IfI Technical Reports

Institute of Computer Science,
Clausthal University of Technology

IfI-05-10

Clausthal-Zellerfeld 2005

Constructive Knowledge: What Agents Can Achieve under Incomplete Information

Wojciech Jamroga[1] and Thomas Ågotnes[2]

[1]Institute of Computer Science, Clausthal University of Technology, Germany
wjamroga@in.tu-clausthal.de
[2]Department of Informatics, University of Bergen, Norway
agotnes@ii.uib.no

Abstract

We propose a non-standard semantics for Alternating-time Temporal Logic with incomplete information, for which no commonly accepted semantics has been proposed yet. In our semantics, formulae are interpreted over sets of states rather than single states. We also propose a new epistemic operator for "practical" or "constructive" knowledge, and we show that the new language is strictly more expressive than existing solutions, while it retains the same model checking complexity. Finally, we study properties of constructive knowledge and other operators in a non-standard semantics like this.

Keywords: Alternating-time Temporal Logic, strategic ability, incomplete information, epistemic logic.

1 Introduction

Modal logics of strategic ability [1, 3, 16, 17] form one of the fields where logic and game theory can successfully meet. The logics have clear possible worlds semantics, are axiomatizable, and have some interesting computational properties. Moreover, they are underpinned by a clear and intuitively appealing conceptual machinery for modeling and reasoning about systems that involve multiple autonomous agents. The basic notions, used here, originate from temporal logic (i.e., the logic of time and computation), and classical game theory [25, 14, 15], which emerged in an attempt to give precise meaning to common-sense notions like choices, strategies, or rationality – and to provide formal models of interaction between autonomous entities, that could be used in further study. Thus, the notions and models were meant to describe real-life phenomena that occur in communities of individual and collective agents (e.g., companies). Of course, the treatment of interaction, given by von Neumann, Morgenstern and Nash, *is* oversimplistic, and its fundamental philosophical merit has also been questioned.[1] One may even argue whether modeling of intelligent agents and their interaction can be done with the tools of mathematics and formal logic at all [26, 18]. However, having a formal model of a problem makes one realize many (otherwise implicit) assumptions underlying his or her approach to this problem. Modal logics that embody basic game theory notions – and at the same time build upon branching-time temporal logics, well known and studied in the context of computational systems – seem a good starting point for investigating multi-agent systems.

Alternating-time Temporal Logic (ATL) [1, 2, 3], is probably the most important logic of strategic ability that has emerged in the recent years. However, ATL considers only agents that possess perfect information about the current state of the world, and such agents seldom exist in reality. On the other hand, incomplete information and knowledge are addressed in epistemic logic in a natural way. A combination of ATL and epistemic logic, called *Alternating-time Temporal Epistemic Logic* (ATEL), was introduced in [22, 23] to enable reasoning about agents acting under incomplete information. Still, it has been pointed out in several places that the meaning of ATEL formulae is somewhat counterintuitive. Most importantly, an agent's ability to achieve property φ should imply that the agent has enough control and knowledge to *identify* and *execute* a strategy that enforces φ. A number of ATEL updates were proposed to overcome this problem, yet none of them seems the ultimate definitive solution. Most of the solutions agree that only *uniform* strategies are really executable. However, in order to identify a successful strategy, the agents must consider not only the courses of action, starting from the current state of the system, but also from states that are indistinguishable from the current one. There are many cases here, especially when group epistemics is concerned: the agents may have common, ordinary or distributed knowledge about a strategy being successful, or they may be hinted the right strategy by a distinguished member (the "boss"), a subgroup ("headquarters committee") or even another group of agents ("consulting company") etc. In other words, there are many subtle cases as for which states might be considered as the (possible) initial situations.

In this paper, we propose a non-standard semantics for the logic of strategic ability and incomplete information. In the semantics, formulae are interpreted over *sets of states* rather than single states. This reflects the intuition that the "constructive" ability to enforce φ means that the agents in question have a single strategy that brings about φ for *all* possible initial situations – and not that a successful strategy exists for *each* initial situation (because those could be different strategies for different situations). To do it in a flexible and general way, the type of the satisfaction relation in our proposal forces one to specify the set of initial states explicitly. In consequence, we write $M, Q \models \langle\langle A \rangle\rangle\varphi$ to express the fact that A must have a strategy which is successful for all states in Q. We also propose a new epistemic operator for "practical" or "constructive" knowledge that yields the set of states for which a single evidence (i.e., a successful strategy) should

[1] Consider this quote from [21]: "Rational Behavior [is]: greed, modified by sloth, constrained by formless fear and justified *ex post* by rationalization."

be presented (instead of checking if the required property holds in each of the states separately, like standard epistemic operators do). We point out that this new operator captures the notion of *knowing "de re"*, while the standard epistemic operators refer to *knowing "de dicto"*.

We begin with a short presentation of Alternating-time Temporal Logic (ATL) and the attempts that have been made to extend ATL to scenarios with incomplete information. Then we present the main contribution of this paper: a new, non-standard semantics for the logic of ability, incomplete information and knowledge. We show that it is strictly more expressive than the existing solutions (with the possible exception of ETSL), while it retains the same model checking complexity. Furthermore, we observe that the classical definition of negation does not seem suitable for reasoning about non-strategic properties in such a semantics. To overcome this, we propose a new negation operator (which we call "strong" or "constructive" negation); we also point out that strong negation can be used to define standard knowledge in terms of constructive knowledge. Finally, we study the properties of constructive knowledge itself. It turns out that, out of the S5 properties, axioms K,D,4,5 (but not T!) hold with classical negation (and implication), while only axioms 4,5 hold under constructive negation. However, we observe that a weak negation or a conjunction immediately following a constructive knowledge operator do not have a distinct meaning. Thus, we can restrict the language, excluding such formulae without losing expressive power — and if we do then the T axiom schema becomes valid for both classical and strong negation.

2 What Agents Can Achieve

ATL [1, 2, 3] was invented to capture properties of *open computer systems* (such as computer networks), where different components can act autonomously, and computations in such systems are effected by their combined actions. Alternatively, ATL can be seen as a logic for systems involving multiple agents, that allows one to reason about what agents can achieve in game-like scenarios. As ATL does not include incomplete information in its scope, it can be seen as a logic for reasoning about agents who always have complete information about the current state of affairs.

2.1 ATL: Ability in Perfect Information Games

ATL can be understood as a generalization of the branching time temporal logic CTL [4], in which path quantifiers are replaced with so called *cooperation modalities*. Formula $\langle\langle A\rangle\rangle\varphi$, where A is a coalition of agents, expresses that A have a collective strategy to enforce φ. ATL formulae include temporal operators: "$\bigcirc$" ("in the next state"), $\Box$ ("always from now on") and $\mathcal{U}$ ("until"). Operator $\Diamond$ ("now or sometime in the future") can be defined as $\Diamond\varphi \equiv \top\mathcal{U}\varphi$. Like in CTL, every occurrence of a temporal operator is preceded by exactly one cooperation modality. Example ATL properties are: $\langle\langle jamesbond\rangle\rangle\Diamond\mathsf{win}$ (James Bond has an infallible plan to eventually win) and

$\langle\langle jamesbond, bondsgirl\rangle\rangle \mathsf{fun}\,\mathcal{U}\,\mathsf{shot\text{-}at}$ (Bond and his current girlfriend have a collective way of having fun until someone shoots at them).

A number of semantics have been defined for ATL, most of them equivalent [5, 6]. In this paper, we use a variant of *concurrent game structures*, which includes a nonempty finite set of all agents $\mathbb{A}\mathrm{gt} = \{a_1, ..., a_k\}$, a nonempty set of states St, a set of atomic propositions Π, a valuation of propositions $\pi : St \to \mathcal{P}(\Pi)$, and the set of (atomic) actions Act. Function $d : \mathbb{A}\mathrm{gt} \times St \to \mathcal{P}(Act)$ defines actions available to an agent in a state, and o is a (deterministic) transition function that assigns the outcome state $q' = o(q, \alpha_1, \ldots, \alpha_k)$ to state q and a tuple of actions $\langle \alpha_1, \ldots, \alpha_k \rangle$ that can be executed by $\mathbb{A}\mathrm{gt}$ in q. A *strategy* of agent a is a conditional plan that specifies what a is going to do for every possible situation ($s_a : St \to Act$ such that $s_a(q) \in d_a(q)$).[2] A *collective strategy* S_A for a group of agents A is a tuple of strategies, one per agent from A. A *path* Λ in M is an infinite sequence of states that can be effected by subsequent transitions, and refers to a possible course of action (or a possible computation) that may occur in the system; by $\Lambda[i]$, we denote the ith position on path Λ. Function $out(q, S_A)$ returns the set of all paths that may result from agents A executing strategy S_A from state q onward.

$out(q, S_A) = \{\lambda = q_0 q_1 q_2 ... \mid q_0 = q$ and for every $i = 1, 2, ...$ there exists a tuple of agents' decisions $\langle \alpha_{a_1}^{i-1}, ..., \alpha_{a_k}^{i-1} \rangle$ such that $\alpha_a^{i-1} = S_A(a)(q_{i-1})$ for each $a \in A$, $\alpha_a^{i-1} \in d(a, q_{i-1})$ for each $a \notin A$, and $o(q_{i-1}, \alpha_{a_1}^{i-1}, ..., \alpha_{a_k}^{i-1}) = q_i\}$.

