A Framework for Reasoning about Rational Agents

Wojciech Jamroga and Nils Bulling
Department of Informatics, Clausthal University
of Technology, Germany
{wjamroga,bulling}@in.tu-clausthal.de

ABSTRACT

We propose an extension of alternating-time temporal logic, that can be used for reasoning about the behavior and abilities of agents under various rationality assumptions.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent Systems*; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*Modal logic*

General Terms

Theory

Keywords

multi-agent systems, game theory, temporal logic, rationality

1. INTRODUCTION

Alternating-time temporal logic (ATL) [1] is a temporal logic that incorporates some basic game theoretical notions. In this paper, we extend ATL with a notion of *plausibility*, which can be used to model and reason about rational behavior of agents. In our approach, some strategies (or rather *strategy profiles*) can be assumed plausible, and one can reason about what can be *plausibly* achieved by agents under such an assumption.

This idea has been inspired by the way in which games are analyzed in game theory. First, game theory identifies a number of *solution concepts* (e.g., Nash equilibrium, undominated strategies, Pareto optimality) that can be used to define rationality of players. Then, we usually *assume that players play rationally* in the sense of one of the concepts, and we *ask about the outcome of the game under this assumption*. Note that solution concepts do not only help to determine the right decision for "our" agent. Perhaps even more importantly, they constrain the possible (predicted) responses of the opponents. For many games the number of all possible outcomes is infinite, although only some of them "make sense". Still, we need a notion of rationality (like subgame-perfect Nash equilibrium) to discard the "less sensible" ones, and determine what should happen had the game been played by ideal players.

There are two possible points of focus in this context. Research within game theory understandably favors work on *characterization* of various types of rationality (and defining most appropriate solution concepts). Applications of game theory, also understandably, tend toward *using* the solution concepts to predict the outcome in a given game (i.e., to "solve" the game). The first issue has been studied in the framework of logic, for example in [2, 4, 10, 11]; more recently, game-theoretical solution concepts have been characterized in dynamic logic [7], dynamic epistemic logic [3], and ATL [13, 8]. The second issue seems to have been neglected in logic-based research: papers by Van Otterloo and his colleagues [14, 16, 15] are the only exceptions we know of (and each of them commits to a particular view of rationality). Here, we try to fill in this gap, and propose a general, modal logic-based framework for reasoning about behavior and abilities of rational agents.

AAMAS'07 May 14–18 2007, Honolulu, Hawai'i, USA.

2. ATL

The language of Alternating-time Temporal Logic [1] is defined over a set $\mathbb{A}\mathrm{gt}$ of agents and a set Π of propositions, and consists of the following formulae:

$$\varphi ::= p \mid \neg\varphi \mid \varphi\wedge\varphi \mid \langle\!\langle A\rangle\!\rangle\gamma, \qquad \gamma ::= \bigcirc\varphi \mid \Box\varphi \mid \varphi\,\mathcal{U}\,\varphi.$$

where $p \in \Pi$ and $A \subseteq \mathbb{A}\mathrm{gt}$. Informally, $\langle\!\langle A\rangle\!\rangle\varphi$ says that agents A have a collective strategy to enforce φ. ATL formulae include the usual temporal operators: $\bigcirc$ ("in the next state"), $\Box$ ("always from now on") and $\mathcal{U}$ (strict "until"). Additionally, $\Diamond$ ("now or sometime in the future") can be defined as $\Diamond\varphi \equiv \top\,\mathcal{U}\,\varphi$.

The semantics of ATL is defined in *concurrent game structures*:

$$M = \langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o\rangle,$$

consisting of: a set $\mathbb{A}\mathrm{gt} = \{1,\dots,k\}$ of *agents*; set Q of *states*; set Π of *atomic propositions*; *valuation of propositions* $\pi : Q \to \mathcal{P}(\Pi)$; set Act of *actions*. Function $d : \mathbb{A}\mathrm{gt}\times Q \to \mathcal{P}(Act)$ indicates the actions available to agent $a \in \mathbb{A}\mathrm{gt}$ in state $q \in Q$. A *move vector* in state q is a tuple $\langle m_1,\dots,m_k\rangle \in d(q)$ where $d(q) = \prod_{a\in\mathbb{A}\mathrm{gt}} d_a(q)$ is the set of all move vectors in q. Finally, o is a *transition function* which maps a state $q \in Q$ and a move vector $\langle m_1,\dots,m_k\rangle \in d(q)$ to another state $q' = o(q,\langle m_1,\dots,m_k\rangle)$.

