wright differentiates from logical sufficiency [38]: 'The successional nature of causation is incorporated in the concept of causal sufficiency, which is defined as the complete instantiation of all the conditions in the antecedent of the relevant causal law'. Moreover, Definition 9 introduces the constraint of necessity and sufficiency minimality which has been proven to be essential for regularity theories of causation [2, 5, 38]. The minimality of C condition ensures that weak necessity and minimality of C at t_i is applied to all elements in the set. In other words, it excludes the possibility to have in C an occurrence of event that has not occurred in one of the time points of the decreasing sequence $t_1, \ldots, t_k$. (ii) Definition 9 formalises weak necessity by subordinating necessity to sufficiency achieving that [38]: 'a causally relevant factor need merely be necessary for the sufficiency of a set of conditions sufficient for the occurrence of the consequence, rather than being necessary for the consequence itself'. It is worth mentioning that the condition $S^\chi(t_i) \nvDash W_i$—intuitively expected when referring to necessity—is included in the minimality condition by the case where $C' = \varnothing$, thus $S^\chi(t_i) = S^\chi(t_i) \triangleright \varnothing \nvDash W_i$. (iii) The actuality of the conditions is assured by the use of actual occurrence of events, which is implied by the presence of $E^\chi(t_i)$ in $C(t_i) = C \cap E^\chi(t_i)$.

Once causal sufficiency and minimality of the partial state W is defined, the causal inquiry is conducted by a recursive reasoning on a partition of W. The goal of this recursive reasoning is to identify the events which occurrence was necessary to the sufficiency of W. This reasoning is done by going back in time and analysing the information given by $\tau^s_{\sigma,\kappa}(t)$ and $\tau^e_{\sigma,\kappa}(t)$. Two limit cases can be identified. The first is when the partition set W_k is empty before its corresponding time t_k is equivalent to $t = 0$, meaning that all occurrences of events necessary for the sufficiency of W have been identified. When this is not the case, it means that there are fluent literals in W that were true in the initial state $S^{\chi}(0)$ and which value has not changed until $S^{\chi}(t_\psi)$. In this second case, the set C will contain the events $ini_l \in E^{\chi}(-1)$ whose l remains in W_k—events which symbolise events in the past beyond the framework of formalisation.

In practice, it is possible to study what will be considered as the direct NESS-causes of the truthfulness of ψ at t_ψ for each form that ψ may take. In the case where ψ is a fluent literal l, the direct NESS-causes will be the last occurrences of events to have made l true before or at t_ψ. In this basic case W is the singleton which unique element is that literal. This basic causal information is the one embedded in Section 3 action language semantics. In the case where ψ is a conjunction $\psi = l_1 \wedge \cdots \wedge l_m$ of fluent literals, the direct NESS-causes will be all the occurrence of events that are direct NESS-causes of the truthfulness of one of the literals l_i in the conjunction at t_ψ. Finally, the case where ψ is a disjunction of fluent literals, and more generally a disjunctive normal form, is by far the more interesting and challenging. Indeed, it is in this case that we can be confronted to situations of overdetermination. Whenever ψ is a disjunctive normal form, this means that there is a minimal causal sufficient backing W for each disjunct. Each of these backings is a possible way to cause the truthfulness of the formula ψ at t_ψ—in the same spirit as Beckers' paths [6]. Example 1 illustrates how Definition 9 handles one of those challenging situations.

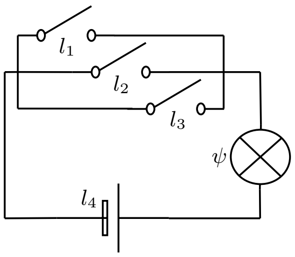

Fig. 1. Electrical circuit consisting of a voltage source, three switches, and an individual connected to electrodes.

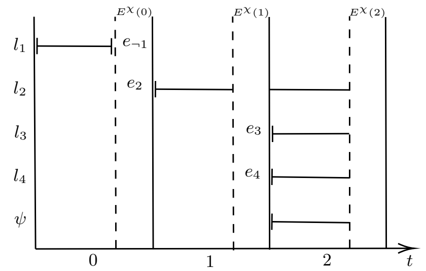

Fig. 2. Evolution of fluents given κ in Example 1.