Now, the semantics of ATL formulae can be given via the following clauses:

$M, q \models p$ iff $p \in \pi(q)$ (where $p \in \Pi$);

$M, q \models \neg\varphi$ iff $M, q \not\models \varphi$;

$M, q \models \varphi \wedge \psi$ iff $M, q \models \varphi$ and $M, q \models \psi$;

$M, q \models \langle\langle A\rangle\rangle \bigcirc \varphi$ 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 $M, \Lambda[i]$ 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 $0 \leq j < i$.

The complexity of ATL model checking is linear in the number of transitions in the model and the length of the formula [3].

[2] This is a deviation from the original semantics of ATL [2, 3], where strategies assign agents' choices to *sequences* of states, which suggests that agents can recall the whole history of each game. It should be pointed out, however, that both types of strategies yield equivalent semantics for ATL [20]. The reason why we use "memoryless" strategies here is that, under incomplete information, model checking strategic abilities of agents with perfect recall becomes undecidable (cf. Section 2.2.7).

2.2 Strategic Ability and Incomplete Information

ATL is unrealistic in a sense: real-life agents seldom possess complete information about the current state of the world. On the other hand, incomplete information and knowledge are handled in epistemic logic in a natural way. A combination of ATL and epistemic logic, called *Alternating-time Temporal Epistemic Logic* (ATEL), was introduced in [22, 23] in order to enable reasoning about agents acting under incomplete information.

2.2.1 ATL with Epistemic Logic.

ATEL [22, 23] enriches the picture with epistemic component, adding to ATL operators for representing agents' knowledge: $K_a\varphi$ reads as "agent a knows that φ". Additional operators $E_A\varphi$, $C_A\varphi$, and $D_A\varphi$ refer to *"everybody knows"*, *common knowledge*, and *distributed knowledge* among the agents from A. Thus, $E_A\varphi$ means that every agent in A knows that φ holds, while $C_A\varphi$ means not only that the agents from A know that φ, but they also know that they know it, and know that they know that they know it, etc. The distributed knowledge modality $D_A\varphi$ expresses that, if the agents could share their individual knowledge, they would be able to infer φ.

Models for ATEL extend concurrent game structures with epistemic accessibility relations $\sim_1, ..., \sim_k \subseteq Q \times Q$ (one per agent) for modeling agents' uncertainty.[3] We will call such models *concurrent epistemic game structures* (CEGS) in the rest of the paper. Agent a's epistemic relation is meant to encode a's inability to distinguish between the (global) system states: $q \sim_a q'$ means that, while the system is in state q, agent a cannot determine whether it is in q or q'. Then:

$$M, q \models K_a\varphi \quad \text{iff } \varphi \text{ holds for every } q' \text{ such that } q \sim_a q'.$$

Relations $\sim_A^E$, $\sim_A^C$ and $\sim_A^D$, used to model group epistemics, are derived from the individual relations of agents from A. First, $\sim_A^E$ is the union of relations $\sim_a$, $a \in A$. Next, $\sim_A^C$ is defined as the transitive closure of $\sim_A^E$. Finally, $\sim_A^D$ is the intersection of all the $\sim_a$, $a \in A$. The semantics of group knowledge can be defined as below (for $\mathcal{K} = C, E, D$):

$$M, q \models \mathcal{K}_A\varphi \quad \text{iff } \varphi \text{ holds for every } q' \text{ such that } q \sim_A^{\mathcal{K}} q'.$$

Example 1 (Gambling robots) *Two robots (a and b) play a simple card game. The deck consists of Ace, King and Queen (A, K, Q); it is assumed that A beats K, K beats Q, but Q beats A. First, the "environment" agent deals a random card to both robots (face down), so that each player can see his own hand, but he does not know the card of the other player. Then robot a can exchange his card for the one remaining in the deck (action exch), or he can keep the current one (keep). At the same time, robot b can change the priorities of the cards, so that A becomes better than Q (action chg) or*

[3]The relations are assumed to be equivalences.

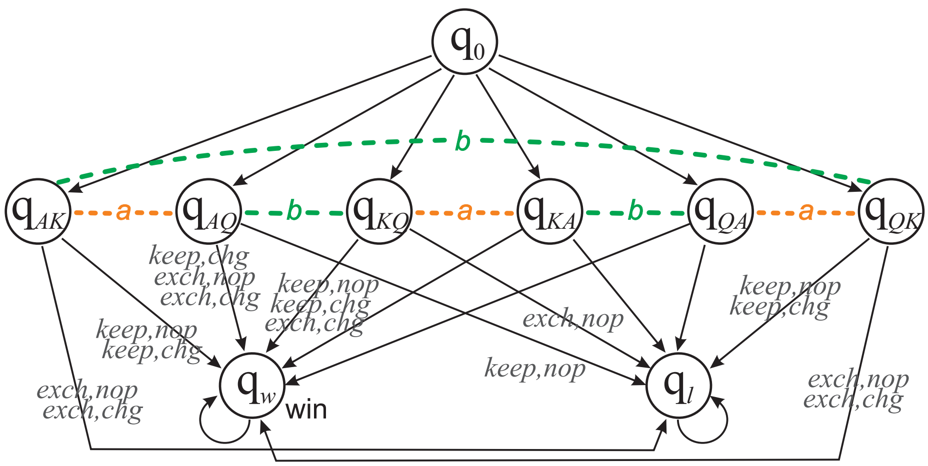

Figure 1: Gambling Robots game

he can do nothing (nop). If a has a better card than b after that, then a win is scored, otherwise the game ends in a "losing" state. A CEGS *for the game is shown in Figure 1. Note that* $q_0 \models \langle\langle a \rangle\rangle \Diamond win$*, although, intuitively, a has no feasible way of ensuring a win. This is a fundamental problem with* ATEL*, which we discuss briefly below.*

2.2.2 Problems with ATEL.

It has been pointed out in several places that the meaning of ATEL formulae is somewhat counterintuitive [7, 9, 11]. Most importantly, one would expect that an agent's ability to achieve property φ should imply that the agent has enough control and knowledge to *identify* and *execute* a strategy that enforces φ (cf. also [20]). ATEL adds to ATL the vocabulary of epistemic logic; still, in ATEL the strategic and epistemic layers are combined as if they were independent. They should be – if we do not ask whether the agents in question are able to identify and execute their strategies. They should not if we want to interpret strategies as *executable plans*, about which the agents *know* that they guarantee achieving the goal.

Moreover, agents in ATEL are assumed some epistemic capabilities when making decisions, and other for epistemic properties like $K_a\varphi$. The interpretation of knowledge operators refers to the agents' capability to distinguish one *state* from another; the semantics of $\langle\langle A \rangle\rangle$ allows the agents to base their decisions upon *histories*, i.e. sequences of states. These tensions between complete vs. incomplete information on one hand, and perfect vs. imperfect recall on the other, has been studied in [9]. It was argued that, when reasoning about what an agent can *enforce*, it seems more appropriate to require the agent to know his winning strategy rather than to know only that such a strategy exists. This problem is closely related to the distinction between knowledge *de re* and knowledge *de dicto*, well known in the philosophy of language [19], as well as research on the interaction between knowledge and action [12, 13, 27]. Several variations on "ATL with incomplete information" have been proposed, yet none of them seems the ultimate definitive solution. We summarize the most important proposals below.

2.2.3 ATL$_{ir}$.

The logic of ATL$_{ir}$ [20] includes the same formulae as ATL, only the cooperation modalities are presented with a subscript: $\langle\langle A\rangle\rangle_{ir}$ to indicate that they address agents with *imperfect information* and *imperfect recall*. Agents are assumed to know their actions, so they must have the same choices in indistinguishable states: if $q \sim_a q'$ then $d(a,q) = d(a,q')$. As a consequence of imperfect recall, agents use memoryless strategies ($s_a : St \to Act$ such that $s_a(q) \in d_a(q)$). As a consequence of imperfect information, they are required to use *uniform* strategies, i.e. ones that specify the same choices in indistinguishable states (if $q \sim_a q'$ then $s_a(q) = s_a(q')$). In other words, agents make choices with respect to their *local* (epistemic) states rather than global states of the system. Speaking a bit informally, formula $\langle\langle A\rangle\rangle_{ir}\varphi$ holds in M, q iff there is a uniform collective strategy S_A such that, for every $a \in A$, q' such that $q \sim_a q'$, and path $\Lambda \in out(q', S_A)$, we have that φ is true for Λ. In other words, there is a strategy such that *everybody in A knows* that executing this strategy will bring about φ. Note that it is not possible to express that A have common knowledge about the successful strategy, or that they are able to identify it if they share their knowledge etc.

