Proceedings of the 23rd International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Reasoning about Welfare-Affecting Capabilities in Concurrent Games

Yinfeng Li , Emiliano Lorini
IRIT, CNRS, University of Toulouse, France
{yinfeng.li, emiliano.lorini}@irit.fr

Abstract

We extend the languages of Coalition Logic (CL) and Alternating-time Temporal Logic (ATL) with new modalities for welfare-affecting capabilities, capturing the benefit and harm that a coalition can bring to another coalition through its strategic choices. These languages are interpreted over concurrent games enriched with agents' preferences. On the conceptual side, we use these languages to formalize the notions of potential benefactor and danger. A potential benefactor corresponds to a coalition having the capability to confer a benefit on another coalition. Danger corresponds to a coalition having the capability to inflict harm on another coalition. On the technical side, we present several results concerning their axiomatization, their relationships with standard CL and ATL, and their computational complexity.

1 Introduction

Understanding how agents can influence each other's outcomes in strategic settings is a central problem in multi-agent systems. In many practical scenarios, agents or coalitions act not only to achieve their own goals but also to affect the welfare of others, either positively or negatively. This motivates the formal study of notions, such as danger and potential benefactor, that capture both the benefits and harms that coalitions can generate through their strategic choices.

While the general theory of danger assumes that dangerousness can be ascribed not only to agents but also to events and situations, in this paper we focus on the dangerousness of a coalition relative to another coalition. Intuitively, a *danger* corresponds to a coalition that can inflict harm on another coalition by selecting a particular strategy. In this sense the first coalition is a danger for the second coalition. Its positive counterpart *potential benefactor* corresponds to a coalition that can confer a benefit on another coalition through its strategic choices. In this sense the first coalition is a potential benefactor for the second coalition. These notions allow us to reason explicitly about the effects of coalition strategies on the welfare of others, capturing scenarios in which outcomes are not merely neutral but have qualitative significance for other agents.

Our analysis aligns with two families of philosophical theories concerning harm (and benefit) and danger. On the one hand, we share with the so-called *non-comparative causal account of harm* (Harman 2009; Bradley 2012) the idea that an agent or coalition harms another agent or coalition when the former brings about a bad outcome for the latter. Conversely, it is beneficial when it contributes to securing a good outcome for the latter. The non-comparative account is traditionally contrasted with the *comparative counterfactual account*, according to which an event or action is harmful to someone insofar as it makes things go worse for them, overall, than they would have been if the event or action had not occurred.

On the other hand, we share with the *modal non-probabilistic account of danger* (Fassio 2025; Garland 2003; Peschard, Benétreau-Dupin, and Wessels 2022) the idea that an agent or coalition constitutes a danger for another agent or coalition when the former has the potential to harm the latter. The modal account is traditionally opposed to the probabilistic account, which identifies danger with risk and contrasts it with safety. Thus, while the probabilistic account conceives of danger as a high *probability* of harm, the modal account focuses instead on mere *possibility*. In particular, we interpret *potential* as the existence of a strategy available to a coalition that, if chosen, would bring harm to another coalition.[1]

The notions of good and bad are central to our analysis and, more broadly, to the theory of harm, benefit, danger, and potential benefactor. In this work, we assume that these two notions can be defined in terms of agents' preferences: an outcome is good (resp. bad) for an agent when it is one that the agent prefers (resp. disvalues).

To formalize this rich conceptual framework, we build on concurrent games enriched with agents' preferences over computations, interpreted as long-term outcomes. We employ this class of structures to define a novel logical language based on Alternating-time Temporal Logic (ATL), which enables reasoning about the beneficial (resp. harmful) strategic capabilities of coalitions. We refer to these as the WA$-$ATL language, denoting the extension of ATL with *welfare-affecting capabilities*. It rests on the assumption that a coalition positively (resp. negatively) affects the welfare of another coalition when the former brings benefit (resp. harm) to the latter through the choice of a particular

[1] In (Williamson 2009), a counterfactual interpretation of this notion of potential or possibility is proposed. We leave for future work a comparison between our analysis and Williamson's counterfactual theory of danger.

strategy. Overall, this framework provides a rigorous basis for representing and reasoning about subtle inter-coalitional relationships, capturing the potentially beneficial or harmful influence that one coalition may exert over another.

We present technical results concerning the axiomatizability and computational complexity of the proposed language. In particular, the main technical contributions of the paper are twofold.

First, we provide an axiomatization of the until-free fragment (also called safety fragment) of the WA−ATL language with respect to the general class of concurrent game models with *regular preferences*, together with a strong notion of beneficial (resp. harmful) strategy. This notion is based on the idea that benefiting (resp. harming) a coalition means guaranteeing its best (resp. worst) outcome. By *regular*, we mean that an agent's preference ordering over computations is stable over time, non-flat, and both well-founded and conversely well-founded. Our completeness proof relies on a novel technique based on the notion of a canonical model under formula context.

Second, we establish a polynomial embedding of the WA−ATL language into standard ATL, under the regular preference assumption, as well as a polynomial embedding of its coalition logic (CL) fragment into standard CL. Thanks to these embeddings, we obtain tight complexity results for satisfiability checking.

The semantics we employ is game-theoretic, as it is based on the notion of concurrent games. However, a key limitation of the game-theoretic approach, compared to our logical framework, is that while it enables the representation of welfare-affecting strategic capabilities, it does not support reasoning about them. This limitation arises because game theory operates solely at the semantic level and does not distinguish between semantics and the level of a logical language. In contrast, our approach allows not only the representation of these concepts, through the semantics of concurrent game frames with regular preferences, but also reasoning about them via a logical language interpreted over this semantic structure.

The paper is organized as follows. Section 2 discusses related work. Section 3 introduces the formal semantics. In Section 4, we use this semantics to formalize the notions of beneficial and harmful strategies. Section 5 presents the WA−ATL language. Section 6 is devoted to the axiomatization of its until-free fragment. Section 7 provides the polynomial embeddings and the complexity results. Finally, Section 8 concludes the paper. Proofs are provided in the extended version of this paper available at https://hal.science/hal-05609815.

2 Related Work

On the formal semantics side, our work draws on two main traditions. On the one hand, it builds upon the concurrent game semantics traditionally used to interpret ATL (Alur, Henzinger, and Kupferman 2002; Goranko and van Drimmelen 2006), CL (Pauly 2002; Goranko 2001), strategy logic (SL) (Mogavero et al. 2014; Chatterjee, Henzinger, and Piterman 2007), and STIT (seeing-to-it-that) logic (Broersen, Herzig, and Troquard 2006; Boudou and Lorini 2018). On the other hand, it follows the tradition of modal logics of preferences, including the pure logic of preferences and their dynamics (van Benthem and Liu 2007; Grossi, van der Hoek, and Kuijer 2022), the interaction between preferences and epistemic states (Lorini 2021; Naumov and Ovchinnikova 2023), and higher-order preferences (Jiang and Naumov 2024). However, these works do not address the strategic or temporal dimensions of preferences.

The extension of concurrent games with preferences has recently attracted attention as a means to model agents' rationality (Li, Lorini, and Mittelmann 2025), to capture the game-theoretic concept of iterated elimination of strictly dominated strategies (IESDS) (Liu et al. 2020), and to establish connections between ATL strategic reasoning and correlated equilibrium (Huang and Ruan 2017). In (Li, Lorini, and Mittelmann 2025), a new class of models, *concurrent games with preferences*, was introduced, offering a sophisticated framework for representing agents' preferences over outcomes in concurrent games, as well as their temporal dynamics. We adopt this class of models in the present work. In contrast, (Liu et al. 2020) and (Huang and Ruan 2017) focus on a simpler notion of dichotomous preference induced by an agent's goal, where the goal is expressed either directly at the model level (Huang and Ruan 2017) or as a parameter of the modalities capturing the outcomes of the IESDS procedure (Liu et al. 2020). A quantitative notion of goals, based on a fuzzy interpretation, has recently been incorporated into ATL and SL (Jamroga et al. 2024; Bouyer et al. 2023), and has been used to represent agents' utilities in the context of mechanism design (Mittelmann et al. 2025).

In the present paper, we use concurrent games with preferences introduced in (Li, Lorini, and Mittelmann 2025) to provide a novel logical investigation of the notions of beneficial and harmful strategies, as well as the related concepts of potential benefactor and danger, which have not previously been explored in the context of logics for multi-agent systems.

On the conceptual side, the work most closely related to ours is (Beckers, Chockler, and Halpern 2024), which offers a causal analysis of the concept of harm based on the semantics of structural equation models. As emphasized in the introduction, there is a close connection between the notion of harm and the notion of danger that we investigate in this paper.

3 Semantics

In this section, we present the semantics that will later be used to define the concepts of *potential benefit* and *potential harm*, and to interpret the formal languages expressing these concepts. We begin with the notion of a *multi-agent transition frame*. Next, we introduce *concurrent game frames*, a specific class of multi-agent transition frames that satisfy a set of semantic constraints. Finally, we extend concurrent game frames by incorporating a notion of *preference over computations*. The notion of a concurrent game with preferences is based on (Li, Lorini, and Mittelmann 2025).