Example 1 (parallel switches and Milgram). Consider Figure 1 simple electric circuit inspired by Milgram's experiment [29]. This circuit is made up of a voltage source, an individual strapped and connected to electrodes, and three switches connected in parallel. The positive literals $l_1, l_2, l_3, l_4 \in Lit_{\mathbb{F}}$ represent the closed state of each switch and the voltage source respectively—their respective complement thus represents the opened state. $\psi = (l_1 \wedge l_4) \vee (l_2 \wedge l_4) \vee (l_3 \wedge l_4)$ where $\psi \in \mathcal{P}$ represents the triggering conditions for the strapped individual being electrocuted. Thus, three backings are possible to cause ψ: $W = \{l_1, l_4\}$, $W' = \{l_2, l_4\}$, and $W'' = \{l_3, l_4\}$. $e_1, e_2, e_3 \in \mathbb{E}$ are the events which intrinsic effect is to close each switch respectively, $e_4 \in \mathbb{E}$ is an event which intrinsic effect is to close the voltage source, and $e_{\neg 1} \in \mathbb{E}$ is the event which intrinsic effect is to open the first switch. We assume that the situation involves five agents: the one strapped and four others—each controlling one of the four components of the circuit. The studied sequences illustrated by Figure 2 and given by $\tau^{e}_{\sigma,\kappa}$ and $\tau^{s}_{\sigma,\kappa}$ are:

$$E^{\chi}(-1) = \left\{ini_{l_1}, ini_{\overline{l_2}}, ini_{\overline{l_3}}, ini_{\overline{l_4}}\right\}$$
$$S^{\chi}(0) = \left\{l_1, \overline{l_2}, \overline{l_3}, \overline{l_4}\right\}, E^{\chi}(0) = \{e_{\neg 1}, e_2\}$$
$$S^{\chi}(1) = \left\{\overline{l_1}, l_2, \overline{l_3}, \overline{l_4}\right\}, E^{\chi}(1) = \{e_3, e_4\}$$
$$S^{\chi}(2) = \left\{\overline{l_1}, l_2, l_3, l_4\right\}$$

Given the above traces, ψ is true at $t = 2$ by both W' and W''.

The question that arises in Example 1 is: what are the causes of ψ being true at $t = 2$? Said in another way, what are the causes of the strapped individual being electrocuted at $t = 2$? Batusov and Soutchanski's proposal will consider $(ini_{l_1}, -1)$ and $(e_4, 1)$ as 'achievement causes', and $(e_2, 0)$ as a 'maintenance cause'—'causes responsible for protecting a previously achieved effect, despite potential threats that could destroy the effect' [4]—this given that we omit to consider $(e_3, 1)$ in the comparison because it occurs at the same time as $(e_4, 1)$ and thus requires definitions that handle concurrency. Considering factuality as an essential feature of a causal inquiry, the presence of $(ini_{l_1}, -1)$ in the causes seems unacceptable. Factually, $(ini_{l_1}, -1)$ plays no role in the truthfulness of ψ at $t = 2$. Definition 9 gives the sets $\{(e_2, 0), (e_4, 1)\}$ and $\{(e_3, 1), (e_4, 1)\}$ which union gives the answer $\{(e_2, 0), (e_3, 1), (e_4, 1)\}$.

The interpretation given by Batusov and Soutchanski of Example 1 is not the only possible divergent interpretation. We wondered whether answer $\{(e_2, 0), (e_4, 1)\}$ alone was not more satisfactory given that, even if both l_2 and l_3 are true at $t = 2$, the precedence of l_2 could be taken into account. However, this intuition appears as conflating causality and responsibility. If we strictly limit ourselves to a factual causal inquiry as prescribed by Wright [36], both $(e_2, 0)$ and $(e_3, 1)$ are causes of the truthfulness of ψ at $t = 2$. The intuition that would induce us to take into account the precedence of $(e_2, 0)$ belongs to Wright's *proximate-cause inquiry* and not to the causal inquiry. Indeed, once $(e_2, 0)$ and $(e_3, 1)$ are identified as causes, there is a policy choice for which the precedence of $(e_2, 0)$ mitigates or eliminates the responsibility of $(e_3, 1)$ for the final effect. We suspect that Batusov and Soutchanski's choices—which led them to consider $(ini_{l_1}, -1)$ as a cause—were influenced by this same intuition, but taken even further.