A *computation* or *path* $\lambda = q_0q_1\dots$ is an infinite sequence of states such that there is a transition between each q_i, q_{i+1}. We define $\lambda[i] = q_i$ to denote the i-th state of λ. $\Lambda_{\mathcal{M}}$ denotes all paths of M. The set of all paths, starting in q, is given by $\Lambda_{\mathcal{M}}(q)$.

A (memoryless) *strategy* of agent a is a function $s_a : Q \to Act$ such that $s_a(q) \in d_a(q)$. The set of strategies for agent a is denoted by Σ_a. A *collective strategy* s_A for team $A \subseteq \mathbb{A}\mathrm{gt}$ specifies an individual strategy for each agent $a \in A$. Now, $\Sigma_A =$

$\prod_{a\in A}\Sigma_a$ denotes the set of collective strategies of A. The set of all *strategy profiles* is given by $\Sigma = \Sigma_{\mathbb{A}\mathrm{gt}}$.[1]

Let $A \subseteq B \subseteq \mathbb{A}\mathrm{gt}$, and let s_B be a collective strategy for B. We use $s_B[A]$ to denote the *substrategy* of s_B for agents A, i.e., strategy t_A such that $t^a_A = s^a_B$ for every $a \in A$. Additionally, for a set of strategy profiles P, $P(s_A)$ denotes all strategy profiles that contain s_A as substrategy (i.e., $P(s_A) = \{s' \in P \mid s'[A] = s_A\}$).

Finally, the *outcome* of strategy s_A in state q is defined as the set of all computations that may result from executing s:

$out(q, s_A) = \{\lambda = q_0q_1q_2... \mid q_0 = q$ and for every $i = 1, 2, ...$ there exists $m \in d(q_{i-1})$ such that $m_a = s_A[a](q_{i-1})$ for each $a \in A$, and $o(q_{i-1}, m) = q_i\}$.

The semantics of cooperation modalities can be given through the following clauses:

$M, q \models \langle\langle A\rangle\rangle\bigcirc\varphi$ iff there is a collective strategy s_A such that, for every $\lambda \in out(q, s_A)$, we have $M, \lambda[1] \models \varphi$;

$M, q \models \langle\langle A\rangle\rangle\Box\varphi$ iff there exists s_A such that, for every $\lambda \in out(q, s_A)$, we have $M, \lambda[i] \models \varphi$ for every $i \geq 0$;

$M, q \models \langle\langle A\rangle\rangle\varphi\,\mathcal{U}\,\psi$ iff there exists s_A such that for every $\lambda \in out(q, s_A)$ there is $i \geq 0$, for which $M, \lambda[i] \models \psi$, and $M, \lambda[j] \models \varphi$ for every $0 \leq j < i$.

3. REASONING ABOUT RATIONAL AGENTS

Agents usually have very limited ability to predict the future. However, some lines of action seem often more sensible or realistic than others. Having defined a rationality criterion, we obtain means to *solve* the game, i.e. to determine the *most plausible plays*, and compute their outcome. In game theory, the outcome consists of the payoffs (or utilities) assigned to players at the end of the game. In temporal logics, the outcome of a play can be seen as the *set of paths* that can occur – which allows for subtler descriptions.

In general, plausibility can be seen as a broader notion than rationality: one may obtain plausibility specifications e.g. from learning or folk knowledge. In this paper, however, we focus on plausibility as rationality in a game-theoretical sense.

3.1 ATL with Plausibility

We extend the language of ATL with operators **Pl**, **Ph**, and (**set-pl** ω). **Pl** restricts the considered strategy profiles to ones that are *plausible* in the given model. Consequently, $\mathbf{Pl}\,\langle\langle A\rangle\rangle\gamma$ means that agents A can enforce γ if only plausible strategy profiles can be used. **Ph** disregards plausibility assumptions, and refers to all *physically* available strategies. Finally, we propose one model update operator: (**set-pl** ω) allows to define (or redefine) the set of plausible strategy profiles Υ to the ones described by plausibility term ω (in this sense, it implements *revision* of plausibility). ω is called a *plausibility term*, and refers to a set of strategy profiles from Σ. We note that, in contrast to [6, 12, 5], the concept of plausibility presented in this paper is *objective*, i.e. it does not vary from agent to agent. This is, again, very much in the spirit of game theory, where rationality criteria are used in an analogous way. Moreover, our plausibility concept is *holistic* in the sense that Υ represents an idea of the plausible behavior *of the whole system* (including the behavior of other agents). Finally, it is *global*, because plausibility sets do not depend on the state of the system.