3.1 Multi-Agent Transition Frame

Concurrent game models can be defined from a more general class of structures, which we call *multi-agent transition frames* (MATFs). An MATF is parameterized by the set of agents $\mathbb{AGT} = \{1, \ldots, n\}$, with elements denoted $a, b, \ldots$

Definition 1 (MATF)**.** *A multi agent transition frame (MATF) with respect to $\mathbb{AGT}$ is a tuple $\mathsf{F} = (\mathbb{S}, \mathbb{ACT}, \mathcal{R})$, where*

- *$\mathbb{S}$ is a non-empty set of states,*
- *$\mathbb{ACT}$ is a set of actions,*
- *$\mathcal{R} : \mathbb{ACT}^{\mathbb{AGT}} \longrightarrow 2^{\mathbb{S}\times\mathbb{S}}$ assigns a transition relation on $\mathbb{S}$ to every action profile, where $\mathbb{ACT}^{\mathbb{AGT}} = \{\delta_{\mathbb{AGT}} \mid \delta_{\mathbb{AGT}} : \mathbb{AGT} \longrightarrow \mathbb{ACT}\}$ is the set of action profiles.*

Given an MATF as in Definition 1, we call the subsets of $\mathbb{AGT}$ *coalitions*, and we call $\mathbb{AGT}$ the *grand coalition*. For any coalition C, a *joint action* of C is a function $\delta_C : C \longrightarrow \mathbb{ACT}$, and the set of all joint actions of C is denoted by $\mathbb{ACT}^C$. Note that the empty coalition has a unique joint action, $\emptyset$ (also denoted $\delta_\emptyset$), and action profiles correspond to the joint actions of the grand coalition. Joint actions of the grand coalition $\mathbb{AGT}$ are also called *action profiles*.

For any joint action δ_C, we define

$$\mathcal{R}_{\delta_C} = \big\{(s, s') \in \mathbb{S} \times \mathbb{S} \mid \text{there exists } \delta_{\mathbb{AGT}} \in \mathbb{ACT}^{\mathbb{AGT}} \text{ such that } \delta_C \subseteq \delta_{\mathbb{AGT}} \text{ and } (s, s') \in \mathcal{R}(\delta_{\mathbb{AGT}})\big\}.$$

Note that in the previous definition we write $\delta_C \subseteq \delta_{\mathbb{AGT}}$ since in set theory a function is representable as a set of ordered pairs. We write $(s, s') \in \mathcal{R}_{\delta_C}$ as $s\mathcal{R}_{\delta_C}s'$, and define

$$\mathcal{R}_{\delta_C}(s) = \big\{s' \in \mathbb{S} \mid s\mathcal{R}_{\delta_C}s'\big\}.$$

It can be verified that $\mathcal{R}_\emptyset = \bigcup_{\delta_{\mathbb{AGT}} \in \mathbb{ACT}^{\mathbb{AGT}}} \mathcal{R}_{\delta_{\mathbb{AGT}}}$.

In this paper, we treat structures and tuples as mathematical objects and use dot-notation to refer to their components or derived operations. For example, $\mathsf{M}.\mathbb{S}$ denotes the set of all states of structure M, and $t.i$ denotes the i-th element of tuple t.

The following definition introduces the concept of available joint action for a coalition. Intuitively, a joint action is available for a coalition at a given state if the coalition can choose it at that state.

Definition 2 (Available choice)**.** *We say that the joint action δ_C is available for a coalition C at a state s if $\mathcal{R}_{\delta_C}(s) \neq \emptyset$.*

The previous notion of available joint action should be interpreted as what a coalition can choose at a given state: $\big\{\delta_C \in \mathbb{ACT}^C \mid \mathcal{R}_{\delta_C}(s) \neq \emptyset\big\}$ is called coalition C's *choice set* at state s.

Two fundamental concepts in the semantics are those of *path* and *computation*, which are formally defined as follows.

Definition 3 (Path and computation)**.** *A* path *in a MATF $\mathsf{F} = (\mathbb{S}, \mathbb{ACT}, \mathcal{R})$ is a sequence $\lambda : \mathbb{D} \longrightarrow \mathbb{S}$ of states, where $\mathbb{D}$ is a nonempty (possibly infinite) initial segment of the natural numbers, such that*

$$\lambda(k)\mathcal{R}_\emptyset\lambda(k+1) \quad \text{for all } k \in \mathbb{D} \text{ such that } k+1 \in \mathbb{D}.$$

The set of all paths in F is denoted by $Path_\mathsf{F}$. The set of all paths starting at a state $s \in \mathbb{S}$ is

$$Path_{\mathsf{F},s} = \{\lambda \in Path_\mathsf{F} \mid \lambda(0) = s\}.$$

A computation *(or* full path*) in F is a path $\lambda \in Path_\mathsf{F}$ such that there is no path $\lambda' \in Path_\mathsf{F}$ for which λ is a proper prefix. The set of all computations in F is denoted by $Comp_\mathsf{F}$. The set of computations starting at a state $s \in \mathbb{S}$ is*

$$Comp_{\mathsf{F},s} = \{\lambda \in Comp_\mathsf{F} \mid \lambda(0) = s\}.$$

A partial path *is a path that is not a computation. The set of all partial paths in F is denoted by $PPath_\mathsf{F}$, and the set of partial paths starting at state $s \in \mathbb{S}$ is*

$$PPath_{\mathsf{F},s} = \{\lambda \in PPath_\mathsf{F} \mid \lambda(0) = s\}.$$

We denote the domain of a function f by $\mathsf{dom}(f)$, the cardinality of $\mathsf{dom}(\lambda)$ by $\mathsf{len}(\lambda)$, and the restriction of λ to $\{k' \in \mathbb{N} \mid k' \leq k\}$ by $\lambda|_k$. Moreover, for a finite path λ, we let $\mathsf{END}(\lambda)$ denote the largest element of $\mathsf{dom}(\lambda)$, and we abbreviate $\lambda(\mathsf{END}(\lambda))$ as $\mathsf{last}(\lambda)$.

The following fact highlights an interesting property of partial paths, which is guaranteed by the fact that the last state of a partial path always has a successor.

Fact 1. *λ is a partial path iff λ is finite and $\mathsf{last}(\lambda)$ is a non-terminal state, i.e., $\mathcal{R}_\emptyset(\mathsf{last}(\lambda)) \neq \emptyset$.*

The following definition introduces the crucial notion of a *joint strategy* for a coalition in a MATF. We define it as a *total* function from partial paths to joint actions of the coalition. The reason for considering it as total rather than partial is that, in the light of Fact 1, the last state of a partial path always has a successor which is reachable through a joint action of the grand coalition.

Definition 4 (Joint strategy)**.** *A* joint strategy *of a coalition C in a MATF $\mathsf{F} = (\mathbb{S}, \mathbb{ACT}, \mathcal{R})$ is a total function $\mathcal{F}_C : PPath_\mathsf{F} \longrightarrow \mathbb{ACT}^C$ such that, for all $\lambda \in \mathsf{dom}(\mathcal{F}_C)$, $\mathcal{F}_C(\lambda)$ is available at $\mathsf{last}(\lambda)$. The set of all joint strategies of coalition C in F is denoted by Str^C_F.*

Given an initial state s and a joint strategy for a coalition C, we can compute the set of computations generated by this strategy starting at s.

Definition 5 (Generated computations)**.** *Let $\mathsf{F} = (\mathbb{S}, \mathbb{ACT}, \mathcal{R})$ be a MATF, $s \in \mathbb{S}$, and $\mathcal{F}_C \in Str^C_\mathsf{F}$. A computation λ is said to be* generated *by $\mathcal{F}_C$ at s if and only if:*

- *$\lambda(0) = s$,*
- *for all $k \in \mathsf{dom}(\lambda)$, if $k+1 \in \mathsf{dom}(\lambda)$ then $\lambda(k)\mathcal{R}_{\mathcal{F}_C(\lambda|_k)}\lambda(k+1)$.*

The set of all computations generated by $\mathcal{F}_C$ at s is denoted by $\mathcal{O}_\mathsf{F}(s, \mathcal{F}_C)$.

3.2 Concurrent Game Frame

We define concurrent game frames (CGFs) as a subclass of MATFs. In particular, a CGF is a MATF which satisfies the three properties of *never-ending interaction*, *independence of choices* and *collective choice determinism*.

Definition 6 (CGF)**.** *A MATF* $\mathsf{F} = (\mathbb{S}, \mathbb{ACT}, \mathcal{R})$ *is a concurrent game frame (CGF), if it satisfies the following three properties:*

Never-ending interaction: *the relation* $\mathcal{R}_\emptyset$ *over* $\mathbb{S}$ *is serial.*[2]

Independence of choices: *for all states* $s \in \mathbb{S}$*, for all coalitions* C, C' *such that* $C \cap C' = \emptyset$*, and for all joint actions* $\delta_C \in \mathbb{ACT}^C$*,* $\delta_{C'} \in \mathbb{ACT}^{C'}$*, if both* δ_C *and* $\delta_{C'}$ *are available at* s*, then their union* $\delta_C \cup \delta_{C'}$ *is also available at* s*.*

Collective choice determinism: *for every action profile* $\delta_{\mathbb{AGT}}$*, the relation* $\mathcal{R}_{\delta_{\mathbb{AGT}}}$ *over* $\mathbb{S}$ *is deterministic.*[3]

Never-ending interaction captures the fact that every state in a CGF has at least one successor. According to *independence of choices*, agents can never be deprived of available actions due to the choices made by other agents. Finally, *collective choice determinism* ensures that the outcome of an available joint action of the grand coalition is uniquely determined. In other words, once all agents have made their choices the next state is uniquely determined.

In a CGF, computations are infinite due to the *never-ending interaction* property. Moreover, by *collective choice determinism*, the set of computations induced by a strategy of the grand coalition $\mathbb{AGT}$ at a given state is a singleton. We denote this unique computation generated by a strategy $\mathcal{F}_{\mathbb{AGT}}$ at state s by $\wr_{\mathsf{F}}(s, \mathcal{F}_{\mathbb{AGT}})$. Finally, the *independence of choices* property ensures that the merge of strategies of two disjoint coalitions yields a strategy for their union.

