*ECAI 2023*
*K. Gal et al. (Eds.)*


*doi:10.3233/FAIA230474*

# Anticipating Responsibility in Multiagent Planning

**Timothy Parker[a;*], Umberto Grandi[a] and Emiliano Lorini[a]**

[a]IRIT, CNRS, University of Toulouse, France

**Abstract.** Responsibility anticipation is the process of determining if the actions of an individual agent may cause it to be responsible for a particular outcome. This can be used in a multi-agent planning setting to allow agents to anticipate responsibility in the plans they consider. The planning setting in this paper includes partial information regarding the initial state and considers formulas in linear temporal logic as positive or negative outcomes to be attained or avoided. We firstly define attribution for notions of active, passive and contributive responsibility, and consider their agentive variants. We then use these to define the notion of responsibility anticipation. We prove that our notions of anticipated responsibility can be used to coordinate agents in a planning setting and give complexity results for our model, discussing equivalence with classical planning. We also present an outline for solving some of our attribution and anticipation problems using PDDL solvers.

## 1 Introduction

In any multi-agent setting, a key concept is that of responsibility. There are two main notions of responsibility, which are forward-looking and backward-looking responsibility [25]. In general, forward-looking responsibility is to have an obligation to bring about or prevent a certain state of affairs, while backward-looking responsibility means to be held accountable for a particular action or state of affairs that occurred. Our paper considers only backward-looking responsibility, which is often used in multi-agent settings to determine appropriate sanctions or rewards for agents. While responsibility attribution is a well-studied problem [1, 2, 4, 12, 19], we focus on the novel concept of responsibility anticipation, which means to determine if a particular plan for a single agent *may* lead to their responsibility for some outcome, given the possible plans of all other agents. We believe that by anticipating responsibility, agents will be better able to coordinate their actions even if they cannot communicate. We consider responsibility in a multi-agent setting with concurrent actions and where outcomes are described in Linear Temporal Logic over finite traces ($\mathsf{LTL}_f$)[7]. Following the work of Lorini et al. [17] we recognise two key components to responsibility, the causal and agentive components. The causal component requires that the actions of the agent in some way contributed to the outcome in question. Lorini et al. identify two different notions of causal responsibility, active and passive responsibility. We formalise both in our model as well as a notion of contributive responsibility defined by Braham and Van Hees [4]. Roughly speaking, given some state of affairs $\omega$, active responsibility means to bring about $\omega$, passive responsibility means to allow $\omega$ to occur, and contributive responsibility means to be part of a coalition that brings about $\omega$. The agentive component requires that the agent is aware that their actions will (or in some cases may) contribute to the outcome. In our setting the agents have full knowledge of the action theory (i.e the capabilities of all agents), but are uncertain regarding the intended actions of other agents and the initial state of the world. This allows us to define agentive notions of active, passive and contributive responsibility.

While our model allows us to attribute responsibility retrospectively (after plan execution), the focus of our work is in anticipating responsibility to aid in plan selection for a single agent. Since agents often cannot be certain about the outcomes of their plans, we introduce a notion of anticipated responsibility based on our previous work [10] which can be applied to any of our introduced notions of responsibility. We show avoiding anticipated responsibility for a negative outcome, agents can often guarantee that the outcome does not occur, even in some cases where the agents cannot communicate and where no single agent can guarantee avoiding the negative outcome.

We intend for our model to be useful in real-world planning applications. This is why we have taken efforts to ensure that the our planning domain is reasonably compact while still being highly expressive. We also outline how our responsibility attribution and anticipation problems can be reduced to PDDL, both to demonstrate how pre-existing planning solvers can be applied to our problems and to encourage implementation of our model.

Our paper is organised as follows. Section 2 situates our paper with reference to related work in responsibility attribution, and compares our work to several similar papers. Section 3 introduces our multi-agent planning domain and presents an explanatory example. Section 4 formalises our notions of responsibility attribution and anticipation and discusses their application to multi-agent planning. Section 5 gives the complexity results for our setting and an outline of a reduction to PDDL. Finally section 6 summarises the paper and outlines directions for future work.

## 2 Related Work

This work contributes primarily to the formalisation of responsibility attribution. It also involves planning with temporally extended goals [3, 8, 5], but since we are not aware of any other work in planning that considers responsibility in plan selection, we will focus this section on responsibility. Our planning model builds on a number of previous papers which are discussed in section 3.

Furthermore, since we are not aware of any other work on responsibility anticipation and its application to planning agents, we will focus on approaches to responsibility attribution in the literature, and discuss how and why they differ from our work.

One approach to formalising responsibility is the work of Alechina et al. [1], which is based on work by Chockler and Halpern [6, 11] on the formalisation of responsibility. Rather than using $\mathsf{LTL}_f$, as in our

* Corresponding Author. Email: timothy.parker@irit.fr

approach, this work uses structural equation modelling (SEM). Their paper focuses specifically on responsibility attribution for the failure of a previously-arranged joint plan, which is a specific sequence of tasks that all agents are expected to follow (but perhaps will not), meaning that the setting of the model is much more specific than our work. Unlike our model, the authors focus only on a single notion of responsibility, this notion allows varying degrees of responsibility for different agents. Alechina et al. also perform a complexity analysis of their model, showing that responsibility attribution is in general NP-Complete (in line with our notion of passive responsibility, see theorem 6) and identify some fragments where responsibility attribution is polynomial.

Halpern and Kleiman-Weiner [12] also use a structural equations model, but focuses on defining the intentions of agents given their actions and their epistemic state (given here as a probability distribution). As this paper does not address causal responsibility there is not much overlap with our model, but it does highlight several interesting concepts that we could attempt to incorporate in future work.

A more general but less compact approach is the work of Baier et al. [2]. Their work covers both forward and backward-looking responsibility attribution, but we will focus on their formalisation of backward-looking responsibility. Whereas our model is based of classical planning, Baier et al. use extensive form games with strategies instead of plans. This makes their model much less compact and more complex than ours, but also more expressive. In their work, a coalition of agents $J$ is causal backwards responsible for some outcome $\omega$ if, fixing the strategies of all other agents (and the random choices of Nature), there exists a strategy for $J$ where $\omega$ does not occur in any possible execution. They also define strategic backward responsibility, which states that $\omega$ occurs, and there is some state in the execution where the coalition of agents has a strategy such that $\omega$ does not occur in any epistemically possible outcome for that strategy (since agents cannot distinguish between some states). Again, this model does not include any other notion of responsibility, but does model an agent's degree of responsibility, which is determined by its membership to one or more responsible coalitions, similarly to our notion of contributive responsibility, though defined on strategies instead of plans. Baier et al. also note that the computational complexity of responsibility attribution is in NP, which is a lower bound than contributive responsibility in our model (see theorem 6), though our model is exponentially more compact.