Example 2 *Coming back to our gambling robots, it is easy to see that* $q_0 \models \neg\langle\langle a\rangle\rangle_{ir}\Diamond win$*, because, for every a's (uniform) strategy, if it guarantees a win in e.g. state* q_{AK} *then it fails in* q_{AQ} *(and similarly for other pairs of indistinguishable states). Let us also observe that* $q_0 \models \neg\langle\langle a,b\rangle\rangle_{ir}\Diamond win$ *(in order to win, a must keep his card in state* q_{AK}*, so he must keep his card in* q_{AQ} *by uniformity and b must play chg in consequence, etc... and, finally, a must play keep in* q_{QK}*, but that leads to the losing state). On the other hand,* $q_{AQ} \models \langle\langle a,b\rangle\rangle_{ir}\bigcirc win$ *(winning strategy:* $s_a(q_{AK}) = s_a(q_{AQ}) = keep, s_a(q_{KQ}) = exch$*,* $s_b(q_{AQ}) = s_b(q_{KQ}) = s_b(q_{AK}) = chg$*;* q_{AK}, q_{AQ}, q_{QK} *are the states that must be considered by a and b in* q_{AQ}*). Still,* $q_{AK} \models \neg\langle\langle a,b\rangle\rangle_{ir}\bigcirc win$*.*

2.2.4 ATOL.

Alternating-time Observational Temporal Logic (ATOL), proposed independently in [9], follows the same perspective as ATL$_{ir}$. However, it includes also epistemic modalities in the object language (like ATEL), and it offers a richer language of strategic operators to express subtle differences between various kinds of collective abilities of teams. In this paper, we use the notation proposed in [10]. The informal meaning of $\langle\langle A\rangle\rangle_{\mathcal{K}(\Gamma)}\varphi$ is: "group A has a (memoryless uniform) strategy to enforce φ, and agents Γ can identify the strategy as successful for A in the epistemic sense $\mathcal{K}$". That is, $M, q \models \langle\langle A\rangle\rangle_{\mathcal{K}(\Gamma)}\varphi$ iff there is S_A for every $a \in A$, q' such that $q \sim^{\mathcal{K}}_{\Gamma} q'$, and path $\Lambda \in out(q', S_A)$, we have that φ is true for Λ.

Example 3 *Re-writing the properties from Example 2, we get:* $q_0 \models \neg\langle\langle a\rangle\rangle_{K(a)}\Diamond win$, $q_0 \models \neg\langle\langle a,b\rangle\rangle_{E(\{a,b\})}\Diamond win$, $q_{AQ} \models \langle\langle a,b\rangle\rangle_{E(\{a,b\})}\bigcirc win$, *and* $q_{AK} \models \neg\langle\langle a,b\rangle\rangle_{E(\{a,b\})}\bigcirc win$. *Moreover,* ATOL *allows to express subtler ways of identifying a winning strategy: we have that* $q_{AQ} \models \langle\langle a,b\rangle\rangle_{D(\{a,b\})}\bigcirc win \wedge \langle\langle a,b\rangle\rangle_{K(a)}\bigcirc win$ *(the robots can identify the strategy if they share knowledge; also, a can be the "boss" who points out the strategy), and* $q_{AK} \models \neg\langle\langle a,b\rangle\rangle_{C(\{a,b\})}\bigcirc win$ *(despite all of them knowing the winning strategy, they do not have common knowledge about it).*

Proposition 1 [20, 8, 9] *Model checking* ATL_{ir} *and* ATOL *is NP-complete in the size of the model and the formula.*

2.2.5 Feasible ATEL.

"Feasible ATEL" [11] is an update of ATEL, in which the "perfect information" cooperation modalities are kept, but the language is extended with new modalities: $\langle\langle A\rangle\rangle^f$, $\langle\langle A\rangle\rangle^f_E$, $\langle\langle A\rangle\rangle^f_C$, $\langle\langle A\rangle\rangle^f_{K_a}$ and $\langle\langle A\rangle\rangle^f_{M_a}$, that represent agents' ability to find a suitable uniform strategy. These new modalities are very similar to the ones of ATOL. The NP-completeness result carries over to "Feasible ATEL" (it subsumes ATL_{ir} and can be seen as a subset of ATOL).

2.2.6 Epistemic Temporal Strategic Logic.

ETSL [24] digs deeper in the repository of game theory, and focuses on the concept of *undominated strategies*. In a way, $\langle\langle A\rangle\rangle\varphi$ in ETSL can be summarized as: "if A play *rationally* to achieve φ (meaning: they never play a dominated strategy), they will achieve φ". This variant of cooperation modalities has a different flavour than the ones from ATL, ATEL, ATOL etc.; we do not discuss it further here.

2.2.7 Agents with Perfect Recall: ATL_{iR} and ATEL-R*.

In the original formulation of ATL, agents were assumed to have perfect recall of the game, in the sense that they could base their decisions on *sequences* of states rather than single states. As agents seldom have unlimited memory, and logics of strategic ability with incomplete information and perfect recall are believed to have undecidable model checking [20], we do not investigate this variant of ability here.

3 New Semantics for Ability and Knowledge

ATOL covers more cases than ATL_{ir} and "Feasible ATEL", and it is not committed to any notion of rationality (unlike ETSL). One major drawback of ATOL is that it vastly increases the number of modal operators necessary to express properties of agents. For team A, a whole family of cooperation modalities $\langle\langle A\rangle\rangle_{\mathcal{K}(\Gamma)}$ is used (instead of a single modality $\langle\langle A\rangle\rangle$ in ATL) to specify who should identify the right strategy for A, in what way etc. It would be much more elegant to modify the semantics of "simple" cooperation modalities $\langle\langle A\rangle\rangle$ and/or epistemic operators, so that they can be composed into sufficiently expressive formulae. The problem with strategic ability under uncertainty is that, when analyzing consequences of their strategies, agents must consider also the outcome paths starting from states other than the current state – namely, all states that *look the same* as the current state. Thus, a property of a strategy being successful with respect to goal φ is *not* local to the current state; *the same* strategy must be successful in all "opening" states being considered. In order to capture this feature of strategic ability under incomplete information, we change the type of the satisfaction relation $\models$, and define what it means for a formula φ to be satisfied in a *set of states* $Q \subseteq St$ of model M. To our best knowledge, nobody has used this kind of semantics yet.

Moreover, we extend the language of ATEL with unary "constructive knowledge" operators $\mathbb{K}_a$, one for each agent a, that yield the set of states, indistinguishable from the current state from a's perspective. Constructive common, "everybody's" and distributed knowledge is formalized via operators $\mathbb{C}_A, \mathbb{E}_A$, and $\mathbb{D}_A$.

3.1 Language and Semantics

The language includes atomic propositions, Boolean connectives, strategic formulae, standard epistemic operators, and constructive knowledge operators:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\langle A\rangle\rangle \bigcirc \varphi \mid \langle\langle A\rangle\rangle \Box \varphi \mid \langle\langle A\rangle\rangle \varphi \mathcal{U} \varphi \mid C_A\varphi \mid E_A\varphi \mid D_A\varphi \mid \mathbb{C}_A\varphi \mid \mathbb{E}_A\varphi \mid \mathbb{D}_A\varphi.$$

The models are concurrent epistemic game structures again, and we consider only memoryless uniform strategies. Now, we define the notion of a formula φ being satisfied by a set of states Q in a model M, written $M, Q \models \varphi$. We will also write $M, q \models \varphi$ as a shorthand for $M, \{q\} \models \varphi$. Note that this is the latter notion of satisfaction (in single states) that we will ultimately be interested in – but that notion is defined in terms of the (more general) satisfaction in sets of states. Let $\mathrm{img}(q, \mathcal{R})$ be the image of state q with respect to relation $\mathcal{R}$, i.e. the set of all states q' such that $q\mathcal{R}q'$. Moreover, we use $out(Q, S_A)$ as a shorthand for $\cup_{q \in Q} out(q, S_A)$, and $\mathrm{img}(Q, \mathcal{R})$ as a shorthand for $\cup_{q \in Q} \mathrm{img}(q, \mathcal{R})$. The new semantics is given through the following clauses.

$M, Q \models p$ iff $p \in \pi(q)$ for every $q \in Q$;

$M, Q \models \neg\varphi$ iff $M, Q \not\models \varphi$;

$M, Q \models \varphi \wedge \psi$ iff $M, Q \models \varphi$ and $M, Q \models \psi$;

$M, Q \models \langle\langle A\rangle\rangle \bigcirc \varphi$ iff there exists S_A such that, for every $\Lambda \in out(Q, S_A)$, we have that $M, \{\Lambda[1]\} \models \varphi$;

$M, Q \models \langle\langle A\rangle\rangle \Box \varphi$ iff there exists S_A such that, for every $\Lambda \in out(Q, S_A)$ and $i \geq 0$, we have $M, \{\Lambda[i]\} \models \varphi$;

$M, Q \models \langle\langle A\rangle\rangle \varphi \mathcal{U} \psi$ 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 $0 \leq j < i$.

$M, Q \models \mathcal{K}_A \varphi$ iff $M, q \models \varphi$ for every $q \in \mathrm{img}(Q, \sim_A^{\mathcal{K}})$ (where $\mathcal{K} = C, E, D$).

$M, Q \models \hat{\mathcal{K}}_A \varphi$ iff $M, \mathrm{img}(Q, \sim_A^{\mathcal{K}}) \models \varphi$ (where $\hat{\mathcal{K}} = \mathbb{C}, \mathbb{E}, \mathbb{D}$ and $\mathcal{K} = C, E, D$, respectively).

Individual knowledge operators can be derived as: $K_a\varphi \equiv C_{\{a\}}\varphi$ and $\mathbb{K}_a\varphi \equiv \mathbb{C}_{\{a\}}\varphi$. Moreover, we define $\varphi_1 \vee \varphi_2 \equiv \neg(\neg\varphi_1 \wedge \neg\varphi_2)$, and $\varphi_1 \rightarrow \varphi_2 \equiv \neg\varphi_1 \vee \varphi_2$. It should be noted that there are many other possibilities for defining negation, disjunction and implication, corresponding to the different ways of quantifying over the set Q. We henceforth call $\neg$ *weak negation*. Another notion of negation is introduced and discussed in Section 4.