[1]In the original semantics [1], strategies assign agents' choices to *sequences* of states. It should be pointed out, however, that both types of strategies yield equivalent semantics for the "pure" ATL.

DEFINITION 1 (ATLP). *The language $\mathcal{L}_{ATLP}$ is defined over nonempty sets: Π of propositions, $\mathbb{A}\mathrm{gt}$ of agents, and Ω of plausibility terms. Let $p \in \Pi$, $a \in \mathbb{A}\mathrm{gt}$, $A \subseteq \mathbb{A}\mathrm{gt}$, and $\omega \in \Omega$. $\mathcal{L}_{ATLP}$ formulae are defined recursively as:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi\wedge\varphi \mid \langle\langle A\rangle\rangle\bigcirc\varphi \mid \langle\langle A\rangle\rangle\Box\varphi \mid \langle\langle A\rangle\rangle\varphi\,\mathcal{U}\,\varphi \mid \mathbf{Pl}\,\varphi \mid \mathbf{Ph}\,\varphi \mid (\textbf{set-pl}\ \omega)\varphi$$

Again $\Diamond\varphi \equiv \top\,\mathcal{U}\,\varphi$. With ATLP, we can for example express that $\mathbf{Pl}\,\langle\langle\varnothing\rangle\rangle\Box(\mathsf{unlocked} \wedge \mathbf{Ph}\,\langle\langle guard\rangle\rangle\bigcirc\neg\mathsf{unlocked})$: "it is plausible to expect that the emergency doors will always remain unlocked, but the guard retains the physical ability to lock them".

3.2 Semantics of ATLP

Models of ATLP extend concurrent game structures with a plausibility set Υ, and a denotation of plausibility terms $\omega \in \Omega$. The denotation is defined via a *plausibility mapping* $[\![\cdot]\!] : Q \to (\Omega \to \mathcal{P}(\Sigma))$: each term is mapped to a *set* of strategy profiles. Note that the denotation depends on the current state of the system. In a way, the state defines the initial "position in the game", which influences the set of rational strategy profiles for most rationality criteria. For example, a strategy profile can be a Nash equilibrium in q_0, and yet it may not be a NE in some of its successors.

DEFINITION 2. *A* concurrent game structure with plausibility (CGSP) *is given by a tuple*

$$M = \langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o, \Upsilon, \Omega, [\![\cdot]\!]\rangle$$

where $\langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o\rangle$ is a CGS, $\Upsilon \subseteq \Sigma$ is a set of plausible strategy profiles; Ω is a set of of plausibility terms, and $[\![\cdot]\!]$ is a plausibility mapping.

The idea behind $\mathbf{Pl}\,\langle\langle A\rangle\rangle\gamma$ is that only plausible strategy profiles can be played. Thus, coalition A can only choose strategies that are *substrategies* of plausible profiles. Moreover, the agents in $\mathbb{A}\mathrm{gt}\backslash A$ can only respond in a way that yields a plausible strategy profile.

DEFINITION 3. *Let M be a CGSP, $A \subseteq \mathbb{A}\mathrm{gt}$ be a set of agents, $q \in Q$ be a state, $s_A \in \Sigma_A$ be a collective strategy of A, and $P \subseteq \Sigma$ be a set of strategy profiles. The set $out(q, s_A, P)$ contains all computations which may result from agents A executing s_A, when only strategy profiles from P can be played. Formally:*

$$out(q, s_A, P) = \{\lambda \in \Lambda_M(q) \mid \exists z \in P(s_A)\forall i\big(\lambda[i+1] = o(\lambda[i], z(\lambda[i]))\big)\}.$$

Note that $out(q, s_A, \Sigma) = out(q, s_A)$. Furthermore, $\Sigma_A(P)$ denotes all profiles of A consistent with P, i.e., the set contains all $s_A \in \Sigma_A$ such that there is a $t \in P$ with $s_A = t[A]$.

