HAL
open science

# Logics of knowledge and action: critical analysis and challenges

Andreas Herzig

## To cite this version:

Andreas Herzig. Logics of knowledge and action: critical analysis and challenges. Autonomous Agents and Multi-Agent Systems, 2015, 29 (5), pp.719-753. 10.1007/s10458-014-9267-z . hal-01285093

**HAL Id: hal-01285093**

**https://hal.science/hal-01285093v1**

Submitted on 8 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 15176

**To link to this article** : DOI :10.1007/s10458-014-9267-z
URL : http://dx.doi.org/10.1007/s10458-014-9267-z

**To cite this version** : Herzig, Andreas *Logics of knowledge and action: critical analysis and challenges.* (2015) Autonomous Agents and Multi-Agent Systems, vol. 29 (n° 5). pp. 719-753. ISSN 1573-7454

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

# Logics of knowledge and action: critical analysis and challenges

**Andreas Herzig**

**Abstract** We overview the most prominent logics of knowledge and action that were proposed and studied in the multiagent systems literature. We classify them according to these two dimensions, knowledge and action, and moreover introduce a distinction between individual knowledge and group knowledge, and between a nonstrategic an a strategic interpretation of action operators. For each of the logics in our classification we highlight problematic properties. They indicate weaknesses in the design of these logics and call into question their suitability to represent knowledge and reason about it. This leads to a list of research challenges.

**Keywords** Logic of action · Logic of knowledge · Common knowledge · Frame problem · Uniform strategy

## 1 Introduction

A central issue in the study of multiagent systems (MAS) is the formal analysis of its *relevant concepts* as well as their interplay. Agents are typically analysed in terms of the concepts of knowledge, belief, desire, goal, and intention; their interaction is typically analysed in terms of the concepts of action, event, and time. Following a well-established tradition in philosophical logic and in the artificial intelligence subfield of knowledge representation, many researchers have directed their attention to *formal logics* in order to account for these concepts and design appropriate logical languages enabling formal reasoning.

In this paper we focus on the logics of two concepts playing a fundamental role in MAS: knowledge and action. They are of utmost importance in any area of research whose aim is to model intelligent agents, be they human or artificial. Typical situations to be modelled are when agents with imperfect knowledge perform actions in order to achieve goals. Often

A. Herzig (✉)
CNRS and University of Toulouse, IRIT, 118 Route de Narbonne, 31062 Toulouse Cedex 9, France
e-mail: herzig@irit.fr
http://www.irit.fr/∼Andreas.Herzig

the resulting models involve not only *individual knowledge* and *individual actions*, but also *group knowledge* and *joint actions*. For example, the joint actions of speakers and hearers in a dialogue build the common ground, which has been analysed in the literature as a kind of group knowledge.

The main logics of knowledge and action that were proposed in the philosophical and computer science literature are *modal logics*: extensions of classical propositional logic by *modal operators*. The latter enable us to talk in a natural way about what an agent knows and what the effects of his actions are. For example, the modal operator of knowledge is written $\mathsf{K}_i$, where $i$ is an agent name and where the formula $\mathsf{K}_i\varphi$ reads "$i$ knows that $\varphi$"; and the modal operator of ability is written $\langle i\rangle$, where the formula $\langle i\rangle\varphi$ reads "$i$ can achieve $\varphi$". These two kinds of operators combine in a natural way: for example, the formula $\langle i\rangle\mathsf{K}_j\varphi$ expresses that $i$ can achieve that $j$ knows that $\varphi$, and the formula $\mathsf{K}_i\neg\langle j\rangle\varphi$ expresses that $i$ knows that $j$ cannot achieve $\varphi$. Furthermore, we distinguish two different readings of action operators. In a *nonstrategic* reading we say that $i$ can achieve $\varphi$ if the other agents do not act; in a *strategic* reading we say that $i$ can achieve $\varphi$ whatever the other agents do. We reserve the traditional 'diamond' modal operator $\langle i\rangle$ for the former reading and use $[\![i]\!]$ for the latter, combining a modal diamond and a modal box in a way that renders its 'exists/for all' quantification. The notation generalises from individual agents to groups.

The first contribution of the present paper is a classification of the main logics of action and knowledge along the knowledge dimension and the action dimension. In the knowledge dimension we distinguish logics that do not account for knowledge at all, logics that only account for individual knowledge, and logics that account for group knowledge; and in the action dimension we distinguish logics that do not account for actions at all, logics that only have nonstrategic action operators, and logics that have strategic action operators. This makes up a $3\times3$ grid. Its simplest element is classical propositional logic: the logic which neither accounts for knowledge nor for action. The other eight categories are extensions of classical propositional logic by particular families of modal operators.

Our second and main contribution is a critical assessment of the most prominent existing logics in each of the eight relevant categories. For all but classical propositional logic we highlight one or more problematic properties. They can be traced back either to the design of the logical language—some natural statements cannot be expressed—or to the design of the logic itself—either too many or too little theorems. The former property affects the suitability of these logics for knowledge representation while the latter affects their suitability for drawing appropriate conclusions from the knowledge once it has been represented. Some of these problems had already been identified in particular in the philosophical literature. However, they were often ignored in the MAS and artificial intelligence literature.

We present each logic *syntactically*: we entirely omit the semantical aspects of the logics (which are quite involved in some cases) and instead list and discuss for each logic under examination its logical principles, i.e., its axioms and inference rules. This suffices to explain the problems that we want to point out, while providing at the same time a concise and uniform presentation. It also squares well with our knowledge representation perspective. This said, we are aware that much of the work on logics for MAS is heavily semantic-driven, foremost in the temporal logic literature where there was a 'theorem proving vs. model checking' debate [49] and where model checking is the most important reasoning technique. The model checking perspective was also adopted by several authors in the epistemic logic literature [75,89]. In the 'theorem proving vs. model checking' debate the main argument for the latter was better chances for decidability and a generally lower complexity. We would however like to separate such more mathematical concerns from our present more philosophical concerns: the conceptual analysis should precede the investigation of computational properties, and we

believe that the conceptual weaknesses that we are going to highlight are important enough to justify our enterprise.

The rest of the paper is organised as follows. In Sect. 2 we introduce two kinds of modal operators of knowledge and two kinds of modal operators of action, leading to a grid of nine classes of MAS logics. In the remaining sections we basically examine one exemplary logic per class, omitting the class having neither action nor knowledge operators. For two of the classes we present two logics. We start by examining logics without knowledge operators: Propositional Dynamic Logic (PDL) and Coalition Logic of Propositional Control (CL-PC) as logics with nonstrategic action operators (Sect. 3) and Alternating-time Temporal Logic (ATL) and Seeing-To-It-That Logic (STIT) as logics with strategic action operators (Sect. 4).The following three sections examine logics with operators of individual knowledge: the basic epistemic logic S5 (Sect. 5), Public Announcement Logic (PAL) as a logic with nonstrategic action operators (Sect. 6), and Alternating-time Epistemic Temporal Logic (ATEL) as a logic with strategic action operators (Sect. 7). The last three sections are about extensions of some underlying logic with the common knowledge operator: $\mathtt{S5}^{\mathsf{CK}}$ extends S5 (Sect. 8), $\mathtt{PAL}^{\mathsf{CK}}$ extends PAL (Sect. 9), and $\mathtt{ATEL}^{\mathsf{CK}}$ extends ATEL (Sect. 10). Based on the critiques of the preceding sections, Sect. 11 concludes by a list of research challenges.

## 2 A classification of MAS logics

We start by introducing a modal language that extends the language of classical propositional logic by *modal operators of knowledge* and *modal operators of action*. We distinguish two kinds of modal operators of knowledge and two kinds of modal operators of action.

Beyond the logical operators our language has the following non-logical symbols: propositional variables, action names, and agent names. More precisely, we suppose given

- a countable set of propositional variables $\mathbb{P} = \{p, q, q_1, \ldots\}$;
- a countable set of action names $\mathbb{A} = \{a, b, b_1, \ldots\}$;
- a finite set of agent names $\mathbb{I} = \{i, j, j_1, \ldots\}$.

Nonempty subsets of $\mathbb{I}$ are called *groups* or *coalitions*.[1] For a given coalition $J$, the elements of its complement $\overline{J} = \mathbb{I} \setminus J$ are called *the opponents of* $J$. Throughout the paper we use the term 'group' in general, while we use 'coalition' only when we reason strategically, i.e., when we take into account what the coalition's opponents might do (cf. the readings of the ability operators in Sect. 2.2).