Definition 9 gives us essential information about causal relations by looking to the actual effects of events. However, the set of direct NESS-causes of an effect may include exogenous events that are not necessarily relevant. This is especially true in a framework such as ours, where we are interested in agent's decisions—thus actions. It is therefore essential to establish a causal chain by going back in time in order to find the set of actions that led to the effect. To this end, we must broaden our vision to look not only at the actual effects of events which are direct NESS-causes, but also (i) at the events that caused those events to be triggered and (ii) at the events that caused those events to have their actual effects.

Example 2 (causing events to have their actual effects). Consider the literals $l_1, l_2, l_3, l_{c_1}, l_{c_3} \in Lit_{\mathbb{F}}$, the formula $\psi = l_1 \wedge l_2 \wedge l_3$, events $e, e' \in \mathbb{E}$ where there respective effects are $eff(e) = \left\{[l_{c_1}]\overline{l_1}, [\top]l_2, [l_{c_3}]l_3\right\}$ and $eff(e') = \left\{[\top]\overline{l_{c_1}}, [\top]l_{c_3}\right\}$. The studied sequences illustrated by Figure 3 and given by $\tau^e_{\sigma,\kappa}$ and $\tau^s_{\sigma,\kappa}$ are:

$$E^{\chi}(-1) = \left\{ini_{l_1}, ini_{\overline{l_2}}, ini_{\overline{l_3}}, ini_{\overline{l_4}}, ini_{l_{c_1}}, ini_{\overline{l_{c_3}}},\right\}$$
$$S^{\chi}(0) = \left\{l_1, \overline{l_2}, \overline{l_3}, l_{c_1}, \overline{l_{c_3}}\right\}, E^{\chi}(0) = \{e'\}$$
$$S^{\chi}(1) = \left\{l_1, \overline{l_2}, \overline{l_3}, \overline{l_{c_1}}, l_{c_3}\right\}, E^{\chi}(1) = \{e\}$$
$$S^{\chi}(2) = \left\{l_1, l_2, l_3, \overline{l_{c_1}}, l_{c_3}\right\}$$

Given the above sequence, Definition 9 gives us the direct NESS-cause relation $C \underset{W}{\rightsquigarrow} (\psi, t_{\psi})$ where C is the set $\{(e, 1), (ini_{l_1}, -1)\}$.

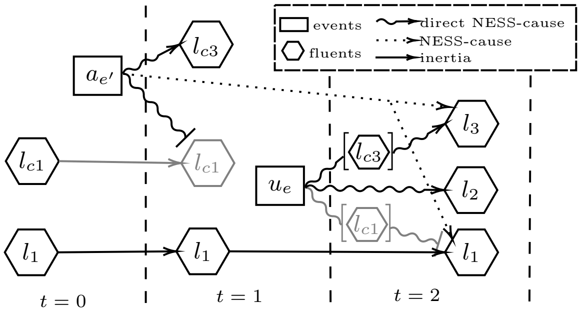

Fig. 3. Causal relations in Example 2.

In Example 2, the actual effects of the occurrence $(e, 1)$ were $actualEff(\{e\}, S^{\chi}(1)) = \{l_2, l_3\}$. In order to determine the desired causal chain, one of the steps requires to ask ourselves what occurrence of events caused $(e, 1)$ to have those effects—inquiry concerning exclusively conditional effects which condition is not $[\top]$. We distinguish two cases both illustrated by Figure 3. The two effects concerned are, $[l_{c_1}]\overline{l_1}$ and $[l_{c_3}]l_3$, each one representing a case. The effect $[l_{c_1}]\overline{l_1}$ corresponds to the case where the complement of the condition $[l_{c_1}]$ has been direct

NESS-caused, thus causing $(e, 1)$ to 'maintain' l_1. The effect $[l_{c_3}]l_3$ corresponds to the case where the condition $[l_{c_3}]$ has been direct NESS-caused, thus causing $(e, 1)$ to 'produce' l_3 as an actual effect. The predicate $after(E, L_p, L_m)$—inspired by Khan and Lespérance's work [23]—gives the formula to direct NESS-cause in order to be considered a cause of an event having its actual effects. In the discussed example this formula is $\psi' = \overline{l_{c_1}} \wedge l_{c_3}$.