A similar definition of responsibility exists in the work of Naumov and Tao [19] whose setting of Imperfect Information Strategic Games is very close to our notion of planning domain, but restricted to plans of length 1. Their notion of blameworthiness says that $i$ is blameworthy for $\omega$ if $\omega$ occurrs and $i$ could have performed an action guaranteeing $\neg\omega$ in all possible states. This is a stronger version of our notion of causal passive responsibility, as we require only that $i$ could have avoided $\omega$ if the state and all actions of other agents were fixed. They also present a notion of "seeing to it" which requires that an agent guarantees in all possible worlds that $\omega$ occurs. This is very close to our notion of agentive active responsibility, the only difference being that in our model there must exist some possible history from the initial state where the outcome does not occur, whereas in their model that history can start at any epistemically possible state for $i$. Also, unlike us Naumov and Tao formalise their notions as operators in logic, allowing for the development of a proof system for these operators (they develop a proof system for their notion of blameworthiness in a previous, perfect-information setting [18]).

Our work is heavily inspired by the work of Lorini et al. [17]. This paper formalises the notions of active and passive responsibility that we use in this paper, as well as the variant of agentive responsibility. The model of Lorini et al. is based on STIT logic in a multi-agent setting using a Kripke possible world semantics. We extend their work to the setting of multi-agent planning, though for simplicity we do not model agents' knowledge of the possible actions of other agents, in our setting all plans of the other agents are considered possible.

Our work is also inspired by the work of Braham and van Hees [4], who analyse responsibility in a game-theoretic framework. One of their conditions for moral responsibility is that an agent's actions must have "causally contributed" to the outcome in question. We adapt the notion of causal contribution into our setting as the notion of causal responsibility.

## 3 Model

In this section we introduce the planning framework in which we will define our notions of responsibility. As many of our definitions are drawn from existing literature, in the interest of space we have chosen to omit some of the less informative formal definitions. We will indicate where we have done this.

### 3.1 Agents, Actions and Histories

The building blocks of our model are a finite set of agents $Agt$ and a countable set of propositions $Prop = \{p, q, \ldots\}$. From $Prop$ we define a set of states $S = 2^{Prop}$, with elements $s, s', \ldots$ Let $Act = \{a, b, \ldots\}$ be a finite non-empty set of action names.

To trace the actions of agents and changing states over time we define a $k$-history to be a pair $H = (H_{st}, H_{act})$ with $H_{st} : \{0, \ldots, k\} \to S$ and $H_{act} : Agt \times \{0, \ldots, k-1\} \to Act$. Time point $k$ is the point in time when all agents have finished acting, so no actions take place at this time. The set of $k$-histories is noted $Hist_k$. The set of all histories is $Hist = \bigcup_{k \in \mathbb{N}} Hist_k$.

### 3.2 Multi-Agent Action Theory

In order to calculate the effects of an agent's actions, we introduce a compact action theory based on situation calculus [21]. We first define $\mathcal{L}_{\mathsf{PL}+}$ (propositional logic with action descriptions) as follows:

$$\varphi \quad ::= \quad p \mid do(i, a) \mid \neg\varphi \mid \varphi \wedge \varphi$$

with $p$ ranging over $Prop$, $i$ ranging over $Agt$ and $a$ ranging over $Act$. Atomic formulas in this language are those that consist of a single proposition $p$ or a single instance of $do(i, a)$.

Semantic interpretation of formulas in $\mathcal{L}_{\mathsf{PL}+}$ is performed relative to a $k$-history $H \in Hist$ and a time point $t \in \{0, \ldots, k\}$ as follows (we omit boolean cases which are defined as usual):

$$\begin{aligned} H, t \models p \quad &\iff p \in H_{st}(t), \\ H, t \models do(i, a) \quad &\iff t < k \text{ and } H_{act}(i, t) = a \end{aligned}$$

We define our action theory as a pair of a positive and negative effect precondition function $\gamma = (\gamma^+, \gamma^-)$, where $\gamma^+ : Agt \times Act \times Prop \to \mathcal{L}_{\mathsf{PL}+}$ and $\gamma^- : Agt \times Act \times Prop \to \mathcal{L}_{\mathsf{PL}+}$. If the formula $\gamma^+(i, a, p)$ holds in a state where action $a$ is executed by agent $i$, proposition $p$ will be *true* in the next state (provided no other action interferes). Similarly, $\gamma^-(i, a, p)$ guarantees that $p$ will be *false* in the next state if action $a$ is executed by $i$ (without interference). In case of conflicts between actions, we use an inertial principle: if two or more actions attempt to enforce different truth values for $p$, then the truth value of $p$ does not change.

To signal that action $a$ is not available to agent $i$ we can simply set $\gamma^+(i,a,p) = \gamma^-(i,a,p) = \bot$ for all $p \in Prop$. We assume the existence of a "do nothing" action $Skip$, defined such that $\gamma^+(i,Skip,p) = \gamma^-(i,Skip,p) = \bot$ for all $i$ and $p$.

We say that history $H$ is a $\gamma$-compatible history for action theory $\gamma = (\gamma^+, \gamma^-)$ if each state respects the actions performed in the previous state. The set of $\gamma$-compatible histories is noted $Hist(\gamma)$. A formal definition can be found in the full version of the paper [20].

### 3.3 Compactness of our Action Theory

A common alternative to our action theory in the field of responsibility attribution is a state transition function $\tau : S \times Act^{Agt} \to S$ (see, e.g., [2, 14]). This takes as input the current state and the actions of all agents, and outputs the next state (which can be any state). It is therefore straightforward to see that for any deterministically consistent history $H$ (meaning the same joint action in the same state always leads to the same outcome), there is some state transition function $\tau$ that can be used to generate $H$ given the start state and the actions of all agents. However, we can show that our action theory is equally as expressive as any state transition function, and strictly more succinct, and this is achieved by our use of action descriptions.