All the logics in this paper will be presented syntactically, in terms of axiomatisations made up of axiom schemas and inference rules. All of them have a classical basis that we do not list explicitly: some axiom schemas for classical propositional logic plus the inference rule of modus ponens. A *deduction* of a formula $\varphi$ from a set of formulas $\Gamma$ is a finite sequence of formulas whose last element is $\varphi$ and which is such that each element of the sequence is either an element of $\Gamma$, or an instance of an axiom schema, or is obtained from preceding elements in the sequence by an inference rule. This is called deduction with global hypotheses in modal logic. One may also define deduction with local hypotheses; the definition is a bit more complicated. When $\varphi$ is deducible from $\Gamma$ with the axiomatisation $\mathcal{L}$ then we write

[1] The set of agent names cannot be infinite if we want the set of all groups of agents to be enumerable—for example, if $\mathbb{I}$ were allowed to be the set of natural numbers then the set of subsets of $\mathbb{I}$ would be uncountable—, and enumerability is necessary to obtain finite axiomatisations. Another way out is to only consider finite subsets of an infinite set of agents. We refer to [48] for non-finite axiomatisations of logics of group knowledge.

$\Gamma \vdash_{\mathcal{L}} \varphi$. When $\Gamma = \{\chi\}$ is a singleton we write $\chi \vdash_{\mathcal{L}} \varphi$, and when $\gamma$ is empty we write $\vdash_{\mathcal{L}} \varphi$ instead of $\emptyset \vdash_{\mathcal{L}} \varphi$. In the latter case the deduction is called a *proof*, and provable formulas are called *theorems*.

## 2.1 Modal operators of knowledge

The modal operators of knowledge are parametrised by either an individual agent or a set of agents. In the former case of individual knowledge, modal formulas take the form $\mathsf{K}_i\varphi$ where $\varphi$ is a formula and are read "agent $i$ knows that $\varphi$". In the latter case we distinguish several kinds of group knowledge: shared, common and distributed.

A formula $\varphi$ is *shared knowledge* in the group of agents $J$, written $\mathsf{EK}_J\varphi$, if each agent in $J$ knows that $\varphi$. $\mathsf{EK}_J\varphi$ is therefore identical to the conjunction of individual knowledge:

$$\mathsf{EK}_J\varphi \leftrightarrow \bigwedge_{i\in J} \mathsf{K}_i\varphi.$$

In most presentations $\mathsf{EK}_J\varphi$ is actually defined as an abbreviation of $\bigwedge_{i\in J} \mathsf{K}_i\varphi$. The concept can therefore be added without any technical difficulties to a logic of individual knowledge, and we do not present it in more detail here.

A formula $\varphi$ is *common knowledge* in $J$, written $\mathsf{CK}_J\varphi$, if it is shared knowledge that $\varphi$, it is shared knowledge that it is shared knowledge that $\varphi$, and so on. Intuitively, $\mathsf{CK}_J\varphi$ is therefore equivalent to an infinite conjunction:

$$\mathsf{CK}_J\varphi \leftrightarrow \mathsf{EK}_J\varphi \wedge \mathsf{EK}_J\mathsf{EK}_J\varphi \wedge \mathsf{EK}_J\mathsf{EK}_J\mathsf{EK}_J\varphi \wedge \cdots .$$

Such an infinite conjunction is however not a well-formed formula: differently from the shared knowledge operator, the common knowledge operator cannot be defined from the modal operators of individual knowledge by means of an abbreviation. The way out is to add the modal operators $\mathsf{CK}_J$ to the language and to give them the status of full-fledged logical connectives.

A formula $\varphi$ is *distributed knowledge* in group $J$ if $\varphi$ follows from the combined knowledge of the agents in $J$. It is somewhat difficult to grasp the concept syntactically beyond that informal definition. We here do so only in an approximate manner and view distributed knowledge as 'hypothetical common knowledge': $\varphi$ is distributed knowledge in the group of agents $J$, written $\mathsf{DK}_J\varphi$, if it is the case that when each agent communicates all his individual knowledge to all the other agents in $J$ then $\varphi$ becomes common knowledge in $J$. We do not go into the details for two reasons: first, the standard way of presentating distributed knowledge is semantical (in terms of the intersection of the accessibility relations of the group members) and cannot be characterised axiomatically (because intersection of accessibility relations is not modally definable); second, its semantics has some known conceptual problems: there are Kripke models where formally there is distributed knowledge that $\varphi$, while intuitively this is not the case [83,94]. Moreover, distributed knowledge was not studied to the same extent as common knowledge. We are going to discuss it briefly in two places: in Sect. 9.3 we explain the shortcomings of our definition of distributed knowledge as hypothetical common knowledge when examining the extension of Public Announcement Logic (`PAL`) with group knowledge, and in Sect. 10 we discuss the role of distributed knowledge when investigating uniform strategies in logics with group knowledge.

Common knowledge implies shared knowledge, shared knowledge implies individual knowledge, individual knowledge implies distributed knowledge, and distributed knowledge implies truth; in formulas, we have $\mathsf{CK}_J\varphi \to \mathsf{EK}_J\varphi$, $\mathsf{EK}_J\varphi \to \mathsf{K}_i\varphi$ if $i \in J$, $\mathsf{K}_i\varphi \to \mathsf{DK}_J\varphi$

**Table 1** The modal operators of knowledge

| Type of operator | Notation | Reading |
|---|---|---|
| Individual | $\mathsf{K}_i\varphi$ | "Agent $i$ knows that $\varphi$" |
| Group | $\mathsf{EK}_J\varphi$ | "It is shared knowledge in $J$ that $\varphi$", |
| | | "Every agent in $J$ knows that $\varphi$" |
| | $\mathsf{CK}_J\varphi$ | "It is common knowledge in $J$ that $\varphi$" |
| | $\mathsf{DK}_J\varphi$ | "It is distributed knowledge in $J$ that $\varphi$" |

if $i \in J$, and $\mathsf{DK}_J\varphi \to \varphi$. In the case of singleton groups, all three concepts of group knowledge coincide with individual knowledge: the formulas $\mathsf{CK}_{\{i\}}\varphi$, $\mathsf{EK}_{\{i\}}\varphi$, $\mathsf{K}_i\varphi$, and $\mathsf{DK}_{\{i\}}\varphi$ are all equivalent.

The different modal operators of knowledge and their readings are listed in Table 1.

## 2.2 Modal operators of action and ability

The modal operators of action relate an agent or group of agents to a formula that is viewed as the outcome of the agent's action.

Most of the MAS logics deal with potential action, as opposed to actual action: they talk about the ability of an agent or a group of agents to achieve some outcome, and not about what they actually do. We are therefore mainly going to deal with *modal operators of ability*.

Modal operators of ability can be interpreted in two ways. In the *nonstrategic* interpretation it is supposed that no other action is performed: the outcome is achieved *ceteris paribus*. In the *strategic* interpretation the outcome is achieved whatever the other agents do. (There seems to be no scholarly name for that opposite of ceteris paribus; *ceteris mutandis*, *ceteris variantis*, or *ceteris agentis* might be suitable.) The strategic interpretation is appropriate to reason about strategic games, where one studies whether an agent or a coalition of agents has a winning strategy.

- In the nonstrategic interpretation, modal formulas take the form $\langle J\rangle\varphi$ and are read "group $J$ can achieve $\varphi$ while the other agents don't act".
- In the strategic interpretation, modal formulas take the form $[\![J]\!]\psi$ and are read "coalition $J$ can achieve $\psi$ whatever the opponents do".[2]

In nonstrategic contexts we talk about groups, while in strategic contexts we talk about coalitions and their opponents.

Only the so-called modal logics of agency deal with actual action. We here limit ourselves to one such operator: the operator of seeing-to-it-that, abbreviated 'stit' [19,64]. Agency formulas take the form $\mathsf{Stit}_J\varphi$ and are read "$J$ achieves $\varphi$ whatever the opponents do". There exists another tradition studying a similar modal operator of agency that is read "$i$ brings it about that" [34,60,90]. We do not discuss it here because the logics of these operators lack the modal operator of historical possibility: the latter will be instrumental for us in Sect. 7, where we relate stit logics to other strategic logics in order to solve the problems occurring there.

All the modal operators of action that we have introduced up to now relate agents and outcomes, but leave implicit the action the agent performs to achieve the outcome. There exist

[2] We have chosen the notation $[\![J]\!]$—the nesting of a modal diamond and a modal box—in order to signal a $\forall\exists$ quantification. It is however not standard in the literature, where one can find $\langle J\rangle$ in Pauly's Coalition Logic CL and $\langle\langle J\rangle\rangle$ in Alternating-time Temporal Logic (ATL).