3.3 Adding Preferences

We now extend a CGF with notions of preferences and regular preferences over computations, yielding the structures of a *CGF with preferences* (CGFP) and a *CGF with regular preferences* (CGFRP). Concretely, a CGF is extended by associating with each agent and each state a preference ordering over the computations starting from that state. Preferences satisfying stability, non-flatness, well-foundedness, and converse well-foundedness are called *regular preferences*. This notion of preference over computations can be interpreted as a preference over long-term outcomes, since in a CGF the set of long-term outcomes available to an agent from a given state coincides with the set of computations originating from that state.

Definition 7 (CGFP and CGFRP)**.** *A CGF with preferences (CGFP) is a tuple* $\mathsf{P} = (\mathsf{F}, \preceq)$*, where:*

- F *is a CGF, and*
- $\preceq$: $\mathbb{AGT} \times \mathbb{S} \longrightarrow 2^{Comp_{\mathsf{F}} \times Comp_{\mathsf{F}}}$ *assigns to each agent* $i \in \mathbb{AGT}$ *and state* $s \in \mathbb{S}$ *a total preorder over* $Comp_{\mathsf{F},s}$*.*

For any agent i*, state* s*, and computations* λ, λ'*, we write* $\lambda \preceq_{i,s} \lambda'$ *to denote that* $(\lambda, \lambda') \in \preceq (i, s)$*, and read it as "at state* s*, agent* i *prefers computation* λ' *at least as much as computation* λ*".*

[2]This means that $\forall s \in \mathbb{S}, \exists s' \in \mathbb{S}$ such that $s\mathcal{R}_\emptyset s'$.

[3]This means that $\forall s, s', s'' \in \mathbb{S}$, if $s\mathcal{R}_{\delta_{\mathbb{AGT}}} s'$ and $s\mathcal{R}_{\delta_{\mathbb{AGT}}} s''$ then $s' = s''$.

A CGF with regular preferences is a CGFP that satisfies the following four properties, for every agent $a \in \mathbb{AGT}$*:*

$$
\begin{aligned}
(\mathbf{SP})\ & \forall s, s' \in \mathbb{S}, \forall \lambda, \lambda' \in Comp_{\mathsf{F},s'} \text{ if } s\mathcal{R}_\emptyset s', \\
& \text{then } \big(\lambda' \preceq_{a,s'} \lambda \text{ iff } s\lambda' \preceq_{a,s} s\lambda\big), \\
(\mathbf{NFP})\ & \forall s \in \mathbb{S}, \exists \lambda, \lambda' \in Comp_{\mathsf{F},s} \text{ such that } \lambda \prec_{a,s} \lambda', \\
(\mathbf{WFP})\ & \forall s \in \mathbb{S}, \exists \lambda \in Comp_{\mathsf{F},s} \text{ such that} \\
& \forall \lambda' \in Comp_{\mathsf{F},s},\ \lambda \preceq_{a,s} \lambda', \\
(\mathbf{CWFP})\ & \forall s \in \mathbb{S}, \exists \lambda \in Comp_{\mathsf{F},s} \text{ such that} \\
& \forall \lambda' \in Comp_{\mathsf{F},s},\ \lambda' \preceq_{a,s} \lambda,
\end{aligned}
$$

where, for every $\lambda, \lambda' \in Comp_{\mathsf{F},s}$*,* $\lambda \prec_{a,s} \lambda'$ *abbreviates* $\lambda \preceq_{a,s} \lambda'$ *and* $\lambda' \not\preceq_{a,s} \lambda$*.*

Property $\mathbf{SP}$ expresses stability of preferences over time. It captures the idea that an agent's preference remains consistent over time: an agent prefers a computation λ to a computation λ' starting at the same state s' if and only if it prefers the precursor of λ (i.e., $s\lambda$) to the precursor of λ' (i.e., $s\lambda'$) at every predecessor s of s'. It is in line with the notion of *time consistency* of preferences studied in economics, as opposed to *time inconsistency* (Frederik, Loewenstein, and O'Donoghue 2002). Property $\mathbf{NFP}$ ensures agents are not indifferent among all outcomes. Properties $\mathbf{WFP}$ and $\mathbf{CWFP}$ guarantee, respectively, the existence of worst and best outcomes.

3.4 From Frames to Models

We conclude this section by moving from frames to models. In particular, we assume a countably infinite set of atomic propositions $\mathbb{P}$ denoting atomic facts and supplement a CGFRP with a valuation function specifying for every state in the CGFRP which propositions are true there.

Definition 8 (CGMRP)**.** *A concurrent game model with regular preferences (CGMRP) with respect to* $\mathbb{AGT}$ *and* $\mathbb{P}$ *is a tuple* $\mathsf{M} = (\mathsf{P}, \mathcal{V})$*, where*

- P *is CGFRP, and*
- $\mathcal{V} : \mathbb{S} \longrightarrow 2^{\mathbb{P}}$ *is a valuation function.*

We illustrate the notion of CGMRP with the help of a concrete example.

Example 1. *We consider a scenario involving three agents, robots* 0*,* 1*, and* 2*, operating in a small environment composed of three rooms: a left room, a middle room, and a right room. Each of the left and right rooms contains a charging station. Robot* 1 *is positioned in front of the exit of the right room, guarding it to prevent entry, while robot* 2 *is stationed in front of the exit of the left room, performing the same role. Initially, robot* 0 *is located in the middle room (proposition* inM_0*). During the course of the game, it can move and change its position, entering either the right room (proposition* inR_0*) or the left room (proposition* inL_0*). The initial configuration is shown in Figure 1.*

Robot 0 *can move from its current room either to the right (action* ri*) or to the left (action* le*). Thus, to move from the middle room to the right room, it must perform the action* ri*; conversely, to move from the right room to the middle room,*

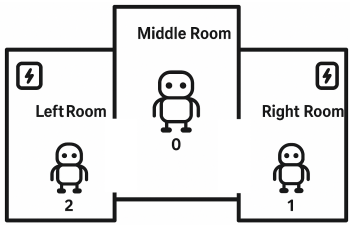
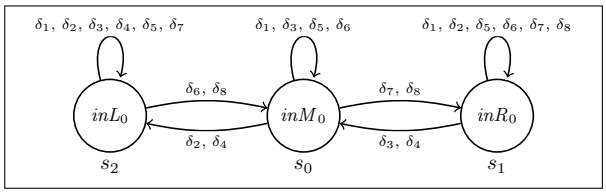


Figure 1: Scenario

it must perform the action le, and so on. If robot 0 is already in the right (resp. left) room and attempts to move further right (resp. left), it remains in place, as it encounters a wall.

Robots 1 and 2 can either block (action bl) or free (action fr) the passage of the room they are guarding. Thus, robot 1 can block or free the passage of the right room, and robot 2 can do the same for the left room. If robot 1 (resp. robot 2) blocks the passage of its respective room, then robot 0 cannot reach the corresponding charging station when it is in the middle room, and cannot leave the room whose passage is blocked when it is located there.

Robot 0's goal is to reach a charging station at some point in the future, while the goal of robots 1 and 2 is to prevent it from doing so. The CGMRP corresponding to this scenario is shown in Figure 2.



(a) State transition diagram.

$$\begin{array}{ll} \delta_1 = (0{:}le, 1{:}bl, 2{:}bl) & \delta_5 = (0{:}ri, 1{:}bl, 2{:}bl) \\ \delta_2 = (0{:}le, 1{:}bl, 2{:}fr) & \delta_6 = (0{:}ri, 1{:}bl, 2{:}fr) \\ \delta_3 = (0{:}le, 1{:}fr, 2{:}bl) & \delta_7 = (0{:}ri, 1{:}fr, 2{:}bl) \\ \delta_4 = (0{:}le, 1{:}fr, 2{:}fr) & \delta_8 = (0{:}ri, 1{:}fr, 2{:}fr) \end{array}$$

(b) Action profiles δ_1–δ_8.

$$\forall s \in \{s_0, s_1, s_2\}, \forall \lambda, \lambda' \in Comp_{\mathsf{F}_{rs},s}, \forall i \in \{0,1,2\},$$
$$\lambda \preceq_{i,s} \lambda \text{ iff } U_0(\lambda') \leq U_0(\lambda) \text{ with}$$
$$U_0(\lambda) = \begin{cases} 0 & \text{if } \forall k \geq 0, \{inR_0, inL_0\} \cap \mathcal{V}(\lambda(k)) = \emptyset, \\ 1 & \text{otherwise;} \end{cases}$$
$$U_1(\lambda) = U_2(\lambda) = \begin{cases} 0 & \text{if } \exists k \geq 0 \text{ such that} \\ & \{inR_0, inL_0\} \cap \mathcal{V}(\lambda(k)) \neq \emptyset, \\ 1 & \text{otherwise.} \end{cases}$$

(c) Utility-induced preference structure.

Figure 2: CGMRP M_{rs} for the robots' scenario: (a) state transitions, (b) action profiles, (c) preference structure.

It consists of three states, s_0, s_1, and s_2, with s_0 as the initial state, and eight possible action profiles, $\delta_1, \ldots, \delta_8$. To make the example self-contained, we assume that each robot's preferences over computations starting from a given state are induced by a dichotomous utility function: this function assigns a value of 1 to a computation that allows the robot to achieve its goal, and a value of 0 otherwise.