**Proposition 1.** *Given a state transition function $\tau$, there exists an action theory $\gamma$ that is equivalent to (generates the same histories as) $\tau$ and is at worst polynomially larger in size.*

**Proposition 2.** *There is a state transition function $\tau_1$ such that if an action theory $\gamma$ is equivalent to $\tau_1$, then $\gamma$ contains $do(i,a)$ in its description.*

Note that the size of $\tau$ is always exponential in the size of $Prop$ and $Agt$, since the number of entries in $\tau$ are fixed. On the other hand, entries for $\gamma$ can in be as small as constant size (for example $\gamma^{\pm}(i,a,p) \in \{\top, \bot\}$). This means $\gamma$ can be as small as $2 \times |Act| \times |Agt| \times |Prop|$. We conjecture that in most applications for this planning model, the action theory $\gamma$ will be polynomial in size in $Prop$, $Agt$ and $Act$. We note in passing that there exist forms of compact actions theories such as the work of Zhu et al. [26].

### 3.4 Planning Domains with Partial Information

We can now define our notion of planning domain. This is a space where agents can create and execute plans, and where the outcomes of those plans can be determined. Since our planning domain includes partial information we make use of epistemic equivalence sets. An epistemic equivalence set $S_i \subseteq S$ is the set of possible start states from the perspective of agent $i$.

**Definition 1** (Partial Information Multi-Agent Planning Domain)**.** *A Partial information multi-agent Planning Domain (PPD) is a tuple $\nabla = (\gamma, s_0, (S_i)_{i \in Agt})$ where $\gamma = (\gamma^+, \gamma^-)$ is an action theory, $s_0$ is an initial state, and for each $i \in Agt$, $S_i$ is the epistemic equivalence set for $i$.*

Our notion of an epistemic equivalence set is straightforward and very general, but not very compact. A more compact alternative would be to give each agent visibility of a certain subset of the propositions in $Prop$ [23]. However, this would be less general as not all epistemic equivalence sets can be expressed in terms of visibility. A more complex and more general approach would be to give each agent a belief base $\mathsf{Bel}_i$ as a set of formulas of $\mathcal{L}_{\mathsf{PL}}$ that describes the beliefs of $i$ regarding the initial state [16]. We use epistemic equivalence sets as this is the simplest notion for defining algorithms and any of the above methods will induce an epistemic equivalence set.

**Example 1** (Crossing a Junction)**.** *The planning domain $\nabla_E$ models an autonomous vehicle (Agent 1) approaching a junction. Agent 1 knows that there is a second vehicle (Agent 2) near the junction, but does not know if Agent 2 has crossed the junction. Each vehicle can either go straight on ($Move$), or do nothing ($Skip$).*

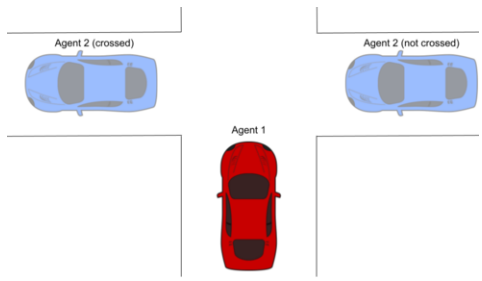


**Figure 1.** A visual representation of $\nabla_E$ in Example 1, showing the start position of Agent 1 and the two possible positions of Agent 2: crossed or not crossed the junction.

*The example is formally defined as follows:*

- $Agt = \{A1, A2\}$
- $Prop = \{crossed_1, crossed_2, collision\}$
- $Act = \{Move, Skip\}$
- $s_0 = \emptyset$, $s_1 = \{crossed_2\}$
- $S_1 = \{s_0, s_1\}$

*The action theory for our example is defined as follows, note that we have already defined the preconditions for Skip in section 3.2:*

$$\begin{aligned}
\gamma^+(A1, Move, crossed_1) =& \neg(\neg crossed_2 \wedge do(A2, Move)) \\
& \wedge \neg collision \\
\gamma^+(A1, Move, collision) =& \neg crossed_1 \wedge \neg crossed_2 \\
& \wedge do(A2, Move) \\
\gamma^+(A2, Move, crossed_2) =& \neg(\neg crossed_1 \wedge do(A1, Move)) \\
& \wedge \neg collision \\
\gamma^{\pm}(i, Move, p) =& \bot \textit{ unless stated otherwise above.}
\end{aligned}$$

*In words, if exactly one agent attempts to cross the junction ($Move$) then they will succeed. If both agents perform $Move$ at the same time then they will collide, which will prevent either from being able to move.*

### 3.5 Action Sequences and Joint Plans

Now that we have defined a planning domain, we can define the notions of action sequence and plan. Given $k \in \mathbb{N}$, a $k$-action-sequence is a function $\pi : \{0, \ldots, k-1\} \to Act$. The set of $k$-action-sequences is noted $Seq_k$. The set of all action sequences is $Seq = \bigcup_{k \in \mathbb{N}} Seq_k$. For a (non-empty) coalition of agents $J \in 2^{Agt} \setminus \emptyset$ we define a joint $k$-plan as a function $\Pi : J \to Seq_k$ (if $J$ is a singleton coalition then $\Pi$ is an individual plan). The set of joint $k$-plans for a coalition $J$ is written $Plan^J_k$. The set of all joint plans for $J$ is $Plan^J = \bigcup_{k \in \mathbb{N}} Plan^J_k$.

Given a joint plan $\Pi$ for coalition $J$ and another coalition $J' \subseteq J$, we can write the sub-plan of $\Pi$ corresponding to $J'$ as $\Pi^{J'}$, we also write $\Pi^{-J'}$ for sub-plan corresponding to $J \setminus J'$. Given two $k$-plans $\Pi_1$ and $\Pi_2$ for disjoint coalitions $J_1, J_2$, we write $\Pi_1 \cup \Pi_2$ for the joint plan for $J_1 \cup J_2$ such that $(\Pi_1 \cup \Pi_2)^{J_1} = \Pi_1$ and $(\Pi_1 \cup \Pi_2)^{J_2} = \Pi_2$. Finally, given two plans $\Pi_1$ and $\Pi_2$, if there exists some plan $\Pi_3$ such that $\Pi_2 = \Pi_1 \cup \Pi_3$ then we say that $\Pi_1$ *is compatible with* $\Pi_2$.