Again, let $P \subseteq \Sigma$. The semantics of ATLP is given by the satisfaction relation $\models_P$ defined as follows:

$M, q \models_P p$ iff $p \in \pi(q)$

$M, q \models_P \neg\varphi$ iff $M, q \not\models_P \varphi$

$M, q \models_P \varphi\wedge\psi$ iff $M, q \models_P \varphi$ and $M, q \models_P \psi$

$M, q \models_P \langle\langle A\rangle\rangle\bigcirc\varphi$ iff there is a strategy profile $s_A \in \Sigma_A(P)$ such that we have $M, \lambda[1] \models_P \varphi$ for all $\lambda \in out(q, s_A, P)$

$M, q \models_P \langle\langle A\rangle\rangle\Box\varphi$ iff there is a strategy profile $s_A \in \Sigma_A(P)$ such that $M, \lambda[i] \models_P \varphi$ for all $\lambda \in out(q, s_A, P)$ and all $i \in \mathbb{N}_0$

$M, q \models_P \langle\langle A\rangle\rangle\varphi\,\mathcal{U}\,\psi$ iff there is a strategy profile $s_A \in \Sigma_A(P)$ and $i \in \mathbb{N}_0$ such that $M, \lambda[i] \models_P \psi$ and for all $j \in \mathbb{N}_0$ with $0 \leq j < i$ we have $M, \lambda[j] \models_P \varphi$

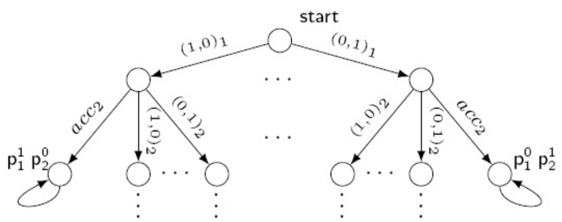

Figure 1: CGS M_2 for the bargaining game

$M, q \models_P \mathbf{Pl}\,\varphi$ iff $M, q \models_\Upsilon \varphi$

$M, q \models_P \mathbf{Ph}\,\varphi$ iff $(M, q) \models \varphi$

$M, q \models_P (\textbf{set-pl}\ \omega)\varphi$ iff $M^\omega, q \models_P \varphi$, where the new model M^ω is equal to M except for the set Υ of plausible strategy profiles, which is now set to $[\![\omega]\!]_q$.

The "absolute" satisfaction relation $\models$ is given by $\models_\Sigma$. Note that an ordinary concurrent game structure (without plausibility) can be interpreted as a CGSP with all strategy profiles assumed plausible, i.e., with $\Upsilon = \Sigma$. This way, satisfaction of ATLP formulae can be extended to ordinary CGS.

We say that formula φ is valid iff $M, q \models \varphi$ for every CGSP M and state $q \in Q_M$; φ is strongly valid iff $M, q \models_P \varphi$ for every CGSP M, $q \in Q_M$, and $P \subseteq \Sigma$; φ is CGS-valid iff $M, q \models \varphi$ for every CGS (without plausibility) M, and $q \in Q_M$. Obviously, strong validity implies validity, which implies CGS-validity. We will be usually interested in the last notion of validity, where plausibility sets must be specified explicitly in the formula. Still, strong validity is important when we want to state that two ATLP formulae are equivalent in the sense of interchangeability.

EXAMPLE 1. *Consider bargaining with discount. Two agents, a_1 and a_2, bargain (in rounds) about how to split goods worth initially $w_0 = 1$ EUR. At each round, the subsequent player makes an offer $\langle x_1, x_2\rangle$, meaning that a_1 takes $x_1 w$ and a_2 gets $x_2 w$, where $x_1 + x_2 = 1$, and w is the current value of the goods. The other player can accept or refuse, but, after each refusal from agent a_i, the worth w decreases by δ_i (called the* discount rate *of player a_i). A CGS M, modeling the game, is presented in Figure 1. The payoffs of agents at the final states are represented with propositions p_i^v (meaning that "player i gets at least the payoff of v").*