**Table 2** The modal operators of ability and action

| Type of operator | Notation | Reading |
|---|---|---|
| Nonstrategic | $\langle \pi \rangle \varphi$ | "Action $\pi$ can achieve $\varphi$ if the other agents do nothing", "There is an execution of program $\pi$ after which $\varphi$" |
| | $\langle J \rangle \varphi$ | "Group $J$ can achieve $\varphi$ while the other agents don't act" |
| Strategic | $\langle\!\langle J \rangle\!\rangle \psi$ | "Coalition $J$ can achieve $\psi$ whatever the opponents do" |
| | $\mathsf{Stit}_J \varphi$ | "Coalition $J$ achieves $\varphi$ whatever the opponents do" |

**Table 3** The grid of MAS logics

| | Knowledge | | |
|---|---|---|---|
| Action | No actions | Nonstrategic action | Strategic action |
| No uncertainty | | CL-PC; PDL | ATL; STIT |
| Individual knowledge | S5 | PAL | ATEL |
| Group knowledge | $\mathtt{S5}^{\mathsf{CK}}$ | $\mathtt{PAL}^{\mathsf{CK}}$ | $\mathtt{ATEL}^{\mathsf{CK}}$ |

logics with *names for actions* in the language. Beyond such atomic actions they also allow to talk about *complex actions*, that may also be called *plans*. Complex actions are built from atomic actions by means of program operators such as sequential composition and iteration. Modal formulas take the form $\langle \pi \rangle \varphi$ and are read "there is an execution of plan $\pi$ after which $\varphi$". In principle such so-called dynamic operators can be interpreted both in a nonstrategic and in a strategic way. The only standard logic in the literature with such dynamic operators, PDL, comes with the first interpretation. However and as we will discuss in Sect. 4, logics with strategic operators $\langle\!\langle J \rangle\!\rangle$ that are indexed by actions committing some agents to perform some actions were proposed recently in the literature. When the committed agents are exactly those of $J$ then we obtain a dynamic operator with a strategic interpretation. While the standard syntax of PDL does not provide agent names, we suppose that each action name $a \in \mathbb{A}$ comes with an agent name $i \in \mathbb{I}$ identifying the author of $a$. This is written $a_i$ or $i{:}a$. The logical language is extended in this way in many papers in the MAS literature, for example in [55,65,84].

The different modal operators of ability and action are summarised in Table 2.

### 2.3 The grid of MAS logics

Our classification of logics according to the knowledge and the action dimension leads to the nine classes of logics that are depicted in the 3×3 grid in Table 3. We do not mention there any logic lacking both knowledge operators and action operators: this would be good old classical propositional logic, which it is not interesting for our enterprise.

For each of the remaining eight classes we give one prominent logic. As we have motivated above, we discuss further representatives for the two classes of logics without knowledge operators: in the nonstrategic category we discuss PDL and in the strategic category we discuss STIT.

In the rest of the paper we are going to critically examine each of the ten logics in our grid.

## 3 No uncertainty, nonstrategic actions: PDL and CL-PC

We start by recalling PDL, which is a logic with names for actions, and show that it suffers from the frame problem. We then present a dialect of PDL solving the frame problem, Dynamic Logic of Propositional Assignments (DL-PA). We finally recall CL-PC, which is a logic without explicit actions, and present its embedding into DL-PA.

### 3.1 Propositional Dynamic Logic (PDL)

The authors of PDL [51,52] were motivated by the verification of computer programs. Later, PDL was also applied in philosophy of action [100]. Its language has modal formulas of the form $\langle\pi\rangle\varphi$, where $\varphi$ is a formula and $\pi$ is a complex action—alias a plan or a program—, read "there exists a possible execution of $\pi$ after which $\varphi$" or "action $\pi$ possibly achieves $\varphi$ (if the opponents do nothing)". The latter reading indicates that in PDL, the interpretation of actions is nonstrategic. The modal operator $\langle\pi\rangle$ comprehends an existential quantification; a dual operator $[\pi]$ with a universal quantification is defined by the abbreviation $[\pi]\varphi \stackrel{\text{def}}{=} \neg\langle\pi\rangle\neg\varphi$. It therefore has to be read "$\varphi$ after every possible execution of $\pi$" or "$\pi$ necessarily achieves $\varphi$ (if the opponents do nothing)". The program $\pi$ is either an atomic program $a$, a sequential composition $\pi;\pi$, a nondeterministic composition $\pi\cup\pi$, a finite iteration $\pi^*$, or a test $\varphi?$, where $\varphi$ is a formula. The standard programming instructions can then be defined as PDL programs: for example, the program $\top?$ is the 'skip' action, $(\varphi?;\pi_1)\cup(\neg\varphi?;\pi_2)$ is nothing but the conditional "if $\varphi$ then $\pi_1$ else $\pi_2$" and $(\varphi?;\pi)^*;\neg\varphi?$ is the loop "while $\varphi$ do $\pi$".

The *star-free fragment of* PDL is the set of formulas without the iteration operator $^*$, that is also called 'Kleene star'.

As we have already said in Sect. 2.2, while the original language of PDL is not designed for multiple agents, one may nevertheless accommodate them by indexing atomic actions with agent names: we suppose that the members of the set of action names $\mathbb{A}$ have the form $i{:}b$ where $i$ is an agent and $b$ is an action name. For example, the formula $\langle i_1{:}b_1\rangle\top\wedge[i_2{:}b_1]\bot$ expresses that $i_1$ is able to perform $b_1$ and $i_2$ is not.

The axiomatisation of PDL is given in Table 4. The principles in the left column, $\mathtt{RE}(\langle\pi\rangle)$, $\mathtt{N}(\langle\pi\rangle)$, $\mathtt{M}(\langle\pi\rangle)$, $\mathtt{C}(\langle\pi\rangle)$, and $\mathtt{Dual}(\langle\pi\rangle)$, are common to all so-called normal modal logics [20,29]; $\mathtt{M}(\langle\pi\rangle)$ and $\mathtt{C}(\langle\pi\rangle)$ can be replaced by the axiom $\mathtt{K}(\langle\pi\rangle)$: $([\pi]\varphi\wedge[\pi](\varphi\to\psi))\to[\pi]\psi$. The right column of Table 4 is about the PDL program operators. Its schemas Seq, Nondet, and Test are equivalences without program operators on the right. It follows that for the star-free fragment, all program operators can be eliminated. For the iteration operator $^*$ there are two axioms: $\mathtt{FP}(\langle\pi^*\rangle)$ is a fixpoint axiom and $\mathtt{LFP}(\langle\pi^*\rangle)$ is a least fixpoint axiom.

When $\varphi$ is deducible from $\Gamma$ in PDL we write $\Gamma\vdash_{\mathtt{PDL}}\varphi$. The following version of the deduction theorem holds:

$$\chi\vdash_{\mathtt{PDL}}\varphi \text{ iff } \vdash_{\mathtt{PDL}}\big[(i_1{:}b_1\cup\cdots\cup i_n{:}b_n)^*\big]\chi\to\varphi,$$

where $\{i_1{:}b_1,\ldots,i_n{:}b_n\}$ is the set of atomic actions occurring in $\varphi$.

The problem of deciding the satisfiability of a PDL formula is PSPACE complete for the star-free fragment and EXPTIME complete for the full language [91].

### 3.2 Reasoning about ability and the frame problem

In PDL, one can describe an atomic action $i{:}b$ by means of formulas stating its preconditions and (possibly conditional) effects. These formulas respectively take the form $\mathit{Precond}\to\langle i{:}b\rangle\top$ and $\mathit{Cond}\to[i{:}b]\mathit{Effect}$. A set of such formulas makes up a theory

**Table 4** Axiomatisation of PDL

| | | | |
|---|---|---|---|
| RE($\langle\pi\rangle$) | $\dfrac{\varphi \leftrightarrow \varphi'}{\langle\pi\rangle\varphi \leftrightarrow \langle\pi\rangle\varphi'}$ | Seq | $\langle\pi_1;\pi_2\rangle\varphi \leftrightarrow \langle\pi_1\rangle\langle\pi_2\rangle\varphi$ |
| N($\langle\pi\rangle$) | $\neg\langle\pi\rangle\bot$ | Nondet | $\langle\pi_1 \cup \pi_2\rangle\varphi \leftrightarrow \langle\pi_1\rangle\varphi \vee \langle\pi_2\rangle\varphi$ |
| M($\langle\pi\rangle$) | $\langle\pi\rangle(\varphi \vee \psi) \rightarrow \big(\langle\pi\rangle\varphi \vee \langle\pi\rangle\psi\big)$ | Test | $\langle\varphi?\rangle\psi \leftrightarrow \varphi \wedge \psi$ |
| C($\langle\pi\rangle$) | $\big(\langle\pi\rangle\varphi \vee \langle\pi\rangle\psi\big) \rightarrow \langle\pi\rangle(\varphi \vee \psi)$ | FP($\langle\pi^*\rangle$) | $\langle\pi^*\rangle\varphi \leftrightarrow \varphi \vee \langle\pi\rangle\langle\pi^*\rangle\varphi$ |
| Dual($\langle\pi\rangle$) | $[\pi]\varphi \leftrightarrow \neg\langle\pi\rangle\neg\varphi$ | LFP($\langle\pi^*\rangle$) | $\langle\pi^*\rangle\varphi \rightarrow \big(\varphi \vee \langle\pi^*\rangle(\neg\varphi \wedge \langle\pi\rangle\varphi)\big)$ |

of actions *ActionTheory*. Given *ActionTheory*, one typically wants to check whether given the description of the initial state *Init*, the group of agents $\{i_1, \ldots, i_n\}$ is possibly able to achieve the goal *Goal* by means of a sequence of atomic actions $i_1{:}b_1; \cdots; i_n{:}b_n$. This is the most basic multiagent version of a prediction task. More elaborate versions with synchronous or asynchronous execution of actions can be designed [35]. Formally, the above prediction task amounts to checking whether $\mathit{Init} \rightarrow \langle i_1{:}b_1; \cdots; i_n{:}b_n\rangle \mathit{Goal}$ can be deduced from *ActionTheory*, i.e., to checking whether

$$\mathit{ActionTheory} \vdash_{\mathtt{PDL}} \mathit{Init} \rightarrow \langle i_1{:}b_1; \cdots; i_n{:}b_n\rangle \mathit{Goal}.$$