The satisfaction relation $\models$ gives us both the traditional notion of satisfaction in a state, and the more general notion of satisfaction in a set of states. As mentioned above, we are usually interested in the former. We say that a formula is *weakly valid* (or simply *valid*) if it is satisfied by all *states* in all models, i.e. if $M, q \models \varphi$ for all models M and states q in M. It is *strongly valid* if it is satisfied by all *sets* in all models; i.e. if for each M and every set of states Q it is the case that $M, Q \models \varphi$. Strong validity implies validity.

3.2 Expressing Agents' Strategic Abilities

The reason why we need to interpret formulae over sets of states is that we need non-standard epistemic operators: $M, q \models \mathbb{K}_a\langle\langle a\rangle\rangle\varphi$ expresses the fact that a has a single strategy that enforces φ from *all* states indiscernible from q, instead of stating that φ can be achieved from *every* such state *separately*. Note that the latter property is very much in the spirit of standard epistemic logic, and indeed can be captured with the standard knowledge operator (via $K_a\langle\langle a\rangle\rangle\varphi$). More generally, the first kind of formulae refers to *having a strategy "de re"* (i.e. having a successful strategy and knowing the strategy), while the latter refers to *having a strategy "de dicto"* (i.e. only knowing that *some* successful strategy is available; cf. [9]). Note also that the property of having a winning strategy for the current state (but not necessarily even knowing *about* it) is simply expressed with $\langle\langle a\rangle\rangle\varphi$. Capturing different ability levels of coalitions is analogous, with various "epistemic modes" of collective recognizing the right strategy.

Example 4 *Robot a has no winning strategy in the starting state of the game:* $q_0 \models \neg\langle\langle a\rangle\rangle\Diamond win$*, which implies that he has neither a strategy "de re" nor "de dicto" (*$q_0 \models \neg\mathbb{K}_a\langle\langle a\rangle\rangle\Diamond win \wedge \neg K_a\langle\langle a\rangle\rangle\Diamond win$*). On the other hand, he has a successful strategy in* q_{AK} *(just play keep) and he knows he has one (because another action, exch, is bound to win in* q_{AQ}*); still, the knowledge is not constructive, since a does not know which strategy is the right one in the current situation:* $q_{AK} \models \langle\langle a\rangle\rangle\bigcirc win \wedge K_a\langle\langle a\rangle\rangle\bigcirc win \wedge \neg\mathbb{K}_a\langle\langle a\rangle\rangle\bigcirc win$.

Other properties of the gambling robots, that we presented in Examples 2 and 3, can be easily expressed in the new logic by combining constructive knowledge with cooperation modalities: $q_0 \models \neg\mathbb{E}_{\{a,b\}}\langle\langle a,b\rangle\rangle\Diamond win$, $q_{AQ} \models \mathbb{E}_{\{a,b\}}\langle\langle a,b\rangle\rangle\bigcirc win$, $q_{AQ} \models \mathbb{D}_{\{a,b\}}\langle\langle a,b\rangle\rangle\bigcirc win \wedge \mathbb{K}_a\langle\langle a,b\rangle\rangle\bigcirc win$, $q_{AK} \models \mathbb{C}_{\{a,b\}}\langle\langle a,b\rangle\rangle\bigcirc win \wedge \neg\mathbb{C}_{\{a,b\}}\langle\langle a,b\rangle\rangle\bigcirc win$ *etc. In the following proposition, we point out that the new logic is expressive enough to embed the previous solutions, and we present a translation.*

Theorem 2 *Let φ be a formula of* ATL$_{ir}$, ATOL *or "Feasible* ATEL*", and let tr be as follows:*

$$
\begin{array}{cc}
tr(p) = p & tr(\neg\varphi) = \neg tr(\varphi) \\
tr(\varphi \wedge \psi) = tr(\varphi) \wedge tr(\psi) & tr(\bigcirc\varphi) = \bigcirc tr(\varphi) \\
tr(\Box\varphi) = \Box tr(\varphi) & tr(\varphi\,\mathcal{U}\,\psi) = tr(\varphi)\,\mathcal{U}\,tr(\psi) \\
tr(\langle\langle A\rangle\rangle_{ir}\varphi) = \mathbb{E}_A\langle\langle A\rangle\rangle\varphi & tr(\langle\langle A\rangle\rangle_{\mathcal{K}(\Gamma)}\varphi) = \hat{\mathcal{K}}_\Gamma\langle\langle A\rangle\rangle\varphi \\
tr(\langle\langle A\rangle\rangle^f\varphi) = \langle\langle A\rangle\rangle\varphi & tr(\langle\langle A\rangle\rangle^f_{\mathcal{K}}\varphi) = \hat{\mathcal{K}}_A\langle\langle A\rangle\rangle\varphi \\
tr(\langle\langle A\rangle\rangle^f_{K_b}\varphi) = \mathbb{K}_b\langle\langle A\rangle\rangle\varphi & tr(\langle\langle A\rangle\rangle^f_{M_b}\varphi) = \neg K_b\neg\langle\langle A\rangle\rangle\varphi \\
tr(\mathcal{K}_A\varphi) = \mathcal{K}_A tr(\varphi) &
\end{array}
$$

where $\mathcal{K} = C, E, D$ *and* $\hat{\mathcal{K}} = \mathbb{C}, \mathbb{E}, \mathbb{D}$*, respectively. Then:*

$$M, q \models \varphi \text{ iff } M, q \models tr(\varphi).$$

Proof. (Structural induction wrt the structure of φ)

- $M, q \models tr(p)$ iff $M, q \models p$.
- $M, q \models tr(\neg\varphi)$ iff $M, q \not\models tr(\varphi)$ iff (by induction) $M, q \not\models \varphi$ iff $M, q \models \neg\varphi$. [Similarly for $\varphi \wedge \psi$.]
- $M, q \models tr(\langle\langle A\rangle\rangle_{E(\Gamma)}\bigcirc\varphi)$ iff $M, q \models \mathbb{E}_\Gamma\langle\langle A\rangle\rangle\bigcirc tr(\varphi)$ iff $M, \mathrm{img}(q, \sim^E_\Gamma) \models \langle\langle A\rangle\rangle\bigcirc tr(\varphi)$ iff $\exists_{S_A}\forall_{\Lambda\in out(\mathrm{img}(q,\sim^E_\Gamma),S_A)} M, \Lambda[1] \models tr(\varphi)$ iff (by induction) $\exists_{S_A}\forall_{\Lambda\in out(\mathrm{img}(q,\sim^E_\Gamma),S_A)} M, \Lambda[1] \models \varphi$ iff $M, q \models \langle\langle A\rangle\rangle_{E(\Gamma)}\bigcirc\varphi$. [Similarly for cooperation modalities with other temporal operators ($\Box$, $\mathcal{U}$) and epistemic modes (K, C, D etc.).]
- $M, q \models tr(\mathcal{K}_A\varphi)$ iff $M, q \models \mathcal{K}_A tr(\varphi)$ iff $\forall_{q'\in\mathrm{img}(q,\sim^{\mathcal{K}}_A)} M, q' \models tr(\varphi)$ iff (by induction) $\forall_{q'\in\mathrm{img}(q,\sim^{\mathcal{K}}_A)} M, q' \models \varphi$ iff $M, q \models \mathcal{K}_A\varphi$.

□

Remark 3 *The new language is* strictly more expressive *than* ATL$_{ir}$, ATOL *etc.: for example, formula* $\mathbb{E}_A\mathbb{E}_A\langle\langle A\rangle\rangle\varphi$ *cannot be expressed in any of the former logics.*

3.3 Model Checking

The *model checking* problem asks whether a given formula φ holds in a given model M and state q. We define *general model checking* as the problem that asks whether formula φ holds in model M and *set of states* Q. Let $mctl(\varphi, M)$ be a CTL model checker that returns the set of all states that satisfy φ in M. Below, we sketch algorithm $mcheck(\varphi, M, Q)$ that returns "yes" if $M, Q \models \varphi$ and "no" otherwise, running in nondeterministic polynomial time.

- Cases $\varphi \equiv p, \varphi \equiv \neg\psi, \varphi \equiv \sim \psi, \varphi \equiv \psi_1 \wedge \psi_2, \varphi \equiv \mathcal{K}_A\psi$: straightforward (proceed as usually).

- Case $\varphi \equiv \hat{\mathcal{K}}_A\psi$: return $mcheck(\psi, M, \mathrm{img}(Q, \sim_A^{\mathcal{K}}))$.

- Case $\varphi \equiv \langle\!\langle A\rangle\!\rangle \bigcirc \psi$: run $mcheck(\psi, M, q)$ for every $q \in St$, and label the states in which the answer was "yes" with an additional proposition yes (not used elsewhere). Then, guess the strategy of A, and "trim" model M by removing all the transitions inconsistent with the strategy (yielding a sparser model M'). Return "yes" iff $Q' \subseteq mctl(\mathsf{A}\bigcirc\mathsf{yes}, M)$. [For other temporal operators: analogous.]