Let ω_{NE} denote the set of Nash equilibria (every payoff can be reached by a Nash equilibrium), and ω_{SPN} the set of subgame perfect Nash equilibria in the game. Then, we have the following formula for every $x \in [0,1]$:

$$M, q_0 \models (\textbf{set-pl}\ \omega_{NE})\langle\!\langle 1,2\rangle\!\rangle \Diamond(\mathsf{p}_1^{x} \wedge \mathsf{p}_2^{1-x}) \wedge$$
$$(\textbf{set-pl}\ \omega_{SPN})\langle\!\langle \emptyset\rangle\!\rangle \Diamond(\mathsf{p}_1^{\frac{1-\delta_2}{1-\delta_1\delta_2}} \wedge \mathsf{p}_2^{\frac{\delta_2(1-\delta_1)}{1-\delta_1\delta_2}}).$$

Indeed, every split of the goods has a corresponding Nash equilibrium, but only the split $\left\langle \frac{1-\delta_2}{1-\delta_1\delta_2}, \frac{\delta_2(1-\delta_1)}{1-\delta_1\delta_2}\right\rangle$ is yielded by a subgame perfect NE [9].

4. CONCLUSIONS

We propose a logic in which one can study the outcome of rational play in a logical framework, under various game-theoretical rationality criteria. To our knowledge, there has been very little work on this issue. Note that we are *not* discussing the merits of this or that rationality criterion, nor the pragmatics of using particular criteria to predict the actual behavior of agents. Our aim, most of all, is to propose a conceptual tool in which the consequences of accepting one or another criterion can be studied. We believe that our concept provides much flexibility and modeling power.

Our ultimate goal is to come up with a logic that would allow us to study strategies, time, knowledge, and plausible/rational behavior under both perfect and imperfect information. However, putting so many dimensions in one framework at once is usually not a good idea – even more so in this case because the interaction between abilities and knowledge is non-trivial. In the companion paper [5], we have investigated *time, knowledge and plausibility*. In this paper, we study *strategies, time and rationality*. We hope to integrate both views into a single powerful framework in the future.

5. REFERENCES

[1] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time Temporal Logic. *Journal of the ACM*, 49:672–713, 2002.

[2] M. Bacharach. A theory of rational decision in games. *Erkenntnis*, 27:17–55, 1987.

[3] A. Baltag. A logic for suspicious players. *Bulletin of Economic Research*, 54(1):1–46, 2002.

[4] G. Bonanno. The logic of rational play in games of perfect information. *Economics and Philosophy*, 7:37–65, 1991.

[5] N. Bulling and W. Jamroga. Agents, beliefs and plausible behavior in a temporal setting. In *Proceedings of AAMAS'07*, 2007.

[6] N. Friedman and J.Y. Halpern. A knowledge-based framework for belief change, Part I: Foundations. In *Proceedings of TARK*, pages 44–64, 1994.

[7] P. Harrenstein, W. van der Hoek, J-J. Meijer, and C. Witteveen. Subgame-perfect Nash equilibria in dynamic logic. In M. Pauly and A. Baltag, editors, *Proceedings of the ILLC Workshop on Logic and Games*, pages 29–30. University of Amsterdam, 2002. Tech. Report PP-1999-25.

[8] W. Jamroga, W. van der Hoek, and M. Wooldridge. Intentions and strategies in game-like scenarios. In *Proceedings of EPIA 2005*, volume 3808 of *LNAI*, pages 512–523. Springer Verlag, 2005.

[9] Tuomas W. Sandholm. Distributed rational decision making. In Gerhard Weiss, editor, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, pages 201–258. The MIT Press, Cambridge, MA, USA, 1999.

[10] R. Stalnaker. On the evaluation of solution concepts. *Theory and Decision*, 37(1):49–73, 1994.

[11] R. Stalnaker. Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy*, 12:133–163, 1996.

[12] K. Su, A. Sattar, G. Governatori, and Q. Chen. A computationally grounded logic of knowledge, belief and certainty. In *Proceedings of AAMAS*, pages 149–156. ACM Press, 2005.

[13] W. van der Hoek, W. Jamroga, and M. Wooldridge. A logic for strategic reasoning. In *Proceedings of AAMAS*, pages 157–164, 2005.

[14] W. van der Hoek, S. van Otterloo, and M. Wooldridge. Preferences in game logics. In *Proceedings of AAMAS*, 2004.

[15] S. van Otterloo and G. Jonker. On Epistemic Temporal Strategic Logic. *Electronic Notes in Theoretical Computer Science*, XX:35–45, 2004. Proceedings of LCMAS'04.

[16] S. van Otterloo and O. Roy. Verification of voting protocols. Working paper, University of Amsterdam, 2005.