The above formula only guarantees that the group is possibly able to achieve $\varphi$. If actions are nondeterministic then we often rather want to know whether the group is necessarily able to achieve $\varphi$. We then have to check the deducibility of the following formula:

$$\mathit{ActionTheory} \vdash \mathit{Init} \rightarrow \big(\langle i_1{:}b_1; \cdots; i_n{:}b_n\rangle\top \wedge [i_1{:}b_1; \cdots; i_n{:}b_n]\mathit{Goal}\big).$$

Let us suppose we want to formulate an action theory for a domain containing the action $i{:}\mathit{moveBlock}_{L_1,L_2}$ of agent $i$ moving some block from location $L_1$ to $L_2$. Suppose the set of propositional variables $\mathbb{P}$ contains *BlockRed*. Then the action theory has to contain the two implications

$$\mathit{BlockRed} \rightarrow [i{:}\mathit{moveBlock}_{L_1,L_2}]\mathit{BlockRed},$$
$$\neg\mathit{BlockRed} \rightarrow [i{:}\mathit{moveBlock}_{L_1,L_2}]\neg\mathit{BlockRed}$$

in order to guarantee that the colour of the block is not altered by moving it. Formulas of that form are called *frame axioms*. Intuitively, action theories should not contain such formulas; instead, they should rather be deducible from it. This however fails to be the case here: the logic PDL is too weak and suffers from the so-called *frame problem* [82]. Formally and according to Reiter [93], an appropriate reasoning about actions formalisms should have action descriptions with size much smaller than $Card(\mathbb{P}) \times Card(\mathbb{I})$ (assuming that $\mathbb{P}$ is finite). His argument appeals to the *principle of inertia*, which basically says that an action only modifies the truth value of very few propositional variables and leaves the others unchanged.

In contrast, the addition of frame axioms to a PDL action theory results in a set of formulas whose size is in the order of $Card(\mathbb{P}) \times Card(\mathbb{I})$.

### 3.3 Reiter's solution to the frame problem

The frame problem not only plagues PDL, but virtually every logic-based framework for reasoning about a dynamic system. Interestingly, it was also investigated during the last decade in theoretical computer science, where separation logic is considered to be a solution that enables local reasoning about memory structures such as heaps or chained lists [87]. In

the artificial intelligence field of reasoning about actions, a lot of dedicated logical formalisms were developed in the eighties in order to solve the frame problem, including the Situation Calculus [82], the Event Calculus [101], the Fluent Calculus [103], and so-called action languages such as $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{C}+$, and most recently $\mathcal{BC}$ [41,71].

One of the most popular solutions to the frame problem is Reiter's [92,93] in terms of what he calls *successor state axioms*. While Reiter's original framework was the Situation Calculus, his solution can be imported into PDL if we augment it by variables for action names, quantifiers and the equality predicate. We call the resulting logic $\mathtt{PDL}^{\forall}$. We know of no result in the literature about the precise relationship with first-order dynamic logic [51,52] where variables do not range over atomic programs but over objects of the world (just as in first-order logic) and just note that the two logics are *a priori* different.

The successor state axiom for the propositional variable *BlockRed* can be formulated in $\mathtt{PDL}^{\forall}$ as

$$\forall x \Big( [x]\mathit{BlockRed} \leftrightarrow \big( x{=}(i{:}\mathit{paintRed}) \vee (\mathit{BlockRed} \wedge x{\neq}(i{:}\mathit{paintBlue})) \big) \Big),$$

where in order to simplify we suppose that $i$ is the only agent who is able to paint and that the only available colours are red and blue. The general form of a successor state axiom for a propositional variable $p$ is

$$\forall x \big( [x]p \leftrightarrow \gamma_p(x) \big),$$

where $x$ is a variable ranging over the set of actions and $\gamma_p(x)$ is a boolean formula. It determines the truth value of $p$ after action $x$ given the truth value of $\gamma_p(x)$ before $x$. This presupposes complete knowledge about the pre- and postconditions of each action $a$. In other words, Reiter's solution does not work when there are integrity constraints, such as that a block cannot be both red and blue and that a block must have a colour. Taking such constraints into account would pose the ramification problem.

Reiter's *basic action theories* have one successor state axiom per propositional variable in $\mathbb{P}$. This solves the frame problem: the cardinality of a basic action theory is linear in that of $\mathbb{P}$, and if we suppose with Reiter that the length of each successor state axiom is bounded then we obtain an action theory whose size is also linear in that of $\mathbb{P}$.

While the basic version of Reiter's solution is for deterministic actions only, nondeterminism can be modelled by means of PDL's operator of nondeterministic composition.

To sum it up, the adoption of Reiter's solution seems to require quantification over action names, which is not available in PDL. However, as we will show in the sequel, Reiter's solution can be captured in a simple variant of PDL that stays in the propositional realm.

### 3.4 DL-PA: a variant of PDL solving the frame problem

It was shown in [114] that Reiter's basic action theories can be captured in a dialect of PDL. Its extension by the PDL program operators, called DL-PA, was further studied in [12,59]. We first describe that logic and then the embedding.

The language of DL-PA is just as that of PDL, except that atomic actions are not just abstract names: they do something concrete, namely to assign formulas $\varphi$ to propositional variables $p$. Such assignments are written $p{:=}\varphi$. The original language only contains assignments of $\top$ and $\bot$; however, the more general assignment $p{:=}\varphi$ can be considered to be an abbreviation of $(\varphi?;\ p{:=}\top) \cup (\neg\varphi?;\ p{:=}\bot)$. For example, $\mathit{BlockAt}_{L_1}{:=}\bot$ makes $\mathit{BlockAt}_{L_1}$ false. We here only present an axiomatisation of star-free DL-PA and refer the reader to the above papers

for the full logic. The four axiom schemas that have to be added to those of PDL are the following:

$$\begin{aligned} \langle p{:=}\varphi\rangle\psi &\leftrightarrow [p{:=}\varphi]\psi \\ \langle p{:=}\varphi\rangle p &\leftrightarrow \varphi \\ q &\leftrightarrow \langle p{:=}\varphi\rangle q \quad \text{for } q \neq p \end{aligned}$$

The first says that assignments are serial and deterministic. The second says that $p$ is true after the assignment of $\varphi$ to $p$ if and only if $\varphi$ was true before. The third says that assignments to $p$ do not modify other variables.

Reasoning in DL-PA has several advantages over PDL: the logic is compact and has the interpolation property; model checking has the same complexity as satisfiability checking; the Kleene star is eliminable (from which it follows that every DL-PA formula is reducible to a boolean formula). Complexity of satisfiability is PSPACE complete for its star-free fragment [59].

The fact that the execution of an assignment $p{:=}\varphi$ leaves unchanged the truth values of all those $q$ that are different from $p$ leads to a solution to the frame problem. For example, the action $i{:}moveBlock_{L_1,L_2}$ can be captured by the DL-PA program

$$i{:}moveBlock_{L_1,L_2} \;=\; (Rep_{i:moveBlock_{L_1,L_2}}?;\, Free?;\, BlockAt_{L_1}{:=}\bot;\, BlockAt_{L_2}{:=}\top),$$

where we suppose that the propositional variables $Rep_{i:moveBlock_{L_1,L_2}}$ and *Free* are true when the action $moveBlock_{L_1,L_2}$ is in agent $i$'s repertoire and when the block is free, respectively. For that action, the formulas

$$\begin{aligned} BlockRed &\rightarrow [i{:}moveBlock_{L_1,L_2}]BlockRed, \\ \neg BlockRed &\rightarrow [i{:}moveBlock_{L_1,L_2}]\neg BlockRed \end{aligned}$$

are both DL-PA theorems.

Let us make this precise. A successor state axiom $\forall x\big([x]p \leftrightarrow \gamma_p(x)\big)$ is *explicit* if for every $a$ that does not occur in $\gamma_p(x)$, $\gamma_p(a) \leftrightarrow p$ is a theorem of classical propositional logic. For such actions $[a]p \leftrightarrow p$ is therefore a theorem, which means that an action that does not occur in $\gamma_p(x)$ cannot modify $p$. Special cases are Reiter's original definition of successor state axioms [92] and local effect successor state axioms [118]. The successor state axiom for *BlockRed* in Sect. 3.3 is clearly explicit.