We can now define the notion of the history generated by a joint $k$-plan $\Pi$ at an initial state $s_0$ under the action theory $\gamma$. It is the $\gamma$-compatible $k$-history along which the agents jointly execute the plan $\Pi$ starting at state $s_0$. We write this as $H^{\Pi, s_0, \gamma}$

### 3.6 Linear Temporal Logic

In our model histories are temporal entities that are always finite in length, therefore the most natural choice to describe properties of histories is Linear Temporal Logic over Finite Traces [7, 8]. This allows us to describe temporal properties such as "$\varphi$ never occurs" or "$\varphi$ always occurs immediately after $\psi$". We write the language as $\mathcal{L}_{\mathsf{LTL}_f}$, defined by the following grammar:

$$\varphi \quad ::= \quad p \mid do(i, a) \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{X}\varphi \mid \varphi \,\mathsf{U}\, \varphi,$$

with $p$ ranging over $Prop$, $i$ ranging over $Agt$ and $a$ ranging over $Act$. Atomic formulas in this language are those that consist of a single proposition $p$ or a single instance of $do(i, a)$. $\mathsf{X}$ and $\mathsf{U}$ are the operators "next" and "until" of $\mathsf{LTL}_f$. Operators "henceforth" ($\mathsf{G}$) and "eventually" ($\mathsf{F}$) are defined in the usual way: $\mathsf{G}\varphi \stackrel{\text{def}}{=} \neg(\top \,\mathsf{U}\, \varphi)$ and $\mathsf{F}\varphi \stackrel{\text{def}}{=} \neg\mathsf{G}\neg\varphi$. We define the semantics for $\mathsf{X}$ and $\mathsf{U}$ as follows, the rest is the same as $\mathcal{L}_{\mathsf{PL}+}$ (for $t \in \{0, \ldots, k\}$).

$$\begin{aligned} H, t \models \mathsf{X}\varphi \quad &\iff t < k \text{ and } H, t+1 \models \varphi, \\ H, t \models \varphi_1 \,\mathsf{U}\, \varphi_2 \quad &\iff \exists t' \geq t : t' \leq k \text{ and } H, t' \models \varphi_2 \text{ and} \\ &\qquad \forall t'' \geq t : \text{ if } t'' < t' \text{ then } H, t'' \models \varphi_1. \end{aligned}$$

## 4 Formalising Responsibility

In order to define responsibility anticipation, we must first define responsibility attribution. Responsibility attribution is a backward-looking notion where, given some fixed history, we seek to determine which agents are responsible for some particular outcome. We distinguish between "agentive" and merely "causal" forms of responsibility. For an agent $i$ to be causally responsible for some outcome $\omega$ simply means that the actions of $i$ were in some way a causal factor in the occurrence of $\omega$. Agentive responsibility requires the additional condition that $i$ *knew* that its actions could or would lead to $\omega$.

Another common notion of responsibility is that of moral responsibility, the kind of responsibility that typically merits praise or blame. We do not formalise moral responsibility in this paper as it is an extremely complex notion, and there is widespread disagreement in the literature regarding exactly what the criteria for moral responsibility are [22]. That said, we do believe that agentive responsibility is a necessary (but not sufficient) condition for moral responsibility.

In the interest of space we have omitted or sketched the proofs of our results, which can be found in the full version of this paper [20].

### 4.1 Causal Responsibility

To be causally responsible for an outcome roughly means to have causally contributed to that outcome occurring. Two main notions of causal responsibility are active and passive responsibility. To be actively responsible means to directly cause the outcome, i.e. to act in a way that guarantees the outcome will occur. To be passively responsible means to allow an outcome to occur while having the ability to prevent it. The following definitions of active and passive responsibility are based on the work of Lorini et al. [17] which uses the logic of STIT (Seeing To It That) and which we previously adapted to a multi-agent planning domain [10].

**Definition 2** (Active Responsibility)**.** *Let $\nabla = (\gamma, s_0, (S_i)_{i \in Agt})$ be a PPD, $i \in Agt$ an agent, and $\Pi_1$ a joint plan. Let $\omega \in \mathcal{L}_{\mathsf{LTL}_f}$. Then, we say that $i$ bears* Causal Active Responsibility *(CAR) for $\omega$ in $(\Pi_1, s_0, \gamma)$, if $H^{\Pi_2, s_0, \gamma} \models \omega$ for all $\Pi_2$ compatible with $\Pi_1^{\{i\}}$ and there exists some joint plan $\Pi_3 \in Plan^{Agt}$ such that $H^{\Pi_3, s_0, \gamma} \not\models \omega$.*

Where $s_0$ and/or $\gamma$ are obvious from context, they are omitted from the statement "$i$ bears CAR for $\omega$ in $(\Pi_1, s_0, \gamma)$" In words, an agent $i$ is causally actively responsible for the occurrence of $\omega$ if, keeping fixed the initial state and the actions of $i$, the other agents could not have acted differently and prevented the occurrence of $\omega$. Note that active responsibility requires that the outcome does not occur in all possible plans. Therefore an agent cannot be actively responsible the sun rising in the morning, as this is inevitable. This corresponds to the deliberative STIT operator of Horty and Belnap [13].

**Definition 3** (Passive Responsibility)**.** *Let $\nabla = (\gamma, s_0, (S_i)_{i \in Agt})$ be a PPD, $i \in Agt$ an agent, and $\Pi_1$ a joint plan. Let $\omega \in \mathcal{L}_{\mathsf{LTL}_f}$. Then, we say that $i$ bears* Causal Passive Responsibility *(CPR) for $\omega$ in $(\Pi_1, s_0, \gamma)$ if $H^{\Pi_1, s_0, \gamma} \models \omega$ and there exists some $\Pi_2$ compatible with $\Pi_1^{-\{i\}}$ such that $H^{\Pi_2, s_0, \gamma} \not\models \omega$.*

An agent $i$ is passively responsible for some outcome $\omega$ if, keeping fixed the initial state and the actions of all other agents, it could have acted differently and prevented the occurrence of $\omega$. In STIT terms this can be writtern as "$\omega \wedge \neg STIT_{Agt \setminus \{i\}} \omega$".

Passive and active responsibility fail in cases of causal overdetermination. For example: suppose three agents push a car off a cliff. Since the car is heavy, two of them are needed to successfully push the car, meaning no agent is actively responsible. Since any agent could have stopped pushing without changing the outcome, none of them are passively responsible. Therefore, we introduce the notion of contributive responsibility based on the work of Braham and van Hees [4], which is a more general notion of causal responsibility.