Definition 10 ($after(E, L_p, L_m)$). *Given a causal setting χ, a set of events $E \in E^\chi(t)$, and partial states $L_m, L_p, W_{\psi'} \subseteq Lit_{\mathbb{F}}$ such that $S^\chi(t) \vDash L_m$ and $S^\chi(t) \nvDash L_p$, the predicate $after(E, L_p, L_m) = \psi'$ with $\psi' = \bigwedge_{l \in W_{\psi'}} l$ such that:*

- *Necessity and minimality of E: $W_{\psi'} \triangleright E \vDash L_p \cup L_m$ and $\forall E' \subset E, W_{\psi'} \triangleright E' \nvDash L_p \cup L_m$.*
- *Monotonicity: $\forall W', W_{\psi'} \subseteq W', W' \triangleright E \vDash L_p \cup L_m$.*

Having introduced the predicate $after(E, L_p, L_m)$, we can now introduce NESS-causes that are found relying on the establishment of the causal chain.

Definition 11 (NESS-causes). *Given a causal setting χ, the direct NESS-cause relation $C \underset{W}{\rightsquigarrow} (\psi, t_\psi)$, and the decreasing sequence $t_1, \dots, t_k$ induced by the existing partition $W_1, \dots, W_k$ of the backing W, the occurrence of events set $C' = \{(e,t),\ e \in E^\chi(t),\ t \in \mathbb{T}\}$ is a sufficient set of NESS-causes of the truthfulness of the formula ψ at t_ψ iff one of the following cases is satisfied:*

- *Base case: $C' = C$.*
- *Recursive case: Given the sets $C_R = C \setminus C'$ and $C_O = C' \setminus C$ of 'removable' and 'overwhelming' occurrence of events respectively, and the partitions of C and C_R matching the decreasing sequence $t_1, \dots, t_k$—$C(t_1), \dots, C(t_k)$ and $C_R(t_1), \dots, C_R(t_k)$ respectively—there is a covering sequence of subsets $C_O = \bigcup_{i \in \{0,\dots,k\}} C_{O_i}$ (not necessarily monotonic in time) such that:*
  $\forall i \in \{1, \dots, k\}, C_R(t_i) \neq \varnothing \implies (e,t) \in C_{O_i}$ are NESS-causes of (ψ', t_i), where $\psi' = tri(C_R(t_i)) \wedge after(C_R(t_i), L_p, L_m)$, $L_p = W_i \cap actualEff(C_R(t_i), S^\chi(t_i))$, and $L_m = [W_i \setminus actualEff(C_R(t_i), S^\chi(t_i))] \cup W_{i+1} \cup \dots \cup W_k$.

(e,t) is a NESS-cause of (ψ, t_ψ) iff $\exists C' \subseteq \mathbb{E} \times \mathbb{T}$ such that $(e,t) \in C'$ and the occurrence of events set C' is a sufficient set of NESS-causes of (ψ, t_ψ). The set of NESS-causes $D = C \setminus C_R \cup C'$ is called a set of decisional causes if $D \subseteq \mathbb{A} \times \mathbb{T}$.

Definition 11 captures the two ways in which the occurrence of an event can have a causal relation with the truthfulness of ψ at t_ψ. First, by being a NESS-cause of the triggering conditions of an occurrence of event that is a NESS-cause of (ψ, t_ψ)—captured by the conjunct $tri(C_R(t_i))$. Second, by being a NESS-cause that the occurrence of an event that is a NESS-cause of (ψ, t_ψ) had its actual effects—captured by the conjunct $after(C_R(t_i), L_p, L_m)$.

Having determined the causal relations linking events and formulas of the language, we can now give a suitable for action languages definition of actual causality.

Definition 12 (actual cause). *Given a causal setting χ and an event $e \in E^{\chi}(t_{\psi})$, the actual causes of (e, t_{ψ}) are the NESS-causes of $(tri(e), t_{\psi})$, i.e. the truthfulness of the triggering conditions of e at t_{ψ}.*

5 Conclusion

The contribution of this paper is to link automated planning and causality by continuing the momentum established by recent papers [4, 8, 21, 24] of proposing an action languages suitable definition of actual causality. By this proposal we address two of what we consider the main remaining limitations of this venture. First, not to yield into the easy confusion between responsibility and causality, our proposal is suitable for a factual and independent of policy choices causal inquiry. Second, not to disregard the much debated cases of overdetermination, our proposal is based on an action language semantics allowing concurrency of events. By taking as a base Wright's NESS test—as done recently in causal calculus [9] and in structural equation framework [6]—we are able to manage these cases satisfactorily. To the best of our knowledge, no other action languages suitable definition of actual causality has been able to handle those complex cases, yet essential. Our approach thus allows agents to handle complex cases of actual causality.