The theorem below basically says that for explicit basic action theories, action $a$'s effects according to Reiter are exactly the effects of the DL-PA assignment program

$$\pi_a \;=\; \big(q_1^a{:=}\gamma_{q_1^a}(a);\, \cdots;\, q_n^a{:=}\gamma_{q_n^a}(a)\big),$$

where $q_1^a, \ldots, q_n^a$ are all variables whose successor state axiom mentions $a$. It is formulated for the extension $\text{PDL}^\forall$ of PDL by action variables, quantification and equality within which we have presented Reiter's solution in Sect. 3.3 and follows from Theorem 18 of [114].[3]

**Theorem 1** *Let ActionTheory be a set of explicit successor state axioms, one per propositional variable $p$. Let $\varphi$ be some* PDL *formula. For each atomic program $a$, let $\{q_1^a, \ldots, q_n^a\}$ be the set of propositional variables $q_k^a$ such that $a$ occurs in the successor state axiom for*

[3] We here give a simplified version; the original formulation is not about the above extension $\text{PDL}^\forall$ of PDL, but about logic ES [70]. The latter is a variant of the Situation Calculus where situation terms are suppressed (and which is therefore closer to modal logics) and which moreover comes with sensing actions and epistemic operators.

*$q_k^a$ and such that $q_k^a$ occurs in $\varphi$. Let $\varphi'$ result from replacing each abstract atomic action $a$ in $\varphi$ by the complex assignment program $\pi_a = \big(q_1^a{:=}\gamma_{q_1^a}(a); \cdots; q_n^a{:=}\gamma_{q_n^a}(a)\big)$. Then*

$$\textit{ActionTheory} \vdash_{\texttt{PDL}^\forall} \varphi \ \textit{ if and only if } \ \vdash_{\texttt{DL-PA}} \varphi'.$$

The length of $\varphi'$ is quadratic in the length of $\varphi$, and the embedding of Reiter's solution to the frame problem is therefore polynomial [114, Theorem 27].

One might argue that the above result is not very impressive: after all, frame axioms can also be encoded in PDL by systematically augmenting action theories with a polynomial number of frame axioms. We have seen in the end of Sect. 3.2 that this however violates Reiter's criterion for a solution to the frame problem because the cardinality of the action theory is in the order of $Card(\mathbb{P}) \times Card(\mathbb{I})$. In contrast, the encoding into DL-PA is satisfactory because it is based on $Card(\mathbb{A})$ action descriptions $\pi_a$, one per action $a \in \mathbb{A}$. Differently from PDL and the Situation Calculus, these action descriptions appear directly in the formula, i.e., there are no action names that are axiomatised separately in an action theory.

We view DL-PA as a sort of Assembler language for reasoning about change. Its applicability to update and revision operations, planning tasks, argumentation theory and answer-set programming has been demonstrated recently [33,36,54,62]. In the rest of the section we show that it captures a logic of ability.

### 3.5 Coalition Logic of Propositional Control (CL-PC) and its embedding into DL-PA

Coalition Logic of Propositional Control (CL-PC) was introduced in [110,112]. It is a MAS logic with nonstrategic actions that does not account for uncertainty. Just as in boolean games [21,53], the idea is that each propositional variable is controlled by some agent. When $i$ controls $p$ then $i$ can modify the truth value of $p$ at will. An action of $i$ is therefore identified with the change of some of $i$'s variables. It is supposed that control is not only exclusive—each variable is controlled by at most one agent—, but also exhaustive —each variable is controlled by at least one agent.

The language of CL-PC relates agents to outcomes. It has formulas of the form $\langle J\rangle\varphi$, read "group $J$ can achieve $\varphi$ by modifying its variables while the opponents don't act". The dual $[J]\varphi$ may be read "group $J$ cannot prevent $\varphi$". The fact that $i$ controls $p$ can then be expressed by the CL-PC formula $\langle i\rangle p \wedge \langle i\rangle\neg p$.

We do not give the entire axiomatics of CL-PC here and refer the reader to [110]. Instead, the principles of the logic will become clear through the embedding into DL-PA that we state as Theorem 2. But let us first discuss two CL-PC theorems in order to limber up:

$$\big(\langle i\rangle p \wedge \langle i\rangle\neg p\big) \rightarrow [i]\big(\langle i\rangle p \wedge \langle i\rangle\neg p\big),$$
$$\big(\langle i\rangle p \wedge \langle i\rangle q\big) \rightarrow \langle i\rangle(p\wedge q).$$

The first theorem says that an agent's action repertoire is not changed by the performance of actions. The second theorem comes from the fact that control is exercised over atomic formulas and not over complex formulas. Note that these are theorems and not theorem schemas; for example, the replacement of $q$ by $\neg p$ in the last formula results in a non-theorem.

The focus of CL-PC is reasoning about nonstrategic (ceteris paribus) ability: given a set of CL-PC formulas *AbilityTheory* describing the abilities of the different agents by means of formulas of the form $\langle i\rangle p \wedge \langle i\rangle\neg p$, one may check whether

$$\textit{AbilityTheory} \vdash_{\texttt{CL-PC}} \textit{Init} \rightarrow \langle\{i_1, \ldots, i_n\}\rangle \textit{Goal}$$

holds, where *Init* describes some initial situation and *Goal* some goal. As the agents' repertoires are not changed by the performance of actions, this is the same as

$$\vdash_{\texttt{CL-PC}} (\mathit{AbilityTheory} \wedge \mathit{Init}) \rightarrow \langle\{i_1, \ldots, i_n\}\rangle \mathit{Goal}.$$

Let $\mathbb{P}_\varphi$ be the set of propositional variables occurring in $\varphi$. For every agent $i \in \mathbb{I}$ and variable $p \in \mathbb{P}_\varphi$, let $\mathit{Ctrl}_{i,p}$ be a fresh propositional variable (not occurring in $\varphi$); the intended meaning of $\mathit{Ctrl}_{i,p}$ is that $i$ controls $p$. Then the formulas

$$\begin{aligned} \mathit{Excl}_\varphi &= \bigwedge_{p \in \mathbb{P}_\varphi} \bigwedge_{i,j \in \mathbb{I}, i \neq j} \neg(\mathit{Ctrl}_{i,p} \wedge \mathit{Ctrl}_{j,p}), \\ \mathit{Exh}_\varphi &= \bigwedge_{p \in \mathbb{P}_\varphi} \bigvee_{i \in \mathbb{I}} \mathit{Ctrl}_{i,p} \end{aligned}$$

describe exclusivity and exhaustivity of the control over $\mathbb{P}_\varphi$.

Given a set of agents $J$ and a propositional variable $p$, we define the following program:

$$\mathit{vary}_{J,p} = \top? \cup \Big(\Big(\bigvee_{i \in J} \mathit{Ctrl}_{i,p_1}\Big)?;\ p_1{:=}\neg p_1\Big).$$

It either does nothing or flips the truth value of $p$ if $p$ is under the control of one of $J$'s members. For example:

$$\begin{aligned} \langle \mathit{vary}_{\{i\},p}\rangle p &= \big\langle \top? \cup (\mathit{Ctrl}_{i,p}?;\ p{:=}\neg p)\big\rangle p \\ &\leftrightarrow \langle \top?\rangle p \vee \langle \mathit{Ctrl}_{i,p}?\rangle\ \langle p{:=}\neg p\rangle p \\ &\leftrightarrow p \vee (\mathit{Ctrl}_{i,p} \wedge \neg p) \\ &\leftrightarrow p \vee \mathit{Ctrl}_{i,p}. \end{aligned}$$

Therefore $[\mathit{vary}_{\{i\},p}]p \leftrightarrow p \wedge \neg \mathit{Ctrl}_{i,p}$. Then for a finite set of propositional variables $P = \{p_1, \ldots, p_n\} \subseteq \mathbb{P}$, we define

$$\mathit{vary}_{J,P} = \mathit{vary}_{J,p_1}; \cdots; \mathit{vary}_{J,p_n}$$

which nondeterministically changes the truth value of some of the variables of $P$ that are under $J$'s control.

The theorem below is a consequence of Theorems 4 and 5 of [59].

**Theorem 2** *Let $\varphi$ be a* CL-PC *formula and let $\varphi'$ be the* DL-PA *formula resulting from replacing each subformula $\langle J\rangle\psi$ of $\varphi$ by $\langle \mathit{vary}_{J,\mathbb{P}_\psi}\rangle\psi$. Then*

$$\vdash_{\texttt{CL-PC}} \varphi \ \textit{ if and only if } \ \vdash_{\texttt{DL-PA}} (\mathit{Exh}_\varphi \wedge \mathit{Excl}_\varphi) \rightarrow \varphi'.$$

Note that the length of $\varphi'$ is polynomial in that of $\varphi$.

### 3.6 Strategic reasoning in CL-PC

It was argued in [110] that in CL-PC one can also reason about strategic ability: as pointed out there, the CL-PC formula $\langle J\rangle[\overline{J}]\varphi$ expresses that $J$ can achieve $\varphi$ next *whatever $J$'s opponents do*. It is claimed that this captures at least the spirit of the formula $\langle\!\langle J\rangle\!\rangle \mathsf{X}\varphi$ of Pauly's Coalition Logic CL. Indeed, if we take a CL theorem and replace its strategic operators $\langle\!\langle J\rangle\!\rangle \mathsf{X}$ by $\langle J\rangle[\overline{J}]$ then the resulting formula is a CL-PC theorem [110, Theorem 4.1].