**Definition 4** (Contributive Responsibility)**.** *Let $\nabla = (\gamma, s_0, (S_i)_{i \in Agt})$ be a PPD, $i \in Agt$ an agent, and $\Pi_1$ a joint plan. Let $\omega \in \mathcal{L}_{\mathsf{LTL}_f}$. Then, we say that $i$ bears* Causal Contributive Responsibility *(CCR) for $\omega$ in $(\Pi_1, s_0, \gamma)$ if $H^{\Pi_1, s_0, \gamma} \models \omega$ and there exists some coalition of agents $J$ such that $i \in J$ and for all $\Pi_2$ compatible with $\Pi_1^J$, $H^{\Pi_2, s_0, \gamma} \models \omega$ and there exists some $\Pi_3$ compatible with $\Pi_1^{J \setminus \{i\}}$ such that $H^{\Pi_3, s_0, \gamma} \not\models \omega$.*

In words, an agent $i$ is contributively responsible for $\varphi$ if it is part of some coalition of agents $J$ such that: a) the actions of $J$ were sufficient to guarantee $\varphi$; and b) the actions of $J \setminus \{i\}$ were not sufficient to guarantee $\varphi$. In terms of STIT this can be written as "$\exists J \subseteq Agt : i \in J \wedge STIT_J \omega \wedge \neg STIT_{J \setminus \{i\}} \omega$".

A notable property of Causal Contributive Responsibility is that it is "complete". This means that for any outcome that occurs in a plan, if that outcome was not guaranteed there is at least one agent who is responsible (i.e. bears CCR) for that outcome.

**Theorem 1.** *Let $\nabla = (\gamma, s_0, (S_i)_{i \in Agt})$ be a PPD, let $\Pi$ be a joint plan and let $H = H^{\Pi, s_0, \gamma}$. Let $\omega \in \mathcal{L}_{\mathsf{LTL}_f}$ such that $H \models \omega$.*

*Then either $H' \models \omega$ for every history $H'$ compatible with $\nabla = (\gamma, s_0, (S_i)_{i \in Agt})$ (meaning $H'$ is of the form $H^{\Pi', s_0, \gamma}$ for some joint plan $\Pi'$), or there exists some $i \in Agt$ such that $i$ bears CCR for $\omega$ in $\Pi$.*

An important property of our notions of responsibility is that no agent can be held in any way causally responsible for an outcome that was inevitable (i.e. occurs in every possible joint plan). This is because all three notions of responsibility require the existence of a joint plan where $\omega$ does not occur.

## 4.2 Agentive Responsibility

To bear agentive responsibility for an outcome, an agent must know that their actions will (or in some cases may) be causally responsible for the outcome occurring. Specifically, we consider the epistemic state of the agent where they have decided their own actions, but do not yet know the actions of others.

**Definition 5** (Agentive Active Responsibility)**.** *Let $\nabla = (\gamma, s_0, (S_i)_{i \in Agt})$ be a PPD, $i \in Agt$ an agent, and $\Pi_1$ a joint plan. Let $\omega \in \mathcal{L}_{\mathsf{LTL}_f}$. Then, we say that $i$ bears* Agentive Active Responsibility *(AAR) for $\omega$ in $(\Pi_1, s_0, \gamma)$ if $i$ is actively responsible for $\omega$ in $\Pi_1$ and for every $\Pi_2$ compatible with $\Pi_1^{\{i\}}$, and every $s_1 \in S_i$, $H^{\Pi_2, s_1, \gamma} \models \omega$.*

Agent $i$ bears agentive active responsibility for $\omega$ if their actions were sufficient to guarantee $\omega$ in any possible outcome (given the possible start states and possible actions of other agents). Furthermore, as with CAR, there must be some joint plan from $s_0$ where $\omega$ does not occur.

Since passive and contributive responsibility both include the notion of "allowing" something to happen rather than "forcing" it to happen, the outcome does not need to be guaranteed from the perspective of the agent, but merely possible. This means that the notions of agentive passive and contributive responsibility are both equivalent to their causal versions, as we assume that agents have full knowledge of the action theory and consider the true initial state to be epistemically possible, meaning any actual outcome $\omega$ must have been considered possible from the perspective of every agent. Therefore note that the acronyms CPR and CCR refer to both the causal *and* agentive variants of passive and contributive responsibility.

A more intuitive notion of agentive passive and contributive responsibility would be to say that $\omega$ must be *reasonably likely* from the perspective of $i$ rather than merely "possible". However, since our model contains no notion of probability, plausibility, or knowledge of the actions of other agents, this is not currently possible, though it does present a direction for future iterations of this model.

**Example 2** (Crossing a Junction - continued)**.** *Consider the following joint plan $\Pi_1$ from start state $s_0$:*

$$A1 : [0 \mapsto Move, 1 \mapsto Move], A2 : [0 \mapsto Move, 1 \mapsto Move]$$

*This will result in a collision. Agent 1 bears CPR (and also CCR) for this outcome ($\omega_1 = \mathsf{F} collision$) since in this case A1 could have avoided a collision by waiting for one step before moving (i.e. $A1 : [0 \mapsto Skip, 1 \mapsto Move]$). However, since Agent 2 also could have waited to avoid a collision, Agent 1 is not actively responsible. Consider an alternative plan where each agent is more cautious:*

$$A1 : [0 \mapsto Skip, 1 \mapsto Skip], A2 : [0 \mapsto Skip, 1 \mapsto Skip]$$

*In this case Agent 1 bears CAR and AAR for the failure (negation) of the goal "Agent 1 eventually crosses the road" ($\omega_2 = \mathsf{F} crossed_1$), since $\neg\omega_2$ occurs in any history compatible with the actions of Agent 1 in $\Pi_2$ starting from $s_0$ or $s_1$.*

## 4.3 Anticipating Responsibility

Responsibility attribution is defined on known joint plans and known initial states. Therefore it cannot be used in planning for single agents, who lack knowledge about the actions of other agents and the initial state. However, an agent can always know if it is *potentially* responsible for that outcome, namely if there is some possible history compatible with that plan where they are responsible.[1]

**Definition 6** (Anticipated Responsibility)**.** *Let $\nabla = (\gamma, s_0, (S_i)_{i \in Agt})$ be a PPD, $i \in Agt$ an agent, and $\Pi$ an individual plan. Let $\omega \in \mathcal{L}_{\mathsf{LTL}_f}$ and $X$ a form of responsibility (CAR, CPR, CCR, AAR). Then, we say that $i$ anticipates $X$ for $\omega$ in $(\Pi, \nabla)$ if there is some $s_1 \in S_i$ and some joint plan $\Pi_1$ compatible with $\Pi$ such that is $i$ bears X for $\omega$ in $(\Pi_1, s_1)$.*

For example, if there is some possible history $H$ compatible with the agent $i$'s individual plan $\Pi$, $i$'s knowledge of the start state and the action theory $\gamma$, such that the agent bears causal passive responsibility in $H$, then the agent anticipates causal passive responsibility in $\Pi$. Since the "actual" history that occurs is not relevant to the determination of anticipated responsibility, this can be done before plan execution, making responsibility anticipation potentially useful in plan selection. We will now show the logical implications between our different forms of responsibility:

**Theorem 2.** *The implications shown in figure 2 are correct.*

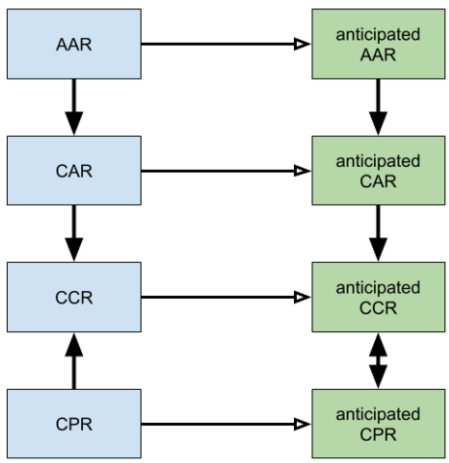


**Figure 2.** A visual representation of the implications between our different forms of responsibility. The horizontal arrows indicate that in any joint plan $\Pi$ where $i$ is attributed some form of responsibility, $i$ can anticipate that form of responsibility in the individual plan $\Pi^{\{i\}}$. A vertical arrow from box $X$ to box $Y$ indicates that in any plan where $i$ is attributed/anticipates X, $i$ also is attributed/anticipates Y.

[1] We could also consider anticipation with universal instead of existential quantification, but being responsible in *every* possible history is a very strong notion and we have not found much use for it.

### 4.4 Responsibility Anticipation in Plan Selection

As previously stated, our hypothesis is that anticipating responsibility can help agents to coordinate towards a common goal, even without communication. Given some goal or value $\varphi$, agents should avoid active responsibility for $\neg\varphi$. This means performing a plan that does not anticipate AAR for $\neg\varphi$. Furthermore, we prove that there is always a plan that does not anticipate AAR for $\neg\varphi$. This means that artificial agents can be formally verified to never be potentially actively responsible for the violation of some value. This could be a useful step in creating provably safe autonomous planning agents.

**Theorem 3.** *Let $\nabla = (\gamma, s_0, (S_i)_{i\in Agt})$ be a PPD, $i \in Agt$, and $\omega$ an $\mathsf{LTL}_f$-formula. Then there exists some individual plan $\Pi$ for $i$ such that $i$ does not anticipate AAR for $\omega$ in $\Pi$.*

*Proof.* (sketch) Either there is some compatible history where $\omega$ does not occur (meaning $i$ does not anticipate AAR) or $\omega$ occurs in every outcome of every plan, so $i$ cannot be responsible. □

Given some value or goal $\varphi$, we want agents to avoid responsibility for $\neg\varphi$, but also to seek responsibility for $\varphi$ (preferably agentive active responsibility, as this guarantees the occurrence of $\varphi$). However, we can show that anticipating agentive active responsibility for $\varphi$ is effectively equivalent to not anticipating causal passive responsibility for $\neg\varphi$ (the dual notion of anticipating CPR).

**Theorem 4.** *Let $\nabla = (\gamma, s_0, (S_i)_{i\in Agt})$ be a PPD, $i \in Agt$, and $\omega$ an $\mathsf{LTL}_f$-formula. If there is some plan $\Pi$ for $i$ such that $i$ anticipates AAR for $\omega$ in $\Pi$, then for any plan $\Pi'$ for $i$, $i$ does not anticipate CPR for $\neg\omega$ in $\Pi'$ if and only if $i$ anticipates AAR for $\omega$ in $\Pi'$.*

*Proof.* (sketch) Given a joint plan $\Pi$ in a planning domain $\nabla$, $i$ is "powerless" with respect to $\omega$ if no alternative plan for $i$ changes the truth value of $\omega$ in $H^{\Pi,s_0,\gamma}$. If $\omega$ occurs in all plans where $i$ is powerless then for all plans $\Pi'$ for $i$, $i$ does not anticipate CPR for $\neg\omega$ in $\Pi'$ if and only if $i$ anticipates AAR for $\omega$ in $\Pi'$. Otherwise, there is no plan $\Pi'$ where $i$ anticipates AAR for $\omega$ in $\Pi'$. □

By "effectively equivalent" we mean that if there exists some plan $\Pi$ for $i$ that anticipates AAR for $\varphi$, then the plans that anticipate AAR for $\varphi$ are exactly the plans that do not anticipate CPR for $\neg\varphi$. However, the notions are not logically equivalent because it is possible that there are some plans for $i$ that do not anticipate CPR for $\neg\varphi$ while there are none that anticipate AAR for $\varphi$.

This also suggests that anticipated CPR is the most important notion of anticipated responsibility, as it is either equivalent or effectively equivalent to every other notion of anticipated responsibility. Finally, we can show that avoiding CPR for $\neg\varphi$ is a potentially powerful method for allowing a group of agents to coordinate on a certain goal, even if those agents cannot communicate.

**Theorem 5.** *Let $\nabla = (\gamma, s_0, (S_i)_{i\in Agt})$ be a PPD and $\omega$ an $\mathsf{LTL}_f$-formula. Let $\Pi$ be a joint plan such that for every agent $i \in Agt$, $i$ does not anticipate CPR for $\neg\omega$ in $\Pi^{\{i\}}$. Then either $H' \models \neg\omega$ for every history compatible with $\nabla$, or $H^{\Pi,s_0,\gamma} \models \omega$.*

*Proof.* (sketch) Suppose for contradiction that $\omega$ occurs in some plan $\Pi'$ and does not occur in $\Pi$. Then by Theorem 1 there is some agent $i$ who bears CCR for $\neg\omega$ in $\Pi$. Then by Theorem 2 it must be the case that $i$ anticipates CPR for $\neg\omega$ in $\Pi^{\{i\}}$, which is a contradiction. □

This shows that even when agents with a shared goal cannot communicate and when no agent can individually guarantee the success of the goal, the application of anticipated responsibility can allow the agents to successfully coordinate their actions and achieve the goal. For instance, in the car example, if each agent chooses to remain stationary (*Skip*) then there will be no collisions. Furthermore, this holds even at a more complex junction with more than two cars, where no driver can individually guarantee that no collision will take place, but each can guarantee not being responsible for a collision by not moving.

## 5 Computing and Implementing Responsibility

In this section we outline a possible implementation of our work in PDDL, and give some foundational complexity results.