4 Beneficial and Harmful Strategy

In this section, we introduce the concepts of beneficial and harmful strategies. The intuition is straightforward: the strategy of a coalition is said to be beneficial (resp. harmful) for another coalition if the choice of that strategy by the first coalition guarantees a good (resp. bad) outcome for the second coalition. We further distinguish between *strongly* and *weakly* beneficial (or harmful) strategies, based on two different interpretations of what constitutes a good or bad outcome for a coalition. A strategy is *strongly beneficial* (resp. *strongly harmful*) for a coalition if it guarantees the best (resp. worst) outcome that the agents in the coalition could obtain. Conversely, a strategy is *weakly beneficial* (resp. *weakly harmful*) for a coalition if it guarantees an outcome that is better (resp. worse) for the agents in the coalition than the one they could obtain if the strategy were not played.

In the semantic interpretation of the welfare-affecting strategic capability modalities we will introduce in Section 5, we will only use the strong notions, since our axiomatization and embedding into standard ATL rely on the notions of best and worst outcomes (whose existence is guaranteed by the regularity assumption on preferences). However, the weak notions are provided here for two reasons: to show that there is no unique way to define beneficial or harmful strategy, and to make the conceptual analysis of beneficial/harmful strategies richer.

The following definition introduces the concepts of strongly beneficial and harmful strategy.

Definition 9 (Strongly beneficial and harmful strategy)**.** *Let $\mathsf{M} = (\mathsf{P}, \mathcal{V})$ be a CGMRP with underlying CGFRP $\mathsf{P} = (\mathsf{F}, \preceq)$. We say a strategy $\mathcal{F}_C$ of coalition C is strongly beneficial for coalition $B \subseteq \mathbb{AGT}$ at the state s in M iff $\mathcal{O}_{\mathsf{F}}(s, \mathcal{F}_C) \subseteq Best_{i,s}$ for every $i \in B$, where $Best_{i,s} = \{\lambda \in Comp_{\mathsf{F},s} \mid \forall \lambda' \in Comp_{\mathsf{F},s}(\lambda' \preceq_{i,s} \lambda)\}$.*

Similarly, we say a strategy $\mathcal{F}_C$ of coalition C is strongly harmful for coalition $H \subseteq \mathbb{AGT}$ at the state s in M iff $\mathcal{O}_{\mathsf{F}}(s, \mathcal{F}_C) \subseteq Worst_{i,s}$ for every $i \in H$, where $Worst_{i,s} = \{\lambda \in Comp_{\mathsf{F},s} \mid \forall \lambda' \in Comp_{\mathsf{F},s}(\lambda \preceq_{i,s} \lambda')\}$.

We denote with $StrSB^C_{\mathsf{M}}(s, B)$ and $StrSH^C_{\mathsf{M}}(s, H)$ the set of strongly beneficial strategies of coalition C for coalition B and the set of strongly harmful strategies of coalition C for coalition H at state s of the CGMRP M, respectively.

Let us now turn to the notion of weakly beneficial and harmful strategies.

Definition 10 (Weakly beneficial and harmful strategy)**.** *Let $\mathsf{M} = (\mathsf{P}, \mathcal{V})$ be a CGMRP with underlying CGFRP $\mathsf{P} = (\mathsf{F}, \preceq)$. We say a strategy $\mathcal{F}_C$ of coalition C is weakly beneficial for coalition $B \subseteq \mathbb{AGT}$ at the state s in M iff $\lambda' \preceq_{i,s} \lambda$ for all $\lambda \in \mathcal{O}_{\mathsf{F}}(s, \mathcal{F}_C)$, $\lambda' \in Comp_{\mathsf{F},s} - \mathcal{O}_{\mathsf{F}}(s, \mathcal{F}_C)$ and $i \in B$.*

Similarly, we say a strategy $\mathcal{F}_C$ of coalition C is weakly harmful for coalition $H \subseteq \mathbb{AGT}$ at the state s in M iff $\lambda \preceq_{i,s} \lambda'$ for all $\lambda \in \mathcal{O}_{\mathsf{F}}(s, \mathcal{F}_C)$, $\lambda' \in Comp_{\mathsf{F},s} - \mathcal{O}_{\mathsf{F}}(s, \mathcal{F}_C)$ and $i \in H$.

We denote with $StrWB^C_{\mathsf{M}}(s,B)$ *and* $StrWH^C_{\mathsf{M}}(s,H)$ *the set of weakly beneficial strategies of coalition* C *for coalition* B *and the set of weakly harmful strategies of coalition* C *for coalition* H *at state* s *of the CGFRP* M*, respectively.*

The intuition behind Definition 10 is that, for a strategy to be weakly beneficial (resp. harmful) for a coalition, the outcome of the strategy should be better (resp. worse) for each member of the coalition than any outcome they could obtain if a different strategy were played.

Let us return to Example 1 presented in Section 3.4 in order to illustrate the preceding definitions.

Example 2 (Cont.)**.** *Consider the following strategy* $\mathcal{F}_{\{1,2\}}$ *for the coalition formed by robots* 1 *and* 2*:*

$$\mathcal{F}_{\{1,2\}}(\lambda) = (1{:}bl, 2{:}bl) \text{ for all } \lambda \in PPath_{\mathsf{F}_{rs}}.$$

This strategy corresponds to blocking the passage of both the right and left rooms indefinitely. It is straightforward to verify that $\mathcal{F}_{\{1,2\}} \in StrSH^{\{1,2\}}_{\mathsf{M}_{rs}}(s_0,\{0\})$*, and moreover,* $\mathcal{F}_{\{1,2\}} \in StrSB^{\{1,2\}}_{\mathsf{M}_{rs}}(s_0,\{1,2\})$*. In words, the strategy* $\mathcal{F}_{\{1,2\}}$ *of coalition* $\{1,2\}$ *is strongly harmful for robot* 0 *at state* s_0*, and strongly beneficial for the coalition made of robots* 1 *and* 2 *at the same state.*

Now consider the following strategy $\mathcal{F}'_{\{1\}}$ *for robot* 1*:*

$$\mathcal{F}'_{\{1\}}(\lambda) = (1{:}bl) \text{ for all } \lambda \in PPath_{\mathsf{F}_{rs}}.$$

This strategy corresponds to robot 1 *blocking the passage of the right room indefinitely. It is straightforward to verify that* $\mathcal{F}'_{\{1\}} \in StrWH^{\{1\}}_{\mathsf{M}_{rs}}(s_0,\{0\})$*, and moreover,* $\mathcal{F}'_{\{1\}} \notin StrSH^{\{1\}}_{\mathsf{M}_{rs}}(s_0,\{0\})$*. In words, the strategy* $\mathcal{F}'_{\{1\}}$ *of robot* 1 *is weakly harmful for robot* 0 *at state* s_0*, but not strongly harmful.*

The following proposition highlights the connection between strongly beneficial/harmful and weakly beneficial/harmful strategies.

Proposition 1. *If a strategy* $\mathcal{F}_C$ *of a coalition* C *is strongly beneficial (resp. harmful) for a coalition at a state, then it is also weakly beneficial (resp. harmful).*

5 Language

In this section, we leverage the semantics presented above to interpret a novel language that extends the language of ATL with a new family of *welfare-affecting strategic capability* modalities. These modalities allow us to reason about the strategic choices of one coalition that affect another coalition, under the assumption that the strategy of one coalition affects another insofar as the first coalition's strategic choice has either a positive or a negative effect on the well-being of the second coalition, by bringing it benefit or causing it harm.

The ATL language with *welfare-affecting capability* modalities, denoted by $\mathcal{L}_{\mathsf{WA-ATL}}$, is defined by the following grammar:

$$\begin{array}{lcl} \text{State formula:} \quad \varphi,\psi & ::= & p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \langle\!\langle C{:}B,H \rangle\!\rangle \gamma \\ \text{Path formula:} \quad \gamma & ::= & \mathsf{X}\varphi \mid \mathsf{G}\varphi \mid (\varphi \,\mathsf{U}\, \psi) \end{array}$$

where p ranges over $\mathbb{P}$, and C, B and H range over $2^{\mathbb{AGT}}$. The other Boolean connectives and constructs $\vee, \rightarrow, \leftrightarrow, \top, \bot$ are defined as abbreviations in the usual way. For simplicity, we sometimes write $\mathcal{L}_{\mathsf{WA-ATL}}$ instead of $\mathcal{L}_{\mathsf{WA-ATL}}(\mathbb{AGT},\mathbb{P})$.

Formulas $\langle\!\langle C{:}B,H \rangle\!\rangle \mathsf{X}\varphi$, $\langle\!\langle C{:}B,H \rangle\!\rangle \mathsf{G}\varphi$, and $\langle\!\langle C{:}B,H \rangle\!\rangle (\varphi \,\mathsf{U}\, \psi)$ capture the notion of a *welfare-affecting strategic capability* to guarantee a certain fact, where this fact is expressed by a temporal formula of LTL (Linear Temporal Logic). In particular, $\langle\!\langle C{:}B,H \rangle\!\rangle \mathsf{X}\varphi$ has to be read "coalition C has a strategy, which is beneficial for group B, harmful to group H, and guarantees that φ is going to be true in the next state"; $\langle\!\langle C{:}B,H \rangle\!\rangle \mathsf{G}\varphi$ has to be read "coalition C has a strategy, which is beneficial for group B, harmful to group H, and guarantees that φ will always be true"; $\langle\!\langle C{:}B,H \rangle\!\rangle (\varphi \,\mathsf{U}\, \psi)$ has to be read "coalition C has a strategy, which is beneficial for group B, harmful to group H, and guarantees that φ will be true until ψ is true".