Note that all the relevant strategies can be guessed *beforehand*, as a single complex (but still polynomial) witness (cf. [8]), which gives us the following result:

Theorem 4 *General model checking for our logic is NP-complete in the size of the model and the formula.*

4 Negation, Disjunction and Knowledge

The semantic role of constructive knowledge operators is to produce sets of states that will appear on the left hand side of the satisfaction relation. In a way, these modalities "aggregate" states into sets, and sets into bigger sets. On the other hand, most of the other operators "split" (or "destroy") sets in the sense that, for evaluating $M, Q \models \varphi$, they require evaluation of subformulae of φ in single states rather than sets of states. Standard epistemic operators (C_A, E_A, D_A) and strong negation (proposed in Section 4.1) are the most straightforward examples (e.g., evaluating $C_A\psi$ in M, Q "splits" into evaluating ψ in each state from $\mathrm{img}(Q, \sim_A^C)$ separately). Cooperation modalities (combined with temporal operators) are "splitting" in a similar way. Besides the "aggregating" and "splitting" operators, there are also "neutral" ones that do not change the set of reference: namely, conjunction ($\wedge$) and weak negation ($\neg$). Below, we study important properties of some of these operators in a non-standard semantics like this.

4.1 Strong Negation and Weak Disjunction

We observe that the semantics of negation presented in Section 3.1 (we call it *weak* negation throughout the rest of the paper) yields a very strong notion of disjunction (and a very weak notion of material implication), as the following proposition states.

Proposition 5

1. $M, Q \models \varphi_1 \vee \varphi_2$ *iff* $M, Q \models \varphi_1$ *or* $M, Q \models \varphi_2$
2. $M, Q \models \varphi_1 \rightarrow \varphi_2$ *iff* $M, Q \models \varphi_1$ *implies* $M, Q \models \varphi_2$.

Proof.

1. $M, Q \models \varphi_1 \vee \varphi_2$ iff $M, Q \models \neg(\neg\varphi_1 \wedge \neg\varphi_2)$ iff $M, Q \not\models \neg\varphi_1 \wedge \neg\varphi_2$ iff $M, Q \not\models \neg\varphi_1$ or $M, Q \not\models \neg\varphi_2$ iff $M, Q \models \varphi_1$ or $M, Q \models \varphi_2$.
2. Straightforward from the above.

□

Such a strong notion of disjunction makes sense when we talk about agents' abilities, i.e. when used inside a $\mathbb{K}_a$ operator. For example: $M, q \models \mathbb{K}_a(\langle\langle A\rangle\rangle\varphi \vee \langle\langle A\rangle\rangle\psi)$ means that A in q can either identify a plan to achieve φ or to achieve ψ. On the other hand, for a disjunction of simpler formulae (e.g. primitive propositions), a weaker notion seems more intuitive: the disjunction should hold in M, Q iff, for any state from Q, at least one of the disjuncts holds (but different disjuncts may hold in different states of Q). To this end, we extend our language with another negation operator $\sim$, which we call *strong* or *constructive* negation, with the following semantics:

$$M, Q \models \sim\varphi \quad \text{iff } M, q \not\models \varphi \text{ for every } q \in Q;$$

Strong negation defines a weak notion of disjunction, and a strong notion of material implication:

$$\begin{aligned} \varphi_1 \parallel \varphi_2 &\equiv \quad \sim(\sim\varphi_1 \wedge \sim\varphi_2) \\ \varphi_2 \rightsquigarrow \varphi_2 &\equiv \quad \sim\varphi_1 \parallel \varphi_2 \end{aligned}$$

Proposition 6

1. $M, Q \models \varphi_1 \parallel \varphi_2$ *iff* $\forall_{q\in Q} M, q \models \varphi_1 \vee \varphi_2$
2. $M, Q \models \varphi_1 \rightsquigarrow \varphi_2$ *iff* $\forall_{q\in Q} M, q \models \varphi_1 \rightarrow \varphi_2$

Proof.

1. $M, Q \models \sim(\sim\varphi_1 \wedge \sim\varphi_2)$ iff $\forall_{q\in Q} M, q \not\models \sim\varphi_1 \wedge \sim\varphi_2$ iff $\forall_{q\in Q} M, q \not\models \sim\varphi_1$ or $M, q \not\models \sim\varphi_2$ iff $\forall_{q\in Q} M, q \models \varphi_1$ or $M, q \models \varphi_2$ iff $\forall_{q\in Q} M, q \models \varphi_1 \vee \varphi_2$.

2. $M, Q \models \varphi_1 \rightsquigarrow \varphi_2$ iff $M, Q \models \sim \varphi_1 \parallel \varphi_2$ iff $\forall_{q \in Q} M, q \models \sim \varphi_1 \vee \varphi_2$ iff $\forall_{q \in Q}(M, q \models \sim \varphi_1$ or $M, q \models \varphi_2)$ iff $\forall_{q \in Q}(M, q \not\models \varphi_1$ or $M, q \models \varphi_2$ iff $\forall_{q \in Q} M, q \models \varphi_1 \rightarrow \varphi_2$.

□

For validity (not strong validity), the two negations, the two disjunctions and the two implications coincide (the proof is immediate from the propositions above):

Proposition 7 *The following are valid (not strongly valid):*

1. $\neg\varphi \leftrightarrow \sim\varphi$
2. $(\varphi_1 \vee \varphi_2) \leftrightarrow (\varphi_1 \parallel \varphi_2)$
3. $(\varphi_1 \rightarrow \varphi_2) \leftrightarrow (\varphi_1 \rightsquigarrow \varphi_2)$

Note that the $\sim$ operator does *not* behave as classical negation: it does not obey the law of double negation under *strong* validity (although the law holds with respect to weak validity by Proposition 7.1). Nevertheless, it preserves the law of excluded middle and the consistency requirement.

Proposition 8

1. $\sim\sim\varphi \leftrightarrow \varphi$ *is not strongly valid.*
2. $\varphi \parallel \sim\varphi$ *is strongly valid.*
3. $\sim(\varphi \wedge \sim\varphi)$ *is strongly valid.*

Proof.

1. Counterexample: $\varphi \equiv \langle\langle A \rangle\rangle \psi$. Then $M, Q \models \varphi$ iff A have one successful strategy for all $q \in Q$, and $M, Q \models \sim\sim\varphi \Leftrightarrow \forall_{q \in Q} M, Q \models \varphi$ iff a have one strategy per each $q \in Q$.
2. $M, Q \models \varphi \parallel \sim\varphi$ iff $M, Q \models \sim(\sim\varphi \wedge \sim\sim\varphi)$ iff $\forall_{q \in Q}(M, q \not\models \sim\varphi \wedge \sim\sim\varphi)$ iff $\forall_{q \in Q}(M, q \models \varphi$ or $M, q \not\models \varphi)$.
3. $M, Q \models \sim(\varphi \wedge \sim\varphi)$ iff $\forall_{q \in Q}(M, q \not\models \varphi \wedge \sim\varphi)$ iff $\forall_{q \in Q}(M, q \not\models \varphi$ or $M, q \models \varphi)$.

□

Finally, a minor remark. We have considered two possible operators for negation. There is a third one which looks quite natural: $M, Q \models \angle\varphi \Leftrightarrow \exists_{q \in Q} M, q \not\models \varphi$, and can be investigated in the future.

4.2 Properties of Constructive Knowledge

In the following proposition we list some properties of constructive knowledge (keep in mind that strong validity implies validity).

Proposition 9 *The following are strongly valid:*

1. $\mathbb{K}_a(\varphi_1 \vee \varphi_2) \leftrightarrow (\mathbb{K}_a\varphi_1 \vee \mathbb{K}_a\varphi_2)$
2. $\mathbb{K}_a\neg\varphi \leftrightarrow \neg\mathbb{K}_a\varphi$
3. $\mathbb{K}_a(\varphi_1 \wedge \varphi_2) \leftrightarrow (\mathbb{K}_a\varphi_1 \wedge \mathbb{K}_a\varphi_2)$
4. $\mathbb{K}_a(\varphi_1 \rightarrow \varphi_2) \leftrightarrow (\mathbb{K}_a\varphi_1 \rightarrow \mathbb{K}_a\varphi_2)$

Proof.

1. $M, Q \models \mathbb{K}_a(\varphi_1 \vee \varphi_2)$ iff $M, \mathrm{img}(Q, \sim_a) \models \varphi_1 \vee \varphi_2$ iff $M, \mathrm{img}(Q, \sim_a) \models \varphi_1$ or $M, \mathrm{img}(Q, \sim_a) \models \varphi_2$ iff $M, Q \models \mathbb{K}_a\varphi_1$ or $M, Q \models \mathbb{K}_a\varphi_2$ iff $M, Q \models \mathbb{K}_a\varphi_1 \vee \mathbb{K}_a\varphi_2$.
2. $M, Q \models \mathbb{K}_a\neg\varphi$ iff $M, \mathrm{img}(Q, \sim_a) \models \neg\varphi$ iff $M, \mathrm{img}(Q, \sim_a) \not\models \varphi$ iff $M, Q \not\models \mathbb{K}_a\varphi$ iff $M, Q \models \neg\mathbb{K}_a\varphi$.
3. $M, Q \models \mathbb{K}_a\varphi_1 \wedge \varphi_2$ iff $M, \mathrm{img}(Q, \sim_a) \models \varphi_1 \wedge \varphi_2$ iff $M, \mathrm{img}(Q, \sim_a) \models \varphi_1$ and $M, \mathrm{img}(Q, \sim_a) \models \varphi_2$ iff $M, Q \models \mathbb{K}_a\varphi_1$ and $M, Q \models \mathbb{K}_a\varphi_2$ iff $M, Q \models \varphi_1 \wedge \varphi_2$.
4. $M, Q \models \mathbb{K}_a(\neg\varphi_1 \vee \varphi_2)$ iff $M, Q \models (\mathbb{K}_a\neg\varphi_1) \vee \mathbb{K}_a\varphi_2$ iff $M, Q \models (\neg\mathbb{K}_a\varphi_1) \vee \mathbb{K}_a\varphi_2$ iff $M, Q \models \mathbb{K}_a\varphi_1 \rightarrow \mathbb{K}_a\varphi_2$.