In the rest of the section we have a look at the other direction.

Consider the following equivalence:

$$\varphi_0 \;=\; \langle J\rangle[\overline{J}](p\vee q) \leftrightarrow \langle J\rangle[\overline{J}]p \vee \langle J\rangle[\overline{J}]q$$

It basically says that if $J$ can achieve $p\vee q$ whatever its opponents do then either $J$ controls $p$ or $J$ controls $q$. We are going to establish that it is a theorem of CL-PC. The difficult part is to prove that $\psi_0 \;=\; [\overline{J}](p\vee q) \leftrightarrow [\overline{J}]p \vee [\overline{J}]q$ is a theorem of CL-PC; the rest follows by principles of normal modal logics. We prove that

$$\psi_0' \;=\; [vary_{\overline{J},\{p,q\}}](p\vee q) \leftrightarrow [vary_{\overline{J},\{p\}}]p \vee [vary_{\overline{J},\{q\}}]q$$

is a theorem of DL-PA, from which it follows by Theorem 2 that $\psi_0$ is a theorem of CL-PC. The following equivalences are all theorems, where we write $Ctrl_{J,p}$ instead of $\bigvee_{i\in J} Ctrl_{i,p}$:

$$\begin{aligned}
[vary_{\overline{J},\{p\}}]p\vee[vary_{\overline{J},\{q\}}]q \;&\leftrightarrow \big(p\wedge\neg Ctrl_{\overline{J},p}\big) \vee \big(q\wedge\neg Ctrl_{\overline{J},q}\big)\\
&\leftrightarrow \big(p \wedge (\neg Ctrl_{\overline{J},p}\vee\neg p)\big) \vee \big(q \wedge (\neg Ctrl_{\overline{J},q}\vee\neg q)\big)\\
&\leftrightarrow (p\vee q) \wedge \big(\neg Ctrl_{\overline{J},q}\vee p\vee\neg q\big) \wedge\\
&\quad \big(\neg Ctrl_{\overline{J},p}\vee\neg p\vee q\big) \wedge \big(\neg Ctrl_{\overline{J},p}\vee\neg Ctrl_{\overline{J},q}\vee\neg p\vee\neg q\big)\\
&\leftrightarrow [\top?][\top?](p\vee q) \wedge [Ctrl_{\overline{J},q}?](p\vee\neg q) \wedge\\
&\quad [Ctrl_{\overline{J},p}?](\neg p\vee q) \wedge [Ctrl_{\overline{J},p}?][Ctrl_{\overline{J},q}?](\neg p\vee\neg q)\\
&\leftrightarrow [\top? \cup \big(Ctrl_{\overline{J},p}?;\, p{:=}\neg p\big)][\top? \cup \big(Ctrl_{\overline{J},q}?;\, q{:=}\neg q\big)]\,(p\vee q)\\
&= [vary_{\overline{J},\{p,q\}}](p\vee q)
\end{aligned}$$

So $\psi_0'$ is a theorem of DL-PA. Therefore $\psi_0$ is a theorem of CL-PC, and so is $\varphi_0$. Let us now reformulate $\varphi_0$ with the modal operator of CL: we get $\langle\!\langle J\rangle\!\rangle\mathsf{X}(p\vee q) \leftrightarrow \langle\!\langle J\rangle\!\rangle\mathsf{X}p \vee \langle\!\langle J\rangle\!\rangle\mathsf{X}q$. This however fails to be a theorem of CL, as we will see in the next section (Sect. 4.1). We thereby complete Theorem 4.1 of [110]: the above $\varphi_0$ exemplifies that the other way round, there are unprovable CL formulas whose CL-PC counterpart is a theorem.

In the next section we are going to have a closer look at the logic of strategic ability ATL, of which CL is a fragment.

## 4 No uncertainty, strategic actions: ATL and STIT

In a strategic interpretation, the sentence "coalition $J$ is able to achieve $\varphi$" is understood as "$J$ is able to achieve $\varphi$ *whatever J's opponents do*". In the present section we discuss one of the main logics with a strategic interpretation of action modalities: Alternating-time Temporal Logic (ATL). We also briefly discuss Belnap's Seeing-To-It-That Logic (STIT), which is not only about what an agent can do, but also about what an agent *does*.

### 4.1 Alternating-time Temporal Logic (ATL)

Alternating-time Temporal Logic (ATL) was introduced as a logic for distributed systems [5]. Just as those of CL-PC, the modal operators of ATL relate actions to outcomes. They combine an action modality with three temporal modalities: the formula $\langle\!\langle J\rangle\!\rangle\mathsf{X}\varphi$ reads "the agents in $J$ have a strategy such that whatever the opponents do, next $\varphi$"; $\langle\!\langle J\rangle\!\rangle\mathsf{G}\varphi$ reads "the agents in $J$ have a strategy such that whatever the opponents do, henceforth $\varphi$"; and $\langle\!\langle J\rangle\!\rangle\varphi\mathsf{U}\psi$ reads "the agents in $J$ have a strategy such that whatever the opponents do, $\varphi$ until $\psi$". These are fused modal operators: the strategy quantifier $\langle\!\langle J\rangle\!\rangle$ is followed by one of the temporal operators $\mathsf{X}$, $\mathsf{G}$, or $\mathsf{U}$. The latter stem from Linear-time Temporal Logic (LTL). Similarly

**Table 5** Axiomatisation of ATL

| | |
|---|---|
| RE($\llbracket$J$\rrbracket$) | $\dfrac{\varphi \leftrightarrow \varphi'}{\llbracket J \rrbracket \mathsf{X}\varphi \rightarrow \llbracket J \rrbracket \mathsf{X}\varphi'}$ |
| M($\llbracket$J$\rrbracket$) | $\llbracket J \rrbracket \mathsf{X}(\varphi \wedge \psi) \rightarrow \big(\llbracket J \rrbracket \mathsf{X}\varphi \wedge \llbracket J \rrbracket \mathsf{X}\varphi\big)$ |
| N($\llbracket$J$\rrbracket$) | $\llbracket J \rrbracket \mathsf{X}\top$ |
| D($\llbracket$J$\rrbracket$) | $\neg\llbracket J \rrbracket \mathsf{X}\bot$ |
| Alt($\llbracket\mathbb{I}\rrbracket$) | $\neg\llbracket \emptyset \rrbracket \mathsf{X}\neg\varphi \rightarrow \llbracket \mathbb{I} \rrbracket \mathsf{X}\varphi$ |
| S($\llbracket$J$\rrbracket$) | $\Big(\llbracket J_1 \rrbracket \mathsf{X}\varphi_1 \wedge \llbracket J_2 \rrbracket \mathsf{X}\varphi_2\Big) \rightarrow \llbracket J_1 \cup J_2 \rrbracket \mathsf{X}(\varphi_1 \wedge \varphi_2) \quad$ if $J_1 \cap J_2 = \emptyset$ |
| FP($\mathsf{G}$) | $\llbracket J \rrbracket \mathsf{G}\varphi \leftrightarrow \varphi \wedge \llbracket J \rrbracket \mathsf{X}\llbracket J \rrbracket \mathsf{G}\varphi$ |
| GFP($\mathsf{G}$) | $\llbracket \emptyset \rrbracket \mathsf{G}\Big(\theta \rightarrow (\varphi \wedge \llbracket J \rrbracket \mathsf{X}\theta)\Big) \rightarrow \llbracket \emptyset \rrbracket \mathsf{G}\big(\theta \rightarrow \llbracket J \rrbracket \mathsf{G}\varphi\big)$ |
| FP($\mathsf{U}$) | $\llbracket J \rrbracket \psi\mathsf{U}\varphi \leftrightarrow \varphi \vee \big(\psi \wedge \llbracket J \rrbracket \mathsf{X}\llbracket J \rrbracket \psi\mathsf{U}\varphi\big)$ |
| LFP($\mathsf{U}$) | $\llbracket \emptyset \rrbracket \mathsf{G}\Big((\varphi \vee (\psi \wedge \llbracket J \rrbracket \mathsf{X}\theta)) \rightarrow \theta\Big) \rightarrow \llbracket \emptyset \rrbracket \mathsf{G}(\llbracket J \rrbracket \psi\mathsf{U}\varphi \rightarrow \theta)$ |

to LTL, the formula $\llbracket J \rrbracket \mathsf{F}\varphi$ can be introduced as an abbreviation of $\llbracket J \rrbracket \top\mathsf{U}\varphi$. It therefore reads "the agents in $J$ have a strategy such that whatever the other agents do, eventually $\varphi$".