We can now define the semantic interpretation of formulas in $\mathcal{L}_{\mathsf{WA-ATL}}(\mathbb{AGT},\mathbb{P})$. As in ATL, formulas are interpreted with respect to *pointed models*. A pointed model is a pair consisting of a CGMRP and one of its states, when interpreting a state formula, or a pair consisting of a CGMRP and one of its computations, when interpreting a path formula.

In Section 4, we introduced two notions of beneficial and harmful strategies: strong and weak. To interpret the welfare-affecting strategic capability modalities $\langle\!\langle C{:}B,H \rangle\!\rangle$, we consider only the strong notion. Specifically, the existential quantification ranges over strongly beneficial or harmful strategies of a coalition. This choice is motivated by our technical results: strong benefit and harm are tailored to best and worst outcomes, and they are the notions for which we provide an axiomatization and obtain an embedding into standard ATL under regular preferences.

The satisfaction relation, denoted by $\models$, between formulas of $\mathcal{L}_{\mathsf{WA-ATL}}$ and pointed models is defined inductively as follows. (The cases for atomic propositions and Boolean connectives are omitted, as they are standard.) Given $\mathsf{M} = (\mathsf{P},\mathcal{V})$ be a CGMRP with $\mathsf{P} = (\mathbb{S},\mathbb{ACT},\mathcal{R},\preceq,\dots)$ as the underlying CGFRP, $s \in \mathbb{S}$ a state, and $\lambda \in Comp_{\mathsf{F},s}$ a history:

$$\begin{array}{rcl} (\mathsf{M},s) \models \langle\!\langle C{:}B,H \rangle\!\rangle \gamma & \Leftrightarrow & \exists \mathcal{F}_C \in StrSB^C_{\mathsf{M}}(s,B) \\ & & \cap StrSH^C_{\mathsf{M}}(s,H) \text{ s.t.} \\ & & \forall \lambda \in \mathcal{O}(s,\mathcal{F}_C), \big(\mathsf{M},\lambda\big) \models \gamma, \\ (\mathsf{M},\lambda) \models \mathsf{X}\varphi & \Leftrightarrow & \big(\mathsf{M},\lambda(1)\big) \models \varphi, \\ (\mathsf{M},\lambda) \models \mathsf{G}\varphi & \Leftrightarrow & \forall k \in \mathsf{dom}(\lambda), \text{ if } k > 0 \\ & & \text{then } \big(\mathsf{M},\lambda(k)\big) \models \varphi, \\ (\mathsf{M},\lambda) \models (\varphi \mathsf{U} \psi) & \Leftrightarrow & \exists k \in \mathsf{dom}(\lambda) \text{ s.t. } k > 0, \\ & & \big(\mathsf{M},\lambda(k)\big) \models \psi, \text{and} \\ & & \forall h \in \mathsf{dom}(\lambda), \text{ if } 0 < h < k \\ & & \text{then } \big(\mathsf{M},\lambda(h)\big) \models \varphi. \end{array}$$

As usual, we say that a formula $\varphi \in \mathcal{L}_{\mathsf{WA-ATL}}$ is valid, if $(\mathsf{M},s) \models \varphi$ for every CGMRP M and every state s in M. We say it is satisfiable if $\neg\varphi$ is not valid.

With the help of the modalities $\langle\langle C{:}B,H\rangle\rangle$, we can elegantly represent the notions of *potential benefactor* and *danger*. In particular, a coalition C is said to be a potential benefactor for another coalition B, denoted $\mathsf{Benef}(C,B)$, if C has a strategy that is beneficial for B. This can be defined as an abbreviation:

$$\mathsf{Benef}(C,B) \stackrel{\mathtt{def}}{=} \langle\langle C{:}B,\emptyset\rangle\rangle \mathsf{X}\top.$$

Analogously, a coalition C is said to be a danger for another coalition H, denoted $\mathsf{Danger}(C,H)$, if C has a strategy that is harmful for H. Formally:

$$\mathsf{Danger}(C,H) \stackrel{\mathtt{def}}{=} \langle\langle C{:}\emptyset,H\rangle\rangle \mathsf{X}\top.$$

Let us go back to our running example introduced in Examples 1 and 2 to illustrate our language.

Example 3 (Cont.)**.** *We have:*

$$(\mathsf{M}_{rs}, s_0) \models \langle\langle \{1,2\}{:}\{1,2\},\{0\}\rangle\rangle \mathsf{G}(\neg inL_0 \wedge \neg inR_0).$$

This means that, at the initial state s_0 of the model M_{rs} corresponding to the robots' scenario, robots 1 and 2 have a strategy that is harmful to robot 0 and beneficial to themselves, which prevents robot 0 from being in the left room or the right room at some point in the future. Consequently, we have:

$$(\mathsf{M}_{rs}, s_0) \models \mathsf{Benef}(\{1,2\},\{1,2\}) \wedge \mathsf{Danger}(\{1,2\},\{0\}).$$

This means that, at the initial state s_0, coalition $\{1,2\}$ is a potential benefactor for itself and a danger for robot 0.

Formulas with nested welfare-affecting strategic modalities are useful to reason about danger-enabling and danger-disabling strategic choices in multi-agent settings. For instance, in the robots scenario, one can express that the coalition $\{0,1\}$ has a strategy that is strongly beneficial to robot 0 and guarantees that in the next state robot 0 becomes a danger to $\{1,2\}$, even though robot 0 is not a danger to $\{1,2\}$ in the current state:

$$(\mathsf{M}_{rs}, s_0) \models \langle\langle \{0,1\}{:}\{0\},\emptyset\rangle\rangle \mathsf{X}\, \mathsf{Danger}(\{0\},\{1,2\}) \wedge \neg\mathsf{Danger}(\{0\},\{1,2\}).$$

Intuitively, robot 0 is initially not a danger because robots 1 and 2 can coordinate to block access to their rooms; however, robot 1 may choose to leave its passage unguarded, allowing robot 0 to enter and then remain in the room indefinitely, thereby preventing robots 1 and 2 from achieving their objective.

Observe that we give a "strict future" interpretation to the temporal modality G: $\mathsf{G}\varphi$ is true if φ holds at all states in the strict future along the actual computation. In standard ATL, a non-strict future interpretation is used. Analogous temporal operators for the non-strict future, in standard ATL style, are definable in the language $\mathcal{L}_{\mathsf{WA-ATL}}$ as follows:

$$\langle\langle C{:}B,H\rangle\rangle \mathsf{G}^{\mathbf{ns}}\varphi \stackrel{\mathtt{def}}{=} \varphi \wedge \langle\langle C{:}B,H\rangle\rangle \mathsf{G}\varphi,$$

$$\langle\langle C{:}B,H\rangle\rangle (\varphi \mathsf{U}^{\mathbf{ns}} \psi) \stackrel{\mathtt{def}}{=} \psi \vee (\varphi \wedge \langle\langle C{:}B,H\rangle\rangle (\varphi \mathsf{U} \psi)).$$

Furthermore, the standard ATL strategic capability modalities $\langle\langle C\rangle\rangle \mathsf{X}\varphi$, $\langle\langle C\rangle\rangle \mathsf{G}^{\mathbf{ns}}\varphi$, and $\langle\langle C\rangle\rangle (\varphi \mathsf{U}^{\mathbf{ns}} \psi)$ are definable as special cases of WA−ATL modalities:

$$\langle\langle C\rangle\rangle \mathsf{X}\varphi \stackrel{\mathtt{def}}{=} \langle\langle C{:}\emptyset,\emptyset\rangle\rangle \mathsf{X}\varphi,$$

$$\langle\langle C\rangle\rangle \mathsf{G}^{\mathbf{ns}}\varphi \stackrel{\mathtt{def}}{=} \langle\langle C{:}\emptyset,\emptyset\rangle\rangle \mathsf{G}^{\mathbf{ns}}\varphi,$$

$$\langle\langle C\rangle\rangle (\varphi \mathsf{U}^{\mathbf{ns}} \psi) \stackrel{\mathtt{def}}{=} \langle\langle C{:}\emptyset,\emptyset\rangle\rangle (\varphi \mathsf{U}^{\mathbf{ns}} \psi).$$

For notational convenience, we denote by $\mathcal{L}_{\mathsf{ATL}}$ the ATL-fragment of the language $\mathcal{L}_{\mathsf{WA-ATL}}$ that includes only modalities of the form $\langle\langle C{:}\emptyset,\emptyset\rangle\rangle$.

We conclude this section by defining two fragments of the language $\mathcal{L}_{\mathsf{WA-ATL}}$. The following grammar defines the until-free *safety* fragment $\mathcal{L}_{\mathsf{WA-ATL}^{-\mathsf{U}}}$:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\langle C{:}B,H\rangle\rangle \mathsf{X}\varphi \mid \langle\langle C{:}B,H\rangle\rangle \mathsf{G}\varphi,$$

while the following grammar defines the *coalition logic* fragment $\mathcal{L}_{\mathsf{WA-CL}}$:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\langle C{:}B,H\rangle\rangle \mathsf{X}\varphi,$$

where p ranges over $\mathbb{P}$, and C, B, and H range over $2^{\mathbb{AGT}}$.

6 Axiomatization

In this section, we introduce a new logic LWASC (Logic of Welfare-Affecting Strategic Capability) defined relative to the until-free fragment $\mathcal{L}_{\mathsf{WA-ATL}^{-\mathsf{U}}}$ and prove its its completeness.

6.1 Logic

The following definition introduces our logic.