In future work we intend to propose a complete and sound translation into logic programming of this actual causation definition suitable for action languages. Then, we intend to extend the definition of causality by including the relation 'prevent' [8]. In Wright's conception of causality, causality can only be sufficient if we take into account—in addition to the positive causes—the conditions that were not true and whose absence was a necessary condition for the occurrence of the result. Then, the events being causes of their absence are also causes of the result. By working on fluent literals, our definition of causation already takes this notion into account. If we extend this reasoning, we could also take the case where the result did not occur because one of these negative conditions was made true. In such a case, the events being causes of the negative condition are causes of the non-occurrence of the result. We intend to define this causal relation given our more complex framework with events concurrency and disjunction.

References

1. Abarca, A.I.R., Broersen, J.M.: A Stit Logic of Responsibility. In: 21st International Conference on Autonomous Agents and Multiagent Systems. pp. 1717–1719. International Foundation for Autonomous Agents and Multiagent Systems, Auckland, New Zealand (May 2022)
2. Andreas, H., Guenther, M.: Regularity and Inferential Theories of Causation. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, fall 2021 edn. (2021)
3. Baral, C., Gelfond, M.: Reasoning About Effects of Concurrent Actions. J. Log. Program. **31**(1-3), 85–117 (1997)

4. Batusov, V., Soutchanski, M.: Situation Calculus Semantics for Actual Causality. In: McIlraith, S.A., Weinberger, K.Q. (eds.) Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18). pp. 1744–1752. AAAI Press, New Orleans, Louisiana, USA (2018)
5. Baumgartner, M.: A Regularity Theoretic Approach to Actual Causation. Erkenntnis **78**(1), 85–109 (Dec 2013)
6. Beckers, S.: The Counterfactual NESS Definition of Causation. Proceedings of the AAAI Conference on Artificial Intelligence **35**(7), 6210–6217 (May 2021)
7. Beckers, S., Vennekens, J.: A principled approach to defining actual causation. Synth. **195**(2), 835–862 (2018)
8. Berreby, F., Bourgne, G., Ganascia, J.G.: Event-Based and Scenario-Based Causality for Computational Ethics. In: 17th International Conference on Autonomous Agents and Multiagent Systems. pp. 147–155. International Foundation for Autonomous Agents and Multiagent Systems, Stockholm, Sweden (Jul 2018)
9. Bochman, A.: Actual Causality in a Logical Setting. In: Lang, J. (ed.) Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018. pp. 1730–1736. Stockholm, Sweden (2018)
10. Bochman, A.: On Laws and Counterfactuals in Causal Reasoning. In: Thielscher, M., Toni, F., Wolter, F. (eds.) Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference. pp. 494–503. AAAI Press, Tempe, Arizona (2018)
11. Bourgne, G., Sarmiento, C., Ganascia, J.G.: ACE modular framework for computational ethics: dealing with multiple actions, concurrency and omission. In: 1st International Workshop on Computational Machine Ethics. CEUR-WS.org (2021)
12. Fox, M., Long, D.: Modelling Mixed Discrete-Continuous Domains for Planning. Journal of Artificial Intelligence Research **27**, 235–297 (Oct 2006)
13. Gelfond, M., Lifschitz, V.: Representing Action and Change by Logic Programs. J. Log. Program. **17**(2/3&4), 301–321 (1993)
14. Gelfond, M., Lifschitz, V.: Action Languages. Electron. Trans. Artif. Intell. **2**, 193–210 (1998)
15. Ghallab, M., Knoblock, C., Wilkins, D., Barrett, A., Christianson, D., Friedman, M., Kwok, C., Golden, K., Penberthy, S., Smith, D., Sun, Y., Weld, D.: PDDL - The Planning Domain Definition Language. Technical Report CVC TR-98-003/DCS TR-1165, Yale Center for Computational Vision and Control (Aug 1998)
16. Giunchiglia, E., Lifschitz, V.: An Action Language Based on Causal Explanation: Preliminary Report. In: Mostow, J., Rich, C. (eds.) Proceedings of the Fifteenth National Conference on Artificial Intelligence and Tenth Innovative Applications of Artificial Intelligence Conference, AAAI 98, IAAI 98, July 26-30, 1998, Madison, Wisconsin, USA. pp. 623–630. AAAI Press / The MIT Press (1998)
17. Hall, N., Paul, L.A.: Causation and Pre-emption. In: Clark, P., Hawley, K. (eds.) Philosophy of Science Today. Oxford University Press, Oxford, New York (May 2003)
18. Halpern, J.Y.: Actual Causality. The MIT Press (2016)
19. Halpern, J.Y.: Actual Causality: A Survey: Joseph Halpern (Jun 2018), https://www.youtube.com/watch?v=hXnCX2pJ0sg, 26:38-27:21
20. Haslum, P., Lipovetzky, N., Magazzeni, D., Muise, C.: An Introduction to the Planning Domain Definition Language. No. 42 in Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers (Apr 2019)
21. Hopkins, M., Pearl, J.: Causality and Counterfactuals in the Situation Calculus. Journal of Logic and Computation **17**(5), 939–953 (Oct 2007)