Table 5 contains the axiomatisation of ATL [45].[4] The modal operators of ATL are not normal: for example, one cannot prove $(\llbracket J \rrbracket \mathsf{X}\varphi \vee \llbracket J \rrbracket \mathsf{X}\psi) \rightarrow \llbracket J \rrbracket \mathsf{X}(\varphi \vee \psi)$, which is the conjunction axiom schema C (cf. Sect. 3.1). ATL is a so-called monotonic modal logic [29]: the monotony schema M holds for $\llbracket J \rrbracket \mathsf{X}$ (that has the form of PDL's schema C, cf. Table 4). The failure of the C schema is due to the fact that $\llbracket J \rrbracket \mathsf{X}$ contains an existential quantification followed by a universal quantification. The superadditivity schema S($\llbracket$J$\rrbracket$) is central. It says that the abilities of disjoint coalitions $J_1$ and $J_2$ can be combined: if $J_1$ is able to achieve $\varphi_1$ and $J_2$ is able to achieve $\varphi_2$ then $J_1$ and $J_2$ together are able to achieve $\varphi_1 \wedge \varphi_2$. Just as PDL, ATL has a deduction theorem:

$$\chi \vdash_{\mathtt{ATL}} \varphi \text{ iff } \vdash_{\mathtt{ATL}} \llbracket \mathbb{I} \rrbracket \mathsf{G}\chi \rightarrow \varphi.$$

To decide the satisfiability of an ATL formula is an EXPTIME complete problem [120].

The language of Coalition Logic (CL) [88] is the fragment of the language of ATL without $\mathsf{G}$ and $\mathsf{U}$. It was proved in [44] that ATL is a conservative extension of CL. To decide the satisfiability of a CL formula is a PSPACE complete problem.

As we have already mentioned in Sect. 3.6, the nonstrategic CL-PC formula $\langle J \rangle [\overline{J}]\varphi$ of CL-PC is similar in spirit to the strategic formula $\llbracket J \rrbracket \mathsf{X}\varphi$. However, they do not have exactly the same meaning. Consider the CL equivalence $\chi_0 = \llbracket J \rrbracket \mathsf{X}(p \vee q) \leftrightarrow (\llbracket J \rrbracket \mathsf{X}p \vee \llbracket J \rrbracket \mathsf{X}q)$. We have seen in Sect. 3.6 that its CL-PC counterpart $\varphi_0$ is a theorem. However, if $p$ and $q$ are different then the left-to-right direction is neither provable in CL nor in ATL. To see this, suppose $\llbracket J \rrbracket \mathsf{X}(p \vee q) \leftrightarrow (\llbracket J \rrbracket \mathsf{X}p \vee \llbracket J \rrbracket \mathsf{X}q)$ was provable in CL. Then $\llbracket \emptyset \rrbracket \mathsf{X}(p \vee \neg p) \leftrightarrow (\llbracket \emptyset \rrbracket \mathsf{X}p \vee \llbracket \emptyset \rrbracket \mathsf{X}\neg p)$ would be provable for any variable $p$ by the rule of uniform substitution, which is derivable in CL. By Axiom N($\llbracket$J$\rrbracket$) and Rule RE($\llbracket$J$\rrbracket$) its left hand side $\llbracket \emptyset \rrbracket \mathsf{X}(p \vee \neg p)$ is a theorem and therefore $\llbracket \emptyset \rrbracket \mathsf{X}p \vee \llbracket \emptyset \rrbracket \mathsf{X}\neg p$ would be provable for any $p$: the empty coalition would be able to achieve either $p$ or $\neg p$, for every $p$. Worse, by the

[4] Actually we give an equivalent axiomatisation that better matches the other axiomatisations in this paper: our RE($\llbracket$J$\rrbracket$) and M($\llbracket$J$\rrbracket$) do not appear in the original presentation, while $\llbracket \emptyset \rrbracket \mathsf{G}\top$ and the inference rules $\dfrac{\varphi}{\llbracket \emptyset \rrbracket \mathsf{G}\varphi}$ and $\dfrac{\varphi \rightarrow \psi}{\llbracket \emptyset \rrbracket \mathsf{X}\varphi \rightarrow \llbracket \emptyset \rrbracket \mathsf{X}\psi}$ are missing in ours. The first can be proved from N($\llbracket$J$\rrbracket$) via RE($\llbracket$J$\rrbracket$) and the Greatest Fixpoint Axiom GFP($\mathsf{G}$). The second can be derived from the first via RE($\llbracket$J$\rrbracket$). The third can be derived from M($\llbracket$J$\rrbracket$) via RE($\llbracket$J$\rrbracket$).

monotony axiom $\mathtt{M}(\langle\!\langle \mathtt{J} \rangle\!\rangle)$ we would obtain the theorem $\big(\bigwedge_{J \subseteq \mathbb{I}} \langle\!\langle J \rangle\!\rangle \mathsf{X} p\big) \vee \big(\bigwedge_{J \subseteq \mathbb{I}} \langle\!\langle J \rangle\!\rangle \mathsf{X} \neg p\big)$: all coalitions would have exactly the same power. This would make CL an uninteresting logic of strategic ability. We conclude that While we agree with van der Hoek and Wooldridge that CL-PC is an interesting logic that captures the spirit of CL, the formulas $\varphi_0$ and $\chi_0$ highlight a mismatch between the two logics that can be traced back to the assumption underlying CL-PC that the agents' actions are modifications of the truth values of propositional variables. We note that $\varphi_0$ is not a schema, i.e., not every instance of it is a theorem. This poses the question whether CL-PC and CL differ by the status of some schema. We conjecture that no such schema exists.

The language of the logic ATL* is more liberal than that of ATL: fused modal operators such as $\langle\!\langle J \rangle\!\rangle \mathsf{X}$ are decomposed into the strategy quantifier $\langle\!\langle J \rangle\!\rangle$ and the temporal operator $\mathsf{X}$. The strategy quantifier can therefore be followed by arbitrary so-called *path formulas*, such as in $\langle\!\langle \{i\} \rangle\!\rangle (p \wedge \mathsf{X} p)$ or $\langle\!\langle \{i\} \rangle\!\rangle \mathsf{G}(p \wedge \neg \mathsf{G} \neg p)$. To decide the satisfiability of an ATL* formula is a 2EXPTIME complete problem [97]. The language of the logic of 'seeing-to-it-that' (STIT) is even more liberal: the strategy quantifier $\langle\!\langle J \rangle\!\rangle$ is decomposed into an existential quantification over historic possibilities that is followed by a universal quantification over the outcomes of $J$'s choices. We will introduce STIT in Sect. 4.3.

### 4.2 Reasoning about strategies in ATL: commitments and irrevocable strategies

One can describe an agent's action repertoire (and therefore his abilities) in ATL by stating the conditioned effects of each of the agent's actions. Such formulas take the form $\mathit{Cond} \rightarrow \langle\!\langle \{i\} \rangle\!\rangle \mathsf{X}\, \mathit{Effect}$. This supposes that actions take a single time step and are performed by a single agent. (More general temporal postconditions and statements of joint ability are however possible, too.) A set of such formulas makes up a theory of ability *AbilityTheory*. Given *AbilityTheory*, one can typically check whether given some description of the initial state *Init*, a coalition of agents $\{i_1, \ldots, i_n\}$ is able to achieve some goal *Goal* in the future, whatever the opponents do. Formally, we check whether

$$\mathit{AbilityTheory} \vdash_{\mathtt{ATL}} \mathit{Init} \rightarrow \langle\!\langle \{i_1, \ldots, i_n\} \rangle\!\rangle \mathsf{F}\, \mathit{Goal}$$

is the case.

While the modal operators of ATL provide an elegant and compact tool to talk about what coalitions are able to achieve against their opponents by means of strategies, strategies remain "unsung heroes" [108]: there are no names for strategies in the language and one cannot really 'grasp' them. Many authors felt that this is not satisfactory and extended ATL. Strategy Logic (SL) [28,85] has strategy variables and quantification over strategies; it is undecidable. The strategy quantifier of ATL *with explicit strategies* [121,123] is indexed by strategy names $\sigma$ committing some of the agents to follow some particular strategy. For example, the formula $\langle\!\langle \{\mathit{Ann}\} \rangle\!\rangle_{\mathit{Bob}:\sigma} \mathsf{G}\, \mathit{Married}$ says that if *Bob* is committed to follow the strategy $\sigma$ of staying faithful then *Ann* has a strategy ensuring that both stay married forever. Commitment to a strategy was later replaced by commitment to an action in ATL *with Explicit Actions* ATLEA [61]. Both ATL with explicit strategies and ATLEA have the same complexity as ATL. The aim of this line of work is to index the ATL strategy quantifier with complex PDL programs. It should in particular provide a dynamic operator with a strategic interpretation; such an operator is missing up to now. The logic ATL with intentions, ATLI, of [68] has a modal operator of commitment to a strategy as a first-class citizen: $(\mathit{str}_i \sigma_i)\, \varphi$ reads "if $i$ intends to play according to strategy $\sigma_i$ then $\varphi$ is true". This is further refined in [27] into a logic ATL with preferences, ATLP, where one can express that agents play some equilibrium.

Introducing commitments into ATL is a way of solving the problem of reasoning with uniform strategies which plagues the epistemic extension of ATL, ATEL. We will discuss this in Sect. 7.