Definition 11 (Logic LWASC)**.** LWASC *consists of the following axioms:*

All tautologies of propositional logic	$(\top)$
$\neg\langle\langle C{:}B,H\rangle\rangle \mathsf{X}\bot$	(A-NAAA)
$\langle\langle C{:}B,H\rangle\rangle \mathsf{X}(\varphi \wedge \psi) \to \langle\langle C{:}B,H\rangle\rangle \mathsf{X}\varphi$	(A-MG)
$\langle\langle C{:}B,H\rangle\rangle \mathsf{X}\varphi \to \langle\langle C'{:}B',H'\rangle\rangle \mathsf{X}\varphi$ *for* $C \subseteq C', B \supseteq B', H \supseteq H'$	(A-MC)
$\langle\langle C:\emptyset,\emptyset\rangle\rangle \mathsf{X}\top$	(A-NEI)
$\big(\langle\langle C{:}B,H\rangle\rangle \mathsf{X}\varphi \wedge \langle\langle C'{:}B',H'\rangle\rangle \mathsf{X}\psi\big) \to \langle\langle C \cup C'{:}B \cup B', H \cup H'\rangle\rangle \mathsf{X}(\varphi \wedge \psi)$ *for* $C \cap C' = \emptyset$	(A-Sup)
$\langle\langle C{:}B,H\rangle\rangle \mathsf{X}(\varphi \vee \psi) \to \big(\langle\langle C{:}B,H\rangle\rangle \mathsf{X}\varphi \vee \langle\langle \mathbb{AGT}{:}B,H\rangle\rangle \mathsf{X}\psi\big)$	(A-Cro)
$\langle\langle C{:}B,H\rangle\rangle \mathsf{G}\varphi \leftrightarrow \langle\langle C{:}B,H\rangle\rangle \mathsf{X}\big(\varphi \wedge \langle\langle C{:}B,H\rangle\rangle \mathsf{G}\varphi\big)$	(A-FP$_\mathsf{G}$)
$\langle\langle \emptyset\rangle\rangle \mathsf{G}\big(\psi \to \langle\langle C{:}B,H\rangle\rangle \mathsf{X}(\varphi \wedge \psi)\big) \to \langle\langle \emptyset\rangle\rangle \mathsf{G}\big(\psi \to \langle\langle C{:}B,H\rangle\rangle \mathsf{G}\varphi\big)$	(A-GFP$_\mathsf{G}$)
$\langle\langle \mathbb{AGT}{:}\emptyset,\{a\}\rangle\rangle \mathsf{X}\top$	(A-WFP)
$\langle\langle \mathbb{AGT}{:}\{a\},\emptyset\rangle\rangle \mathsf{X}\top$	(A-CWFP)
$\neg\langle\langle C{:}B,H\rangle\rangle \mathsf{X}\top$ *for* $B \cap H \neq \emptyset$	(A-BHC)

and the following rules of inference:

$$\frac{\varphi, \varphi \to \psi}{\psi} \quad \text{(MP)}$$

$$\frac{\varphi \leftrightarrow \psi}{\langle\langle C{:}B,H\rangle\rangle \mathsf{X}\varphi \leftrightarrow \langle\langle C{:}B,H\rangle\rangle \mathsf{X}\psi} \quad \text{(RE)}$$

$$\frac{\varphi}{\langle\langle C\rangle\rangle \mathsf{G}\varphi} \quad \text{(Gen)}$$

The names of axioms and inference rules reflect their underlying intuitions. Axiom A-NAAA is called *no absurd available action*, as it expresses that a coalition's available joint action cannot ensure a logically absurd result. Axiom A-MG is called *monotonicity of goals*, reflecting that capabilities are monotonic with respect to goals. Similarly, Axiom A-MC captures *monotonicity of coalitions*, meaning that capabilities are monotonic with respect to coalitions. Axiom A-NEI captures the *never-ending interaction* property in Definition 6. Axiom A-Sup captures *superadditivity*. Axiom A-Cro is the so-called *crown* axiom: it was named this way in (Goranko, Jamroga, and Turrini 2013) because the effectivity function of a game has the shape of a crown. Axiom A-FP$_{\mathsf{G}}$ is the *fixpoint axiom for* G, capturing the fixpoint character of the operator G, while Axiom A-GFP$_{\mathsf{G}}$ is the *greatest fixpoint axiom for* G, reflecting its greatest postfixpoint nature. Axioms A-WFP and A-CWFP capture the *well-foundedness* and *converse well-foundedness* of preferences, respectively. Finally, Axiom A-BHC captures *benefit-harm consistency*: a coalition cannot have a strategy that is both beneficial and harmful for the same coalition at the same time, which relates to the property that agents' preferences are non-flat. The inference rules are *modus ponens* (MP), *replacement of equivalences* (RE), and *generalization* (Gen).

As usual, for every $\varphi \in \mathcal{L}_{\mathsf{WA-ATL}^{-\cup}}$, we write $\vdash \varphi$ to mean that φ is derivable in LWASC, that is, there is a sequence of formulas $(\varphi_1, \ldots, \varphi_m)$ such that:

- $\varphi_m = \varphi$, and
- for every $1 \leq k \leq m$, either φ_k is an instance of one of the axiom schema of LWASC or there are formulas $\varphi_{k_1}, \ldots, \varphi_{k_t}$ such that $k_1, \ldots, k_t < k$ and $\frac{\varphi_{k_1}, \ldots, \varphi_{k_t}}{\varphi_k}$ is an instance of some inference rule of LWASC.

We say that a finite set of formulas X is consistent if $\nvdash \bigwedge_{\varphi \in X} \to \bot$. We say a formula φ is consistent if the singleton set $\{\varphi\}$ is consistent.

It is straightforward to verify that all axioms of LWASC are valid and that all inference rules preserve validity. Hence, the soundness of LWASC follows. For completeness, we employ a technique based on canonical models under formula contexts, which is detailed in the next section.

6.2 Completeness

We are going to prove completeness by canonical models under formula contexts. The following definition introduces the notion of formula context.

Definition 12 (Formula context)**.** *Given a formula ψ, we denote the set of all subformulas of ψ by $sub(\psi)$. Let $sub^{+}(\psi) = \{\top, \bot\} \cup sub(\psi) \cup \{\neg\varphi \mid \varphi \in sub(\psi)\}$, $cl(\psi) = sub^{+}(\psi) \cup \{\langle\langle C{:}B,H\rangle\rangle \mathsf{X}\varphi \mid C \subseteq \mathbb{AGT}, B \subseteq \mathbb{AGT}, H \subseteq \mathbb{AGT}, \varphi \in sub^{+}(\psi)\}$, and $cl^{+}(\psi) = cl(\psi) \cup \{\neg\varphi \mid \varphi \in cl(\psi)\}$. We call*

$$cxt(\psi) = cl^{+}(\psi) \cup \{ \bigvee_{X \in \Gamma} \bigwedge_{\chi \in X} \chi \mid \Gamma \subseteq 2^{cl^{+}(\psi)} \}$$

the ψ-context.

We fix a formula ψ and in the rest of this section we work under the ψ-context. The following definition introduces the notion of maximal consistent set under a ψ-context.

Definition 13 (MCS under ψ-context)**.** *Let $X \subseteq cxt(\psi)$. We say X is a maximal consistent set (MCS) under the ψ-context iff X is consistent, and for every $X' \subseteq ecl(\psi)$, if $X \subset X'$ then X' is not consistent.*

The set of all maximal consistent sets under the ψ-context is denoted by $\mathbb{MCS}^{\psi}$. The following definitions introduce the basic structure of the canonical model.

Definition 14 (Canonical states)**.** *We call $s = (X, B, H) \in \mathbb{MCS}^{\psi} \times 2^{\mathbb{AGT}} \times 2^{\mathbb{AGT}}$ a canonical state under the ψ-context iff $B \cap H = \emptyset$. We denote the set of all canonical states under the ψ-context by $\mathbb{S}^{\psi}$.*

We now introduce the notion of a canonical action.

Definition 15 (Canonical action)**.** *Given a formula ψ, we define*

$$\mathbb{LP} = \{\langle\langle C{:}B,H\rangle\rangle \mathsf{X}\chi \mid \chi \in cxt(\psi) \text{ and } C, B, H \in 2^{\mathbb{AGT}}\},$$

$$\mathbb{CF} = \{cf : 2^{\mathbb{S}^{\psi}}/\{\emptyset\} \longrightarrow \mathbb{S}^{\psi} \mid \forall \Omega \in 2^{\mathbb{S}^{\psi}}/\{\emptyset\},\ cf(\Omega) \in \Omega\},$$

$$\mathbb{ACT}^{\psi} = \mathbb{LP} \times \mathbb{AGT} \times \mathbb{CF}.$$

$\mathbb{LP}$ is the set of local powers, $\mathbb{CF}$ the set of choice functions, and $\mathbb{ACT}^{\psi}$ is the set of canonical actions.

The next definition introduces the notions of consensus.

Definition 16 (Consensus)**.** *Given an action profile $\delta : \mathbb{AGT} \longrightarrow \mathbb{ACT}^{\psi}$, we define $\mathcal{CON}(\delta) = \{\langle\langle C{:}B,H\rangle\rangle \mathsf{X}\chi \in \mathbb{LP} \mid \forall a \in C,\ (\delta(a)).1 = \langle\langle C{:}B,H\rangle\rangle \mathsf{X}\chi\}$. We call $\mathcal{CON}(\delta)$ the agents' consensus set relative to δ.*

In the next definition we use the notion of consensus set to define the notion of candidate successor.