### 5.1 PDDL Implementation

As previously mentioned, our model is designed to be practically useful in real-world planning problems. Therefore we outline how our model can be implemented in the multi-agent extension of PDDL 3.1 proposed by Kovacs [15].

PDDL solvers take two inputs: a domain and a problem. The domain gives the object types, actions and predicates, whereas the problem gives the objects, initial state and goal. Below is some simplified PDDL code for a multi-agent planning domain involving a number of immobile agents and some tables that is inspired by the example of Kovacs [15]. The agents can lift tables that they are next to, or do nothing (*Skip*). Our example involves two tables (table1 and table2) and two agents (A1 and A2).

```
(define (domain responsibility-attribution)

(:requirements :equality :negative preconditions :
    typing :multi-agent)
(:types agent table)
(:predicates (lifted ?o - object) (at ?a - agent ?o
     object))
(:action lift :agent ?a - agent :parameters (?o -
    object)
:precondition (and(not(lifted ?o))
                  (at ?a ?o))
:effect (lifted ?o))

(:action skip :agent ?a - agent :parameters ()
:precondition ()
:effect ()))
```

**Listing 1.** Example PDDL Domain

Consider the history where each agent starts next to a separate table, A1 performs the action *Skip* and A2 performs *Lift*. The following code illustrates how we can use PDDL to check if A1 bears CPR for $\omega = \neg\mathsf{FG}(\text{lifted table1} \wedge \text{lifted table2})$.

Running the first problem checks if $\omega$ actually occurs, the second problem fixes the actions of all agents besides A1 and checks if A1 could have acted differently and avoided $\omega$. If a plan is found, then A1 bears CPR for $\omega$ as A1 will fulfil the conditions in definition 3.[2]

```
(define (problem causal-passive-responsibility-1)
(:domain responsibility-attribution)
(:objects A1 A2 - agent table1 table2 - table)
(:init (at A1 table1) (at A2 table2))
(:goal (and (lifted table1) (lifted table2)
            (do(A1 skip 1))(do(A2 lift 1)))))
```

**Listing 2.** Checking that the outcome occurs.

[2] Note that most code in this section includes the actions for some agents, as responsibility anticipation and attribution are performed relative to the actions of some or all agents. This means that a real-world application of anticipation or attribution could not be written entirely in PDDL, and would require another program to edit the PDDL problem before solving it.

```
(:goal (and (lifted table1) (lifted table2) (do(A2
    lift 1)))))
```

**Listing 3.** Checking for Causal Passive Responsibility

To describe the plans of agents in PDDL goals we use $do(i, a, t)$ which is true whenever agent $i$ does action $a$ at time $t$.[3]

In terms of the outcomes that we can attribute or anticipate responsibility for, PDDL 3 supports any boolean combination of predicates as goals, and also features temporal operators for $\mathsf{LTL}_f$ outcomes [9]. However, since PDDL does not support nesting of temporal operators, we do not have the full expressiveness of $\mathsf{LTL}_f$.

The following problems demonstrate how to check CAR for A1 and $\omega$. Firstly, we have to check if $\omega$ is inevitable by attempting to find a joint plan that achieves $\neg\omega$. Then we have to check if the actions of A1 are sufficient to guarantee $\omega$.

```
(:goal (and (lifted table1) (lifted table2))))

(:goal (and (lifted table1) (lifted table2) (do(a
    skip 1)))))
```

**Listing 4.** Checking inevitability and responsibility for CAR attribution

For checking AAR we first have to follow the procedure for checking CAR, but then we also have to check that the actions of A1 are sufficient to guarantee $\omega$ in every epistemically possible world for A1. In this example we will suppose that $(S_{A1}) = \{\{$at A1 table1, at A2 table2$\}, \{$at A1 table1, at A2 table1$\}\}$ modelling that A1 does not know where A2 is.

```
(define (problem causal-active-responsibility-2)
(:domain responsibility-attribution)
(:objects a b - agent table1 table2 - table)
(:init (at a table1) (at b table1))
(:goal (and (lifted table1) (lifted table2) (do(a
    skip 1)))))
```

**Listing 5.** AAR Attribution

For anticipating CAR or AAR the process is much the same as attribution, since attribution depends only on the actions of A1, meaning the actions of all other agents do not need to be defined. We simply have to repeat the procedure for CAR or AAR attribution once for each epistemically possible start state. If A1 bears CAR/AAR in any start state, then they anticipate CAR/AAR. The process for anticipating CPR is more complex. This is because we need to find start state and a plan for all agents besides A1 such that the intended plan for A1 leads to $\omega$ but there exists some other plan for A1 that leads to $\neg\omega$.

```
(define (problem causal-active-responsibility-2)
(:domain responsibility-attribution)
(:objects a b a-1 b-1 - agent table1 table2 table1
    -1 table2-1- table)
(:init (at a table1) (at b table2) (at a-1 table1
    -1) (at b-1 table2-1))
(:goal (and (lifted table1) (lifted table2)
            (do(a skip 1))
            (not (and (lifted table1-1)
                      (lifted table2-1)))
            (henceforth (and (do(b skip) ->
                               do(b-1 skip))
                             (do(b lift)->
                               do(b-1 lift)))))))
```

**Listing 6.** CPR Anticipation

This can be solved in a single planning problem (at least, one problem per possible start state) by creating a duplicate copy of each object, allowing us to effectively run two copies of the planning domain in parallel, with the goal enforcing that the actions of all agents besides A1 must be the same in both copies. If a plan is found for any possible start state, then A1 anticipates CPR for $\omega$.

The procedure for attributing CCR is even more complex, as which agent's actions we have to fix varies depending on which coalition we are testing, and there are exponentially many coalitions to check. Fortunately, since anticipated CCR is equivalent to anticipated CPR (Theorem 2), the procedure for checking that is relatively straightforward.

### 5.2 Complexity Analysis

In this section we will demonstrate the computational complexity of determining various kinds of responsibility. Full proofs of our results can be found in the full version of this paper [20]. We define X-ATTRIBUTION as the problem of determining if $i$ bears X$\in$ {CAR,CPR,CCR,AAR} for $\omega$ in $\Pi$ and X-ANTICIPATION as the problem of determining if $i$ *anticipates* X for $\omega$ in $\Pi$.