□

Furthermore, it turns out that standard knowledge is definable by constructive knowledge and strong negation:

Theorem 10 $K_a\varphi \leftrightarrow \mathbb{K}_a \sim\sim \varphi$ *is strongly valid.*

Proof. $M, Q \models \mathbb{K}_a \sim\sim \varphi$ iff $M, \mathrm{img}(Q, \sim_a) \models \sim\sim \varphi$ iff $\forall_{q' \in \mathrm{img}(Q, \sim_a)} M, q' \not\models \sim \varphi$ iff $\forall_{q' \in \mathrm{img}(Q, \sim_a)} M, q' \models \varphi$ iff $M, Q \models K_a\varphi$. □

Theorem 10 shows that, when we have strong negation in the language, standard knowledge is a special case of constructive knowledge.

4.2.1 Is $\mathbb{K}_a$ an Epistemic Operator?

Do the S5 properties of knowledge hold for constructive knowledge? It might depend on the whether we use weak or strong implication and negation, but in both cases the answer is *no*. Particularly, the truth axiom does not hold. Also, the version of the **K** axiom with strong implication does not hold.

Theorem 11 *Below, we list the constructive knowledge versions of the S5 properties (plus some other properties) with weak implication/negation, and with strong implication/negation. "Yes" means that the schema is strongly valid; "No" means that it is not even weakly valid (incidentally, none of the properties turned out to be weakly but not strongly valid). While the truth axiom does not hold, we observe that the introspection axioms can be strengthen to equivalences.*

$\mathbf{K}$	$\mathbb{K}_a(\varphi \to \psi) \to (\mathbb{K}_a\varphi \to \mathbb{K}_a\psi)$	*Yes*	$\widetilde{\mathbf{K}}$	$\mathbb{K}_a(\varphi \rightsquigarrow \psi) \rightsquigarrow (\mathbb{K}_a\varphi \rightsquigarrow \mathbb{K}_a\psi)$	*No*
$\mathbf{D}$	$\mathbb{K}_a\varphi \to \neg\mathbb{K}_a\neg\varphi$	*Yes*	$\widetilde{\mathbf{D}}$	$\mathbb{K}_a\varphi \rightsquigarrow \sim\mathbb{K}_a\sim\varphi$	*No*
$\mathbf{D}^+$	$\mathbb{K}_a\varphi \leftrightarrow \neg\mathbb{K}_a\neg\varphi$	*Yes*	$\widetilde{\mathbf{D}}^+$	$\mathbb{K}_a\varphi \leftrightarrow \sim\mathbb{K}_a\sim\varphi$	*No*
$\mathbf{T}$	$\mathbb{K}_a\varphi \to \varphi$	*No*	$\widetilde{\mathbf{T}}$	$\mathbb{K}_a\varphi \rightsquigarrow \varphi$	*No*
$\mathbf{4}$	$\mathbb{K}_a\varphi \to \mathbb{K}_a\mathbb{K}_a\varphi$	*Yes*	$\widetilde{\mathbf{4}}$	$\mathbb{K}_a\varphi \rightsquigarrow \mathbb{K}_a\mathbb{K}_a\varphi$	*Yes*
$\mathbf{4}^+$	$\mathbb{K}_a\varphi \leftrightarrow \mathbb{K}_a\mathbb{K}_a\varphi$	*Yes*	$\widetilde{\mathbf{4}}^+$	$\mathbb{K}_a\varphi \rightsquigarrow \mathbb{K}_a\mathbb{K}_a\varphi$	*Yes*
$\mathbf{5}$	$\neg\mathbb{K}_a\varphi \to \mathbb{K}_a\neg\mathbb{K}_a\varphi$	*Yes*	$\widetilde{\mathbf{5}}$	$\sim\mathbb{K}_a\varphi \rightsquigarrow \mathbb{K}_a\sim\mathbb{K}_a\varphi$	*Yes*
$\mathbf{5}^+$	$\neg\mathbb{K}_a\varphi \leftrightarrow \mathbb{K}_a\neg\mathbb{K}_a\varphi$	*Yes*	$\widetilde{\mathbf{5}}^+$	$\sim\mathbb{K}_a\varphi \rightsquigarrow \mathbb{K}_a\sim\mathbb{K}_a\varphi$	*Yes*
$\mathbf{B}$	$\varphi \to \mathbb{K}_a\neg\mathbb{K}_a\neg\varphi$	*No*	$\widetilde{\mathbf{B}}$	$\varphi \rightsquigarrow \mathbb{K}_a\sim\mathbb{K}_a\sim\varphi$	*No*

Before proving Theorem 11, we establish an intermediate result.

Lemma 12

1. *There is a model M, state q, agent a and formula φ such that $M, \mathrm{img}(q, \sim_a) \models \varphi$ and $M, q \not\models \varphi$.*
2. *There is a model M, state q, agent a and formula φ such that $M, \mathrm{img}(q, \sim_a) \not\models \varphi$ and $M, q \models \varphi$.*

Proof.

1. Let $\varphi = \neg p$, where p is a primitive proposition, and let M be a model with two states q, q' such that $q \sim_a q'$, $\pi(q) = \{p\}$ and $\pi(q') = \emptyset$. $p \notin \pi(q) \cap \pi(q')$; $M, \{q, q'\} \not\models p$; $M, \mathrm{img}(q, \sim_a) \models \varphi$. But $p \in \pi(q)$, so $M, q \models p$; $M, q \not\models \varphi$.
2. Let M be as above, and let $\varphi = p$. $p \in \pi(q)$, so $M, q \models \varphi$. But $p \notin \pi(q) \cap \pi(q')$, so $M, \mathrm{img}(q, \sim_a) \not\models \varphi$.

□

Proof. [of Theorem 11]

$\mathbf{4}^+/\mathbf{4}$: $M, Q \models \mathbb{K}_a\mathbb{K}_a\varphi$ iff $M, \mathrm{img}(Q, \sim_a) \models \mathbb{K}_a\varphi$ iff $M, \mathrm{img}(\mathrm{img}(Q, \sim_a), \sim_a) \models \varphi$ iff $M, \mathrm{img}(Q, \sim_a) \models \varphi$ (since $\mathrm{img}(\mathrm{img}(Q, \sim_a), \sim_a) = \mathrm{img}(Q, \sim_a)$) iff $M, Q \models \mathbb{K}_a\varphi$.

$\mathbf{K}$: Immediate by Proposition 9.

$\mathbf{T}$: Let M, q, a, φ be as in Lemma 12.1. $M, q \models \mathbb{K}_a\varphi$, but $M, q \not\models \varphi$, so **T** is not weakly (and hence not strongly) valid.

$5^+/5$: $M, Q \models \neg\mathbb{K}_a\varphi$ iff $M, Q \not\models \mathbb{K}_a\varphi$ iff, by $\mathbf{4}^+$, $M, Q \not\models \mathbb{K}_a\mathbb{K}_a\varphi$ iff, by Proposition 9, $M, Q \models \mathbb{K}_a\neg\mathbb{K}_a\varphi$.

$\mathbf{D}^+/\mathbf{D}$: $M, Q \models \mathbb{K}_a\varphi$ iff $M, Q \models \neg\neg\mathbb{K}_a\varphi$ iff, by Proposition 9, $M, Q \models \neg\mathbb{K}_a\neg\varphi$.

B: Let M, q, a, φ be as in Lemma 12.2. $M, \mathrm{img}(q, \sim_a) \not\models \varphi$, so $M, q \models \mathbb{K}_a\neg\varphi$. By $\mathbf{4}^+$, $M, q \models \mathbb{K}_a\mathbb{K}_a\neg\varphi$, so $M, q \not\models \neg\mathbb{K}_a\mathbb{K}_a\neg\varphi$, and by Proposition 9 $M, q \not\models \mathbb{K}_a\neg\mathbb{K}_a\neg\varphi$. But $M, q \models \varphi$. Thus, **B** is not weakly (and hence not strongly) valid.