Definition 17 (Candidate successor)**.** *Given a canonical state s and action profile $\delta : \mathbb{AGT} \longrightarrow \mathbb{ACT}^{\psi}$, we call a state s' the candidate successors of δ at s, iff:*

- *for all $\langle\langle C{:}B,H\rangle\rangle \mathsf{X}\chi \in \mathcal{CON}(\delta)$, if $\vdash \bigwedge_{\varphi \in s.1} \varphi \to \langle\langle C{:}B,H\rangle\rangle \mathsf{X}\chi$, then $\vdash \bigwedge_{\varphi \in s'.1} \varphi \to \chi$, $B \subseteq s'.2$, and $H \subseteq s'.3$;*
- *for all $\chi \in cxt(\psi)$ and $B, H \in 2^{\mathbb{AGT}}$, if $\nvdash \bigwedge_{\varphi \in s.1} \varphi \to \langle\langle \mathbb{AGT}{:}B,H\rangle\rangle \mathsf{X}\chi$, then $\vdash \bigwedge_{\varphi \in s'.1} \varphi \to \neg\chi$, $B \nsubseteq s'.2$, or $H \nsubseteq s'.3$.*

The set of all candidate successors of δ at s is denoted by $\mathcal{CS}(s, \delta)$.

The transition relation in the canonical model is definable through the previous notion of candidate successor. It can be verified that $\mathcal{CS}(s, \delta) \neq \emptyset$. This guarantees that the transition relation in the canonical model will satisfy the never-ending interaction property. However, the set of candidate successors may be not singleton. We need some mechanism to select a unique successor from the candidate successors. To this aim, we define the following notion of election.

Definition 18 (Election)**.** *Given a action profile $\delta : \mathbb{AGT} \longrightarrow \mathbb{ACT}^{\psi}$, we define $\mathcal{EL}(\delta) =$*

$\Big(\big(\sum_{a\in\mathbb{AGT}}\delta(a).2\big) \bmod n\Big) + 1$, *where* $n = |\mathbb{AGT}|$. *The agent* $\mathcal{EL}(\delta)$ *is called the elected agent of the election relative to the action profile* δ.

We have now all elements to the define the notion of canonical concurrent game model.

Definition 19 (Canonical CGM)**.** *Given a formula* ψ*, define the canonical concurrent game model as the structure* $\mathsf{CM}^\psi = (\mathbb{S}^\psi, \mathbb{ACT}^\psi, \mathcal{R}^\psi, \mathcal{V}^\psi)$ *such that:*

- $\mathbb{S}^\psi$ *is the set of all canonical states under the* ψ*-context;*
- $\mathbb{ACT}^\psi$ *is the set of all canonical actions under* ψ*-context;*
- $\mathcal{R}^\psi$ *assigns to each action profile* δ *a transition relation* $\mathcal{R}_\delta$ *on* $\mathbb{S}$ *such that for all* $s, s' \in \mathbb{S}^\psi$ *and* $\delta : \mathbb{AGT} \longrightarrow \mathbb{ACT}^\psi$*, we have* $s\mathcal{R}_\delta s'$ *iff* $s' = \big(\delta(a).3\big)\big(\mathcal{CS}(s,\delta)\big)$*, with* $a = \mathcal{EL}(\delta)$*;*
- $\mathcal{V}^\psi(p) = \{(X,B,H) \in \mathbb{S}^\psi \mid p \in X\}$ *for each atomic proposition* p *in* $ecl(\varphi)$.

To define the prefereces of agents in the canonical model, we need to define a utility function and the order on them.

Definition 20 (Utility function in canonical model)**.** *Given a canonical CGM* $\mathsf{CM}^\psi = (\mathbb{S}^\psi, \mathbb{ACT}^\psi, \mathcal{R}^\psi, \mathcal{V}^\psi)$*, for each agent* a *and canonical state* s*, we define a utility function* $U_{a,s} : Comp_{\mathsf{M}^\psi,s} \longrightarrow \mathbb{Z} \cup \{-\infty, +\infty\}$ *such that for any computation* $\lambda \in Comp_{\mathsf{CM}^\psi,s}$*:*

$$U_{a,s}(\lambda) = \begin{cases} +\sup(\mathbb{B}) & \text{if } a \in \lambda(1).2, \\ -\sup(\mathbb{H}) & \text{if } a \in \lambda(1).3, \\ 0 & \text{otherwise.} \end{cases}$$

where sup *is the supremum function,* $\mathbb{B} = \{k \in \mathbb{N} \mid \forall 0 < k' \leq k : a \in \lambda(k').2\}$*, and* $\mathbb{H} = \{k \in \mathbb{N} \mid \forall 0 < k' \leq k : a \in \lambda(k').3\}$*. Note* $s'.2 \cap s'.3 = \emptyset$ *for every* $s' \in \mathbb{S}^\psi$*, so* $U_{a,s}$ *is well-defined.*

Now we can define the canonical CGMRP.

Definition 21 (Canonical CGMRP)**.** *Given a formula* ψ*, define the canonical CGMRP as the structure* $\mathsf{M}^\psi = (\mathbb{S}^\psi, \mathbb{ACT}^\psi, \mathcal{R}^\psi, \preceq^\psi, \mathcal{V}^\psi)$ *such that:*

- $(\mathbb{S}^\psi, \mathbb{ACT}^\psi, \mathcal{R}^\psi, \mathcal{V}^\psi)$ *is the canonical CGM defined in Definition 19;*
- $\preceq^\psi$ *assigns to each agent-state pair* (a, s) *a total preorder on computations starting at* s *such that for all* $\lambda, \lambda' \in Comp_{\mathsf{M}^\psi,s}$*:*
  $\lambda \preceq^\psi_{a,s} \lambda'$ *iff* $U_{a,s}(\lambda) \leq U_{a,s}(\lambda')$.

The following lemma guarantees that every consistent formula has a witness state in the canonical model. By the lemma, we have the completeness.

Lemma 1. *Let* $\varphi \in \mathcal{L}_{\mathsf{WA-ATL}^{-\cup}}$ *be a consistent formula. Then, there is a state* $s \in \mathbb{S}^\varphi$ *in the canonical model* M^φ *such that* $\mathsf{M}^\varphi, s \models \varphi$.

Theorem 1 (Soundness and completeness)**.** *A formula* $\varphi \in \mathcal{L}_{\mathsf{WA-ATL}^{-\cup}}$ *is valid if and only if* $\vdash \varphi$.

7 Tree-like Property and Embedding

In this section, we show that satisfiability for the language $\mathcal{L}_{\mathsf{WA-ATL}}$ satisfies the tree-like model property. Based on this result, we provide polynomial embeddings of $\mathcal{L}_{\mathsf{WA-ATL}}$-satisfiability into ATL-satisfiability, and of $\mathcal{L}_{\mathsf{WA-CL}}$-satisfiability into CL-satisfiability.

7.1 Tree-like Property

Let $\mathcal{R}^*$, $\mathcal{R}^-$ and $\mathcal{R}^+$ be, respectively, the reflexive and transitive closure, the inverse, and the transitive closure of $\mathcal{R}_\emptyset$.

Definition 22 (Tree-like CGF)**.** *Let* $\mathsf{F} = (\mathbb{S}, \mathbb{ACT}, \mathcal{R})$ *be a CGF. We say that:*

- F *has a unique root iff there is a unique* $w_0 \in W$ *(called the* root*), such that, for every* $v \in W$*,* $w_0\mathcal{R}^*v$*;*
- F *has unique predecessors iff for every* $v \neq w_0$*, the cardinality of* $\mathcal{R}^-(v)$ *is at most one;*
- F *has no cycles iff* $\mathcal{R}^+$ *is irreflexive;*
- F *is tree-like iff it has a unique root, unique predecessors and no cycles.*

The previous definition of tree-like CGF extends to tree-like CGFRP and tree-like CGMRP in a natural way. The tree-like model property is stated in the following lemma.

Theorem 2. *For* $\varphi \in \mathcal{L}_{\mathsf{WA-ATL}}$*,* φ *is satisfiable iff* φ *is satisfied on a pointed tree-like CGMRP.*

7.2 Embedding

Our embeddings rely on the following translation from $\mathcal{L}_{\mathsf{WA-ATL}}$ to the ATL fragment extended with special atomic formulas:

$$tr : \mathcal{L}_{\mathsf{WA-ATL}}(\mathbb{AGT}, \mathbb{P}) \longrightarrow \mathcal{L}_{\mathsf{ATL}}(\mathbb{AGT}, \mathbb{P}^+),$$

with $\mathbb{P}^+ = \mathbb{P} \cup \{\heartsuit_i \mid i \in \mathbb{AGT}\} \cup \{\clubsuit_i \mid i \in \mathbb{AGT}\}$:

$$\begin{aligned}
tr(p) &= p, \\
tr(\neg\varphi) &= \neg tr(\varphi), \\
tr(\varphi \wedge \psi) &= tr(\varphi) \wedge tr(\psi), \\
tr(\langle\!\langle C{:}B,H\rangle\!\rangle \mathsf{X}\varphi) &= \langle\!\langle C\rangle\!\rangle \mathsf{X}\big(\heartsuit B \wedge \clubsuit H \wedge tr(\varphi)\big), \\
tr(\langle\!\langle C{:}B,H\rangle\!\rangle \mathsf{G}\varphi) &= \langle\!\langle C\rangle\!\rangle \mathsf{G}\big(\heartsuit B \wedge \clubsuit H \wedge tr(\varphi)\big), \\
tr\big(\langle\!\langle C{:}B,H\rangle\!\rangle(\varphi \mathsf{U} \psi)\big) &= \langle\!\langle C\rangle\!\rangle\big((\heartsuit B \wedge \clubsuit H \wedge tr(\varphi))\mathsf{U} \\
&\qquad (\heartsuit B \wedge \clubsuit H \wedge tr(\psi))\big).
\end{aligned}$$

with $\heartsuit B \stackrel{\text{def}}{=} \bigwedge_{i\in B} \heartsuit_i$ and $\clubsuit H \stackrel{\text{def}}{=} \bigwedge_{i\in H} \clubsuit_i$. The special atomic formula $\heartsuit_i$ is read as "agent i is benefited," and $\clubsuit_i$ as "agent i is harmed". The idea behind the above translation is to label, at the language level, beneficial (resp. harmful) strategies of a coalition C for another coalition B (resp. H) with the special constructs $\heartsuit B$ (resp. $\clubsuit H$).