22. Hume, D.: Enquête sur l'entendement humain. No. 1305 in GF, Flammarion, Paris, 2006 edn. (1748)
23. Khan, S.M., Lespérance, Y.: Knowing Why: On the Dynamics of Knowledge about Actual Causes in the Situation Calculus. In: Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems. pp. 701–709. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (May 2021)
24. LeBlanc, E.C., Balduccini, M., Vennekens, J.: Explaining Actual Causation via Reasoning About Actions and Change. In: Calimeri, F., Leone, N., Manna, M. (eds.) Logics in Artificial Intelligence - 16th European Conference, JELIA 2019, Rende, Italy, May 7-11, 2019, Proceedings. Lecture Notes in Computer Science, vol. 11468, pp. 231–246. Springer (2019)
25. Lewis, D.: Causation. Journal of Philosophy **70**(17), 556–567 (1973), publisher: Oxford Up
26. Lorini, E., Longin, D., Mayor, E.: A logical analysis of responsibility attribution: emotions, individuals and collectives. J. Log. Comput. **24**(6), 1313–1339 (2014)
27. Mackie, J.L.: The Cement of the Universe: A Study of Causation. Clarendon Library of Logic and Philosophy, Oxford University Press, Oxford (1980)
28. Menzies, P., Beebee, H.: Counterfactual Theories of Causation. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, winter 2020 edn. (2020)
29. Milgram, S.: Behavioral Study of obedience. The Journal of Abnormal and Social Psychology **67**(4), 371–378 (1963), publisher: American Psychological Association
30. Mill, J.S.: A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation, Cambridge Library Collection - Philosophy, vol. 1. Cambridge University Press, Cambridge, 2011 edn. (1843)
31. Mueller, E.T.: Commonsense Reasoning: An Event Calculus Based Approach. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2 edn. (2014)
32. Pearl, J.: Causality: models, reasoning, and inference. Cambridge University Press, Cambridge, U.K. ; New York (2000)
33. Pednault, E.P.D.: ADL: exploring the middle ground between STRIPS and the situation calculus. In: Proceedings of the first international conference on Principles of knowledge representation and reasoning. pp. 324–332. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (Dec 1989)
34. Russell, S., Norvig, P.: Artificial Intelligence - A Modern Approach. Prentice Hall Series, Pearson Education, third edn. (2010)
35. Vincent, N.A.: A Structured Taxonomy of Responsibility Concepts. In: Vincent, N.A., van de Poel, I., van den Hoven, J. (eds.) Moral Responsibility, Library of Ethics and Applied Philosophy, vol. 27. Springer Netherlands, Dordrecht (2011)
36. Wright, R.W.: Causation in Tort Law. California Law Review **73**(6), 1735–1828 (1985), publisher: California Law Review, Inc.
37. Wright, R.W.: Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts. Iowa Law Review **73**, 1001 (Dec 1988)
38. Wright, R.W.: The NESS Account of Natural Causation: A Response to Criticisms. In: Goldberg, R. (ed.) Perspectives on Causation. Social Science Research Network, Rochester, NY, hart publishing edn. (Jul 2011)