$\widetilde{\mathbf{K}}$: We construct a counterexample. Let M be a model with states q_1, q_2 and agent a, such that $q_1 \sim_a q_2$, $\pi(q_1) = \{r\}$ and $\pi(q_2) = \{p\}$. Let $\varphi = \neg p$ and $\psi = r$. $p \notin \pi(q_1) \cap \pi(q_2)$, so $M, \mathrm{img}(q_1, \sim_a) \models \neg\varphi$ and $M, q_1 \models \mathbb{K}_\varphi$. $r \notin \pi(q_1) \cap \pi(q_2)$, so $M, \mathrm{img}(q_1, \sim_a) \not\models \psi$ and $M, q_1 \not\models \mathbb{K}_a\psi$. Thus, $M, q_1 \not\models \mathbb{K}_a\varphi \rightarrow \mathbb{K}_a\psi$ and by Proposition 7: $M, q_1 \not\models \mathbb{K}_a\varphi \rightsquigarrow \mathbb{K}_a\psi$ (*). Since both $M, q_1 \models \varphi \rightarrow \psi$ and $M, q_2 \models \varphi \rightarrow \psi$, by Proposition 6, $M, \mathrm{img}(q_1, \sim_a) \models \varphi \rightsquigarrow \psi$ and thus $M, q_1 \models \mathbb{K}_a(\varphi \rightsquigarrow \psi)$. Together with (*), we get that $M, q_1 \not\models \mathbb{K}_a(\varphi \rightsquigarrow \psi) \rightarrow (\mathbb{K}_a\varphi \rightsquigarrow \mathbb{K}_a\psi)$ and, by Proposition 7, $M, q_1 \not\models \mathbb{K}_a(\varphi \rightsquigarrow \psi) \rightsquigarrow (\mathbb{K}_a\varphi \rightsquigarrow \mathbb{K}_a\psi)$. Thus, $\widetilde{\mathbf{K}}$ is not weakly (and hence not strongly) valid.

$\widetilde{\mathbf{T}}$: Let M, q, a, φ be as in Lemma 12.1. $M, q \models \mathbb{K}_a\varphi$ and $M, q \not\models \varphi$, so $M, q \not\models \mathbb{K}_a\varphi \rightsquigarrow \varphi$ by Proposition 7. Thus, $\widetilde{\mathbf{T}}$ is not weakly (and hence not strongly) valid.

$\widetilde{\mathbf{4}}^+/\widetilde{\mathbf{4}}$: $M, Q \models \mathbb{K}_a\varphi \rightsquigarrow \mathbb{K}_a\mathbb{K}_a\varphi$ iff, by Proposition 6, $\forall_{q\in Q}(M, q \models \mathbb{K}_a\varphi \Leftrightarrow M, q \models \mathbb{K}_a\mathbb{K}_a\varphi)$ iff, by $\mathbf{4}^+$, $\forall_{q\in Q}(M, q \models \mathbb{K}_a\varphi \leftrightarrow M, q \models \mathbb{K}_a\varphi)$.

$\widetilde{\mathbf{5}}^+/\widetilde{\mathbf{5}}$: $M, Q \models \sim \mathbb{K}_a\varphi \rightsquigarrow \mathbb{K}_a \sim \mathbb{K}_a\varphi$ iff, by Proposition 6, $\forall_{q\in Q}(M, q \models \sim \mathbb{K}_a\varphi \Leftrightarrow M, q \models \mathbb{K}_a \sim \varphi)$ iff $\forall_{q\in Q}(M, \mathrm{img}(q, \sim_a) \not\models \varphi \Leftrightarrow M, \mathrm{img}(q, \sim_a) \models \sim \varphi)$ iff $\forall_{q\in Q}(M, \mathrm{img}(q, \sim_a) \not\models \varphi \Leftrightarrow \forall_{q'\in \mathrm{img}(q,\sim_a)} M, \mathrm{img}(q', \sim_a) \not\models \varphi)$ which is true, since $\mathrm{img}(q', \sim_a) = \mathrm{img}(q, \sim_a)$ for any $q' \in \mathrm{img}(q, \sim_a)$.

$\widetilde{\mathbf{D}}$: $M, q \models \mathbb{K}_a\varphi \rightsquigarrow \sim \mathbb{K}_a \sim \varphi$ iff $M, \mathrm{img}(q, \sim_a) \models \varphi \Rightarrow M, q \models \sim \mathbb{K}_a \sim \varphi$ iff $M, \mathrm{img}(q, \sim_a) \models \varphi \Rightarrow M, q \not\models \mathbb{K}_a \sim \varphi$ iff $M, \mathrm{img}(q, \sim_a) \models \varphi \Rightarrow M, \mathrm{img}(q, \sim_a) \not\models \sim \varphi$ iff $M, \mathrm{img}(q, \sim_a) \models \varphi \Rightarrow \forall q' \in \mathrm{img}(q, \sim_a) M, q' \models \varphi$. Let M, q, a, φ be as in Lemma 12.1, and let $q' = q$. $M, \mathrm{img}(q, \sim_a) \models \varphi$ but $M, q' \not\models \varphi$. Thus, $\widetilde{\mathbf{D}}$ is not weakly (and hence not strongly) valid.

$\widetilde{\mathbf{B}}$: $M, q \models \varphi \rightsquigarrow \mathbb{K}_a \sim \mathbb{K}_a \sim \varphi$ iff $M, q \models \varphi \Rightarrow M, \mathrm{img}(q, \sim_a) \models \sim \mathbb{K}_a \sim \varphi$ iff $M, q \models \varphi \Rightarrow \forall_{q'\in \mathrm{img}(q,\sim_a)} M, q' \models \mathbb{K}_a \sim \varphi$ iff $M, q \models \varphi \Rightarrow \forall_{q'\in \mathrm{img}(q,\sim_a)} M, \mathrm{img}(q', \sim_a) \models \sim \varphi$ iff $M, q \models \varphi \Rightarrow \forall_{q'\in \mathrm{img}(q,\sim_a)} \forall_{q''\in \mathrm{img}(q',\sim_a)} M, q'' \models \varphi$ iff $M, q \models \varphi \Rightarrow \forall_{q'\in \mathrm{img}(q,\sim_a)} M, q' \models \varphi$. Let M, q, a, φ be as in Lemma 12.2. $q \in \mathrm{img}(q, \sim_a)$ but $M, \mathrm{img}(q, \sim_a) \not\models \varphi$. But $M, q \models \varphi$. Thus, $\widetilde{\mathbf{B}}$ is not weakly (and hence not strongly) valid.

□

4.2.2 In Quest for the Truth Axiom

We have just showed that, out of the S5 properties, axioms K,D,4,5 (but not T!) hold wrt weak negation, while only 4,5 hold for constructive negation.

However, it also turns out that if we restrict the language so that a constructive knowledge operator is not immediately followed by weak negation nor conjunction operator, then the corresponding T axiom of both types of negation becomes strongly valid. Let $\mathcal{L}^-$ be the subset of our logical language in which operators $\mathbb{C}_A, \mathbb{E}_A, \mathbb{D}_A$ are never immediately followed by neither $\neg$ nor $\wedge$.[4]

Theorem 13 *Every $\mathcal{L}^-$ instance of schema $\mathbf{T}$ (i.e., $\mathbb{K}_a\varphi \to \varphi$), and every $\mathcal{L}^-$ instance of schema $\widetilde{\mathbf{T}}$ ($\mathbb{K}_a\varphi \rightsquigarrow \varphi$) are strongly valid.*

Proof.

($\mathbf{T}$) Note that $M,Q \models \mathbb{K}_a\varphi \to \varphi$ iff $M,Q \models \mathbb{K}_a\varphi \Rightarrow M,Q \models \varphi$ (Proposition 5). Let $M,Q \models \mathbb{K}_a\varphi$ and $Q' = \mathrm{img}(Q,\sim_a)$, then $M,Q' \models \varphi$. Note that $Q \subseteq Q'$ (because all $\sim_A^{\hat{\mathcal{K}}}$ are reflexive). Thus: (1) for $\varphi \equiv p$ or φ beginning with a "splitting" operator, we have $M,Q \models \varphi$ immediately (because then the satisfaction in Q "splits" to satisfaction in for each $q \in Q$, and vice versa); (2) for $\varphi \equiv \hat{\mathcal{K}}^1{}_{B_1}...\hat{\mathcal{K}}^n{}_{B_n}\psi$, where ψ begins with a "splitting" operator: let $IMG(R) = \mathrm{img}(\mathrm{img}(...\mathrm{img}(R,\sim_{B_1}^{\hat{\mathcal{K}}^1}),...),\sim_{B_n}^{\hat{\mathcal{K}}^n})$. Then $M, IMG(Q') \models \psi$ and $IMG(Q) \subseteq IMG(Q')$, which implies that $M, IMG(Q) \models \psi$, so $M,Q \models \varphi$.

($\widetilde{\mathbf{T}}$) Note that $\forall_{q\in Q} M,q \models \mathbb{K}_a\varphi \to \varphi$ (by $\mathbf{T}$), which implies that $M,Q \models \mathbb{K}_a\varphi \rightsquigarrow \varphi$ (by Proposition 6).

□

Thus, both versions of the T axioms hold for $\mathcal{L}^-$. It might be argued that these weaker versions of T are more appropriate properties of constructive knowledge than the full truth axioms, since the semantics of e.g. (weak) negation immediately following a constructive knowledge operator is different from the semantics of negation following a traditional knowledge operator. Note that, by Proposition 9, the meaning of weak negation or conjunction in the immediate scope of a constructive knowledge operator is the same as if the operator was immediately outside the constructive knowledge operator. Corresponding results can also be shown for the $\mathbb{C}, \mathbb{D}, \mathbb{E}$ operators (the proof is essientially the same as for Proposition 9). In other words, every formula in our full logical language is equivalent to one in $\mathcal{L}^-$. Thus, we can restrict the logical language to $\mathcal{L}^-$ without loosing expressive power. Appearently we then "get" the T axiom (i.e., if we take $\mathcal{L}^-$ to be the logical language), but it must be noted that even though the two languages are expressively equivalent, the extension of the schema T is *different* in

[4] In fact, it is enough to require that, between every occurence of constructive knowledge ($\mathbb{C}_A, \mathbb{E}_A, \mathbb{D}_A$) and weak negation ($\neg$), there is always at least one operator other than $\wedge$.