The key observation underlying the embedding for the full language $\mathcal{L}_{\mathsf{WA-ATL}}$ is that, in a tree-like model, whether a coalition's strategy is beneficial or harmful for an agent depends only on whether the next state is the second element of a best (resp. worst) computation for that agent. Stability of preferences ensures that the $\heartsuit B/\clubsuit H$-labeling assigned at the current state remains consistent over time.

As the following theorem shows, by using the translation tr, satisfiability for the language $\mathcal{L}_{\mathsf{WA-ATL}}$ is reducible to ATL-satisfiability, and satisfiability for the fragment $\mathcal{L}_{\mathsf{WA-CL}}$ is reducible to CL-satisfiability.

Theorem 3. *Let $\varphi \in \mathcal{L}_{\mathsf{WA-ATL}}$ and $\psi \in \mathcal{L}_{\mathsf{WA-CL}}$. Then,*

- *φ is satisfiable iff*

$$tr(\varphi) \wedge \bigwedge_{a \in \mathbb{AGT}} (\langle\!\langle \mathbb{AGT} \rangle\!\rangle \mathsf{X} \heartsuit_a \wedge \langle\!\langle \mathbb{AGT} \rangle\!\rangle \mathsf{X} \diamondsuit_a) \wedge \langle\!\langle \emptyset \rangle\!\rangle \mathsf{G}\Big(\bigwedge_{a \in \mathbb{AGT}} (\langle\!\langle \mathbb{AGT} \rangle\!\rangle \mathsf{X} \heartsuit_a \wedge \langle\!\langle \mathbb{AGT} \rangle\!\rangle \mathsf{X} \diamondsuit_a \wedge \neg(\heartsuit_a \wedge \diamondsuit_a))\Big)$$

  is satisfiable;

- *ψ is satisfiable iff*

$$tr(\psi) \wedge \bigwedge_{a \in \mathbb{AGT}} \big(\langle\!\langle \mathbb{AGT} \rangle\!\rangle \mathsf{X} \heartsuit_a \wedge \langle\!\langle \mathbb{AGT} \rangle\!\rangle \mathsf{X} \diamondsuit_a \wedge \langle\!\langle \emptyset \rangle\!\rangle \mathsf{X} \neg(\heartsuit_a \wedge \diamondsuit_a)\big)$$

  is satisfiable.

By Theorem 3, the fact that the size of $tr(\varphi)$ is polynomial, and the fact that satisfiability checking for ATL is EXPTIME-complete (van Drimmelen 2003) and satisfiability checking for CL is PSPACE-complete (Pauly 2002), we obtain the following tight complexity result.

Corollary 1. *Checking satisfiability of formulas in $\mathcal{L}_{\mathsf{WA-ATL}}$ is* EXPTIME*-complete. It is* PSPACE*-complete when restricting to the fragment $\mathcal{L}_{\mathsf{WA-CL}}$.*

8 Conclusion

We have introduced a novel language based on ATL that enables reasoning about beneficial and harmful strategies, as well as the related notions of potential benefactor and danger. Our analysis is supported by several results concerning the axiomatizability and computational complexity of the proposed language.

The axiomatizability of the full language $\mathcal{L}_{\mathsf{WA-ATL}}$, including the 'until' modality, remains open. We plan to address this issue in future work. We also intend to extend our analysis to a risk-based notion of danger, according to which a coalition is a danger to another coalition if it has a strategy that, if chosen, brings harm to the latter with high probability. This extension will require enriching CGMRPs with a probabilistic component. Finally, we plan to combine the notion of danger investigated in this paper with a game-theoretic notion of rationality in order to define the concept of *strong danger*: a coalition C is a strong danger to another coalition H if it has a *rational* strategy that, if chosen, brings harm to H. It is called strong because it is realistic to assume that the agents in C may choose the harmful strategy, provided they act rationally.

Acknowledgments

This work is supported by the TIRIS project CaRe (Caring about Others: AI and Psychology Meet to Model and Automate Collective Reasoning) and the ANR PRCE project DesiRes (Designing Artificial Reasoners for Communication).

AI Declaration

We used generative AI tools solely to identify typographical and grammatical errors. All text, scientific content, definitions, proofs, and claims were produced and verified by the authors, who take full responsibility for the paper.

References

Alur, R.; Henzinger, T. A.; and Kupferman, O. 2002. Alternating-time temporal logic. *Journal of the ACM* 49(5):672–713.

Beckers, S.; Chockler, H.; and Halpern, J. Y. 2024. A causal analysis of harm. *Minds and Machines* 34(3):1–24.

Boudou, J., and Lorini, E. 2018. Concurrent game structures for temporal STIT logic. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2018)*, 381–389. IFAAMAS / ACM.

Bouyer, P.; Kupferman, O.; Markey, N.; Maubert, B.; Murano, A.; and Perelli, G. 2023. Reasoning about quality and fuzziness of strategic behaviors. *ACM Transactions on Computational Logic* 24(3):1–38.

Bradley, B. 2012. Doing away with harm. *Philosophy and Phenomenological Research* 85(2):390–412.

Broersen, J. M.; Herzig, A.; and Troquard, N. 2006. Embedding alternating-time temporal logic in strategic STIT logic of agency. *Journal of Logic and Computation* 16(5):559–578.

Chatterjee, K.; Henzinger, T. A.; and Piterman, N. 2007. Strategy logic. In *Proceedings of the 18th International Conference on Concurrency Theory (CONCUR 2007)*, volume 4703 of *LNCS*, 59–73. Springer.

Fassio, D. 2025. On danger. *Philosophy and Phenomenological Research*. Version of Record online: 08 August 2025.

Frederik, S.; Loewenstein, G.; and O'Donoghue, T. 2002. Time discounting and time preference: A critical review. *Journal of Economic Literature* 40(2):351–401.

Garland, D. 2003. The rise of risk. In Ericson, R. V., and Doyle, A., eds., *Risk and Morality*. University of Toronto Press. 48–86.

Goranko, V., and van Drimmelen, G. 2006. Complete axiomatization and decidability of alternating-time temporal logic. *Theoretical Computer Science* 353:93–117.

Goranko, V.; Jamroga, W.; and Turrini, P. 2013. Strategic games and truly playable effectivity functions. *Journal of Autonomous Agents and Multi-Agent Systems* 26(2):288–314.

Goranko, V. 2001. Coalition games and alternating temporal logics. In *Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2001)*, 259–272. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Grossi, D.; van der Hoek, W.; and Kuijer, L. B. 2022. Reasoning about general preference relations. *Artificial Intelligence* 313:103793.

Harman, E. 2009. Harming as causing harm. In Wasserman, D., and Roberts, M., eds., *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*. Springer. 137–154.

Huang, X., and Ruan, J. 2017. Atl strategic reasoning meets correlated equilibrium. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, 1102–1108. IJCAI.

Jamroga, W.; Mittelmann, M.; Murano, A.; and Perelli, G. 2024. Playing quantitative games against an authority: On the module checking problem. In *Proc. of the 23rd Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2024)*, 926–934. IFAAMAS.

Jiang, J., and Naumov, P. 2024. A logic of higher-order preferences. *Synthese* 203(6):1–26.

Li, Y.; Lorini, E.; and Mittelmann, M. 2025. Rational capability in concurrent games. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025)*, 1309–1317. International Foundation for Autonomous Agents and Multiagent Systems / ACM.

Liu, Z.; Xiong, L.; Liu, Y.; Lespérance, Y.; Xu, R.; and Shi, H. 2020. A modal logic for joint abilities under strategy commitments. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI 2020)*, 1805–1812. IJCAI.

Lorini, E. 2021. A qualitative theory of cognitive attitudes and their change. *Theory and Practice of Logic Programming* 21(4):428–458.

Mittelmann, M.; Maubert, B.; Murano, A.; and Perrussel, L. 2025. Formal verification and synthesis of mechanisms for social choice. *Artificial Intelligence* 339:104272.

Mogavero, F.; Murano, A.; Perelli, G.; and Vardi, M. Y. 2014. Reasoning about strategies: On the model-checking problem. *ACM Transactions on Computational Logic* 15(4):34:1–34:47.

Naumov, P., and Ovchinnikova, A. 2023. An epistemic logic of preferences. *Synthese* 201:1–36.

Pauly, M. 2002. A modal logic for coalitional power in games. *Journal of Logic and Computation* 12(1):149–166.

Peschard, I.; Benétreau-Dupin, Y.; and Wessels, C. 2022. *Philosophy and Science of Risk: An Introduction*. New York: Routledge.

van Benthem, J., and Liu, F. 2007. Dynamic logic of preference upgrade. *Journal of Applied Non Classical Logics* 17(2):157–182.

van Drimmelen, G. 2003. Satisfiability in alternating-time temporal logic. In *Proceedings of the Eighteenth Annual IEEE Symposium on Logic in Computer Science (LICS 2013)*, 208–217. IEEE.

Williamson, T. 2009. Probability and danger. *Amherst Lecture in Philosophy* 1–35.