**Theorem 6.** *CAR-ATTRIBUTION is a member of $P^{NP[2]}$, CPR-ATTRIBUTION is NP-Complete, CCR-ATTRIBUTION is a member of $\Sigma_2^P$ and AAR-ATTRIBUTION is a member of $\Delta_2^P$.*

**Theorem 7.** *CAR-ANTICIPATION is a member of $\Delta_p^2$, CPR-ANTICIPATION is NP-Complete, CCR-ANTICIPATION is NP-Complete, and AAR-ANTICIPATION is a member of $\Delta_2^P$,.*

These results are only intended to give an introduction to the complexity analysis of this setting. Two problems that deserve further study are the task of identifying if a plan exists that does/does not anticipate responsibility for some outcome $\omega$ (decision problem) and finding such a plan if one exists (search problem). The problem of identifying if a CPR-anticipating plan exists should be NP-complete given Theorem 7, as NP allows us to simply guess a plan, and then check for anticipated responsibility. This puts us in line with the computational complexity of single-agent planning with propositional goals, which is also NP-complete [24].

## 6 Conclusions and Future Work

In this paper we have presented our model for responsibility attribution and anticipation in a multi-agent planning setting with partial information regarding the initial state. We have presented both causal and agentive versions of active, passive and contributive responsibility. We have demonstrated how our notions of anticipated responsibility could be useful for plan selection in a multi-agent setting, and have given a complexity analysis of our model. Finally, we have outlined a PDDL implementation of our model.

For future work, a full PDDL implementation would allow us to test how useful our concepts of responsibility are when applied to real-world planning problems. Furthermore, we could expand our notions of responsibility to handle additional factors. Some interesting extensions are including beliefs about the likely actions of other agents, in line with Lorini [17], or considering intentions, probabilities and/or degrees of responsibility, in line with Halpern and Kleiman-Weiner [12]. Finally, since agents may have multiple goals or values that they may be held responsible for satisfying or violating, it would be useful to extend our model to allow plan comparison based on anticipated responsibility for multiple different outcomes as outlined in our previous work [10].

[3] We do not define $do(i, a, t)$ here as its definition is quite complex and straightforward, it can be found in the full version of the paper [20].

## Acknowledgements

Support from the ANR-3IA Artificial and Natural Intelligence Toulouse Institute (ANITI) is gratefully acknowledged.

## References

[1] N. Alechina, J. Halpern, and B. Logan, 'Causality, responsibility and blame in team plans', in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, (AAMAS)*, (2017).
[2] C. Baier, F. Funke, and R. Majumdar, 'A game-theoretic account of responsibility allocation', in *Proceedings of the 30th International Joint Conference on Artificial Intelligence, (IJCAI)*, (2021).
[3] M. Bienvenu, C. Fritz, and S. Mcilraith, 'Planning with qualitative temporal preferences.', in *Proceedings of the 10th International Conference on Principles of Knowledge Representation and Reasoning, (KR)*, (2006).
[4] M. Braham and M. van Hees, 'An anatomy of moral responsibility', *Mind*, **121**(483), (2012).
[5] A. Camacho, E. Triantafillou, C. J. Muise, J. A. Baier, and S. A. McIlraith, 'Non-deterministic planning with temporally extended goals: LTL over finite and infinite traces', in *Proceedings of the 31st Conference on Artificial Intelligence, (AAAI)*, (2017).
[6] H. Chockler and J. Halpern, 'Responsibility and blame: A structural-model approach', *Journal of Artificial Intelligence Research*, **22**, (2004).
[7] G. De Giacomo and M. Vardi, 'Linear temporal logic and linear dynamic logic on finite traces', in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*, (2013).
[8] G. De Giacomo and M. Vardi, 'Synthesis for LTL and LDL on finite traces', in *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, (2015).
[9] A. Gerevini and D. Long, 'Plan constraints and preferences in PDDL3', *Proceedings of the International Conference on Automated Planning & Scheduling (ICAPS)*, (2005).
[10] Umberto Grandi, Emiliano Lorini, and Timothy Parker, 'Moral planning agents with ltl values', in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI*, (2023).
[11] J. Halpern, 'A modification of the halpern-pearl definition of causality', in *Proceedings of the 24th International Joint Conference on Artificial Intelligence, (IJCAI)*, (2015).
[12] J. Halpern and M. Kleiman-Weiner, 'Towards formal definitions of blameworthiness, intention, and moral responsibility', in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence, (AAAI)*, (2018).
[13] J. Horty and N. Belnap, 'The deliberative stit: A study of action, omission, ability, and obligation', *Journal of Philosophical Logic*, **24**(6), (1995).
[14] Robert M. Keller, 'Formal verification of parallel programs', *Communications of the ACM*, **19**(7), (1976).
[15] D. L. Kovacs, 'A multi-agent extension of PDDL3.1', in *Proceedings of the 3rd Workshop on the International Planning Competition (IPC), 22nd International Conference on Automated Planning and Scheduling, (ICAPS)*, (2012).
[16] E. Lorini, 'Rethinking epistemic logic with belief bases', *Artificial Intelligence*, **282**, (2020).
[17] E. Lorini, D. Longin, and E. Mayor, 'A logical analysis of responsibility attribution: emotions, individuals and collectives', *Journal of Logic and Computation*, **24**(6), (2014).
[18] P. Naumov and J. Tao, 'Blameworthiness in strategic games', in *The 33rd Conference on Artificial Intelligence, (AAAI)*, (2019).
[19] P. Naumov and J. Tao, 'Two forms of responsibility in strategic games', in *Proceedings of the 30th International Joint Conference on Artificial Intelligence, (IJCAI)*, (2021).
[20] Timothy Parker, Umberto Grandi, and Emiliano Lorini, 'Anticipating responsibility in multiagent planning', in *arXiv:2307.16685*, (2023).
[21] R. Reiter, *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*, The MIT Press, 07 2001.
[22] M. Talbert, 'Moral Responsibility', in *The Stanford Encyclopedia of Philosophy*, (2022).
[23] A. Torreño, E. Onaindia, and O. Sapena, 'An approach to multi-agent planning with incomplete information', in *Proceedings of the 20th European Conference on Artificial Intelligence, (ECAI)*, (2012).
[24] Hudson Turner, 'Polynomial-length planning spans the polynomial hierarchy', in *Proceedings of Logics in Artificial Intelligence, European Conference, (JELIA)*, (2002).
[25] I. van de Poel, 'The relation between forward-looking and backward-looking responsibility', in *Moral Responsibility: Beyond Free Will and Determinism*. Springer Netherlands, (2011).
[26] Shufang Zhu, Lucas M. Tabajara, Jianwen Li, Geguang Pu, and Moshe Y. Vardi, 'Symbolic ltlf synthesis', in *Proceedings of the 26th International Joint Conference on Artificial Intelligence, (IJCAI)*, (2017).