$\mathcal{L}^-$ and the full language (for example, $\mathbb{K}_a\neg p \rightarrow \neg p$ is a full language instance of T, but even though it is equivalent to the $\mathcal{L}^-$ formula $\neg\mathbb{K}_a p \rightarrow \neg p$ the latter is *not* a $\mathcal{L}^-$ instance of T). Consequently, the axiom schemata K, D and 5 cannot be written as in Theorem 11, but they can still be expressed with equivalent formulae.

Moreover, it is hard to pinpoint intuitive meaning of weak negation following immediately constructive knowledge: note that e.g. $\mathbb{K}_a\neg\langle\langle a\rangle\rangle\varphi$ should be read as "*a* has *constructive* knowledge about being *unable* to achieve φ". ($\mathbb{K}_a\langle\langle a\rangle\rangle\neg\varphi$, on the other hand, makes perfect sense: it refers to *a*'s constructive *ability* to prevent φ.) To sum up the discussion about $\mathbf{T}$ (and $\widetilde{\mathbf{T}}$), it seems, first, that the weaker versions given in Theorem 13 are more proper for constructive knowledge, and second, that it might be a good idea to consider the logical language for our logic of constructive knowledge to be limited to $\mathcal{L}^-$.

4.3 Expressing Weak Negation via Strong Negation

It seems worth pointing out that, if we are interested only in weak (initial) validity of formulae (i.e. we want to evaluate the formulae in single states; note, though, that this may require evaluation of subformulae in sets of states), weak negation ($\neg$) is not really necessary in the language. The reasons are as follows:

- We have already argued that $\|/\sim$ are more natural as the disjunction/negation operators in the propositional case.
- Strong negation ($\sim$) is sufficient to define K_a from $\mathbb{K}_a$.
- Basically, we want to initially evaluate the truth of a formula with respect to a single state. The only circumstance in which we have to evaluate a subformula with respect to a set of states, is when it occurs within the scope of a $\mathbb{K}_a$ operator (or a $\mathbb{C}_A, \mathbb{D}_A, \mathbb{E}_A$ operator). But according to Proposition 9 (and corresponding results for the $\mathbb{C}, \mathbb{D}, \mathbb{E}$ operators), we can move all $\neg$ and $\wedge$ operators *outside* that scope. Thus, we never need to evaluate them with respect to a non-singular set. By Proposition 7, the interpretation of $\neg$ and $\wedge$ in a single state is the same as the interpretation of $\sim$and $\|$. For instance:

$$M, q \models \mathbb{K}_a(\langle\langle A\rangle\rangle\Box\varphi \vee \langle\langle A\rangle\rangle\Box\psi) \Leftrightarrow \text{(by Prop. 9)}$$
$$M, q \models \mathbb{K}_a\langle\langle A\rangle\rangle\Box\varphi \vee \mathbb{K}_a\langle\langle A\rangle\rangle\Box\varphi \Leftrightarrow \text{(by Prop. 7)}$$
$$M, q \models \mathbb{K}_a\langle\langle A\rangle\rangle\Box\varphi \parallel \mathbb{K}_a\langle\langle A\rangle\rangle\Box\varphi.$$

- For subformulae "directly" inside a cooperation modality, the strong/weak operators also coincide since cooperation modalities split sets of states. For example, $(\mathbb{K}_a\langle\langle A\rangle\rangle\Box\neg\varphi) \leftrightarrow (\mathbb{K}_a\langle\langle A\rangle\rangle\Box \sim\varphi)$ is valid.

We have argued informally that weak negation does not add to the expressiveness with respect to (weak) validity. A full formal treatment of this, and related questions, is going to be reported in a future work.

5 Conclusions

In this paper, we propose a non-standard semantics for the modal logic of strategic ability under incomplete information, in which formulae are interpreted over *sets of states* rather than single states. We also propose new epistemic operators for "constructive" knowledge. It turns out that, in this new semantics, simple cooperation modalities $\langle\langle A \rangle\rangle$ can be combined with "constructive" epistemic operators into sufficiently expressive formulae. Indeed, the new logic is strictly more expressive than most existing ATL versions for incomplete information, while it retains the same model checking complexity as the least costly of them. The philosophical dimension of constructive knowledge is also natural: the constructive knowledge operators capture the notion of *knowing "de re"*, while the standard epistemic operators refer to *knowing "de dicto"*. Moreover, it turns out that standard (traditional) knowledge can be expressed through a combination of constructive knowledge and a negation operator that we call "strong" negation.

We argue that, if we are ultimately only interested in (weak) validity (i.e., traditional satisfaction in single states), then the weak negation operator is redundant. Moreover, most of the usual S5 properties hold (with the notable exception of the truth axiom T), and if we restrict the syntax so that constructive knowledge is never immediately followed by a "neutral" operator, we do not lose expressive power and the schema T becomes a validity.

We believe that we have finally obtained a satisfying logic of agents' strategies under uncertainty, and at the same time came up with novel, meaningful epistemic operators that capture important properties of the interaction between knowledge, action and ability. In future work, we plan to investigate further the expressive power of various operators in our semantics.

References

[1] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time Temporal Logic. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 100–109. IEEE Computer Society Press, 1997.

[2] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time Temporal Logic. *Lecture Notes in Computer Science*, 1536:23–60, 1998.

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

[4] E. A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 995–1072. Elsevier Science Publishers, 1990.

[5] V. Goranko. Coalition games and alternating temporal logics. In J. van Benthem, editor, *Proceedings of TARK VIII*, pages 259–272. Morgan Kaufmann, 2001.

[6] V. Goranko and W. Jamroga. Comparing semantics of logics for multi-agent systems. *Synthese*, 139(2):241–280, 2004.

[7] W. Jamroga. Some remarks on alternating temporal epistemic logic. In B. Dunin-Keplicz and R. Verbrugge, editors, *Proceedings of Formal Approaches to Multi-Agent Systems (FAMAS 2003)*, pages 133–140, 2003.

[8] W. Jamroga and J. Dix. Model checking strategic abilities of agents under incomplete information. In M. Coppo, E. Lodi, and G.M. Pinna, editors, *Proceedings of ICTCS 2005*, volume 3701 of *Lecture Notes in Computer Science*, pages 295–308. Springer Verlag, 2005.

[9] W. Jamroga and W. van der Hoek. Agents that know how to play. *Fundamenta Informaticae*, 63(2–3):185–219, 2004.

[10] W. Jamroga and W. van der Hoek. Strategic ability under uncertainty. Technical Report IfI-06-05, Clausthal University of Technology, 2005.

[11] G. Jonker. Feasible strategies in Alternating-time Temporal Epistemic Logic. Master thesis, University of Utrecht, 2003.

[12] R.C. Moore. A formal theory of knowledge and action. In J. Hobbs and R.C. Moore, editors, *Formal Theories of the Commonsense World*. Ablex Publishing Corp., 1985.

[13] L. Morgenstern. Knowledge and the frame problem. *International Journal of Expert Systems*, 3(4), 1991.

[14] J.F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences U.S.A.*, 36:48–49, 1950.

[15] M. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[16] M. Pauly. Game logic for game theorists. Technical Report INS-R0017, CWI, 2000.

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

[18] R. Pfeifer and C. Scheier. *Understanding Intelligence*. MIT Press : Cambridge, Mass, 1999.

[19] W. Quine. Quantifiers and propositional attitudes. *Journal of Philosophy*, 53:177–187, 1956.

[20] P. Y. Schobbens. Alternating-time logic with imperfect recall. *Electronic Notes in Theoretical Computer Science*, 85(2), 2004.

[21] M. Shubik. Game theory, complexity, and simplicity. Part III: Critique and prospective. Technical Report 98-04-029, Santa Fe Institute, 1998.

[22] W. van der Hoek and M. Wooldridge. Tractable multiagent planning for epistemic goals. In C. Castelfranchi and W.L. Johnson, editors, *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-02)*, pages 1167–1174. ACM Press, New York, 2002.

[23] W. van der Hoek and M. Wooldridge. Cooperation, knowledge and time: Alternating-time Temporal Epistemic Logic and its applications. *Studia Logica*, 75(1):125–157, 2003.

[24] S. van Otterloo and G. Jonker. On Epistemic Temporal Strategic Logic. In *Proceedings of LCMAS*, pages 35–45, 2004.

[25] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press: Princeton, NJ, 1944.

[26] T. Winograd and F. Flores. *Understanding Computers and Cognition*. Ablex Publishing, 1986.

[27] M. Wooldridge. *Reasoning about Rational Agents*. MIT Press : Cambridge, Mass, 2000.

Impressum

Publisher: Institut für Informatik, Technische Universität Clausthal
Julius-Albert Str. 4, 38678 Clausthal-Zellerfeld, Germany
Editor of the series: Jürgen Dix
Technical editor: Wojciech Jamroga
Contact: wjamroga@in.tu-clausthal.de
URL: http://www.in.tu-clausthal.de/~wjamroga/techreports/

The IfI Review Board