Commitments can also be related to the problem of revocability of ATL strategies. Indeed, in ATL an agent $i$ may adopt some strategy and cancel ('revoke') it at a later stage. In other words, $i$ is never really committed to any strategy. This can be illustrated by the example of Ann and Bob who have in their repertoire the joint action of divorcing, as expressed by the ATL formula $\langle\!\langle \{Ann, Bob\} \rangle\!\rangle \mathsf{X} \neg Married$. Their repertoire also contains the action of staying married (that should perhaps better be called a non-action): $Married \rightarrow \langle\!\langle \{Ann, Bob\} \rangle\!\rangle \mathsf{X} Married$. When *AbilityTheory* contains these two formulas then we have

$$AbilityTheory \vdash_{\mathtt{ATL}} Married \rightarrow \langle\!\langle \{Ann, Bob\} \rangle\!\rangle \mathsf{G}\Big(Married \wedge \langle\!\langle \{Ann, Bob\} \rangle\!\rangle \mathsf{X} \neg Married\Big).$$

So according to ATL, once Ann and Bob are married they have a strategy that (1) makes them stay married forever, and (2) allows them to abandon that strategy and to divorce at any time point. In other words, when we look at the outcome of Ann and Bob's strategy to stay together forever then we never exclude that they deviate from that strategy. Several authors felt that this is an undesirable property. Beyond the above logics with names for strategies there are two approaches that avoid the introduction of names for strategies: ATL *with irrevocable strategies* [1] and ATL *with strategy contexts* [22]. The latter makes the adoption and canceling of strategies explicit. Its satisfiability problem was proved to be undecidable in [104].

### 4.3 Seeing-To-It-That Logic (STIT)

While PDL and ATL are about what agents can do, STIT [19,64] is both about what agents can do and about what they actually do. Such logics are called logics of agency in the philosophy of action literature. We here present the so-called Chellas STIT version.

The central device in the logical language are formulas of the form $\mathsf{Stit}_J \varphi$, read "by their *current choices* the agents in $J$ guarantee that $\varphi$ is true, whatever the other agents do". There is also an operator of historical possibility $\Diamond\varphi$, read "it is historically possible that $\varphi$". Its dual is defined as $\Box\varphi \stackrel{\text{def}}{=} \neg\Diamond\neg\varphi$ and is read "$\varphi$ is historically true" or "$\varphi$ is settled true". The language finally has the linear-time temporal operators $\mathsf{F}$ ('eventually') and $\mathsf{G}$ ('henceforth'). Several more recent publications also include the temporal 'next' operator $\mathsf{X}$ [24].

The operator of historical possibility can be used to quantify over the agents' choices: the formula $\Diamond\mathsf{Stit}_J\varphi$ says that $J$ is able to achieve $\varphi$ whatever the other agents do. So this construction resembles the ATL path quantifier $\langle\!\langle J \rangle\!\rangle$, and it is not very surprising that Coalition Logic CL becomes a fragment of Chellas STIT if we translate $\langle\!\langle J \rangle\!\rangle \mathsf{X}\varphi$ into $\Diamond\mathsf{Stit}_J\mathsf{X}\varphi$ [24]. An embedding of the entire language of ATL into Horty's strategic variant of STIT can be found in the same paper. It has to be noted that the operator of historic possibility is actually redundant if one admits empty coalitions: $\Box\varphi$ can then be viewed as an abbreviation of $\mathsf{Stit}_\emptyset\varphi$. So $\varphi$ is historically necessary if the empty coalition sees to it that $\varphi$.

The agency operators $\mathsf{Stit}_J$ are normal modal box operators satisfying the principles of S5. The operator of historic possibility is a normal modal diamond operator also satisfying the principles of S5. Historic necessity of $\varphi$ implies that every coalition sees to it that $\varphi$: the formula $\Box\varphi \rightarrow \mathsf{Stit}_J\varphi$ is a theorem of Chellas STIT. Just as in ATL, agents and disjoint coalitions of agents are supposed to be independent. This is axiomatised for individuals by a family of axiom schemas of independence of agents [19, Chapter 17] that resembles ATL's

superadditivity:

$$\mathtt{AIA_n}(\Diamond, \mathsf{Stit_i})(\Diamond\mathsf{Stit}_{i_1}\varphi_1 \wedge \cdots \wedge \Diamond\mathsf{Stit}_{i_n}\varphi_n) \to \Diamond(\mathsf{Stit}_{i_1}\varphi_1 \wedge \cdots \wedge \mathsf{Stit}_{i_n}\varphi_n)$$

A somewhat surprising consequence of this axiom is the theorem $\mathsf{Stit}_i\mathsf{Stit}_j\varphi \to \Box\varphi$, for $i \neq j$, in words: $i$ can make $j$ see to it that $\varphi$ only if $\varphi$ is settled true. It even turns out that a generalisation of that implication can replace the above axiom of independence of agents [11].

In STIT one can reason about responsibility and causality: one may check whether the coalition $J$ is responsible for some state of affairs *Fact* in context *Cond* by checking whether the formula

$$\mathit{Cond} \to \mathsf{Stit}_J\,\mathit{Fact}$$

is a STIT theorem. When this is the case then it is also said that $J$ is agentive for *Fact* under the circumstances *Cond*.

Several authors have argued that a better account of agency is provided by the so-called deliberative stit operator that can be defined from Chellas's stit operator by means of the following abbreviation:

$$\mathsf{DStit}_J\,\varphi \stackrel{\text{def}}{=} \mathsf{Stit}_J\varphi \wedge \Diamond\neg\varphi.$$

Their argument is that $\mathsf{Stit}_J\top$ is a theorem of Chellas STIT: every group is agentive for a tautology. This is not the case for the deliberative stit operator: on the contrary, $\neg(\mathsf{Stit}_J\top \wedge \Diamond\neg\top)$ is a theorem of Chellas STIT, i.e., no group can be agentive for a tautology.

There is a price to pay for generality: satisfiability of STIT formulas is undecidable already for the language without temporal operators [56]. It is decidable if coalitions are restricted to individuals, and deciding satisfiability is NEXPTIME complete [11].

## 5 Individual knowledge, no actions: S5

We now turn to something completely different: the logic of individual knowledge. We briefly recall the logic that is generally considered to be the standard logic of knowledge, viz. the modal logic S5, and show that it is too strong.

### 5.1 S5 as an epistemic logic

The logic of individual knowledge has modal operators $\mathsf{K}_i$, one per agent $i$ in $\mathbb{I}$. The formula $\mathsf{K}_i\varphi$ reads "agent $i$ knows that $\varphi$". Sometimes a dual modal operator $\hat{\mathsf{K}}_i$ is used: $\hat{\mathsf{K}}_i\varphi$ is equivalent to $\neg\mathsf{K}_i\neg\varphi$ and reads "$\varphi$ is consistent with $i$'s knowledge".

Starting with Moore [86] and Fagin et al. [35], most of the publications in computer science adopt the modal logic S5 as the logic of knowledge. We give its axiomatisation in Table 6.

As we have already mentioned in Sect. 3.1, the axiom schemas $\mathtt{M}(\mathsf{K_i})$ and $\mathtt{C}(\mathsf{K_i})$ can be replaced by the $\mathtt{K}(\mathsf{K_i})$-axiom $\big(\mathsf{K}_i\varphi \wedge \mathsf{K}_i(\varphi\to\psi)\big) \to \mathsf{K}_i\psi$. Each of the principles $\mathtt{RE}(\mathsf{K_i})$, $\mathtt{N}(\mathsf{K_i})$, and $\mathtt{K}(\mathsf{K_i})$ corresponds to an aspect of the *omniscience problem*: $\mathtt{RE}(\mathsf{K_i})$ says that an agent's knowledge is closed under logical equivalence; $\mathtt{N}(\mathsf{K_i})$ says that the agents know the tautologies; $\mathtt{K}(\mathsf{K_i})$ says that an agent's knowledge is closed under modus ponens. Axiom $\mathtt{T}(\mathsf{K_i})$ says that knowledge implies truth and distinguishes knowledge from belief. Axiom $\mathtt{4}(\mathsf{K_i})$ is the axiom of positive introspection and $\mathtt{5}(\mathsf{K_i})$ is the axiom of negative introspection.

**Table 6** Axiomatisation of the epistemic logic S5

| | |
|---|---|
| RE($\mathsf{K}_i$) | $\dfrac{\varphi \leftrightarrow \varphi'}{\mathsf{K}_i\varphi \leftrightarrow \mathsf{K}_i\varphi'}$ |
| N($\mathsf{K}_i$) | $\mathsf{K}_i\top$ |
| M($\mathsf{K}_i$) | $\mathsf{K}_i(\varphi \wedge \psi) \to (\mathsf{K}_i\varphi \wedge \mathsf{K}_i\psi)$ |
| C($\mathsf{K}_i$) | $(\mathsf{K}_i\varphi \wedge \mathsf{K}_i\psi) \to \mathsf{K}_i(\varphi \wedge \psi)$ |
| T($\mathsf{K}_i$) | $\mathsf{K}_i\varphi \to \varphi$ |
| 4($\mathsf{K}_i$) | $\mathsf{K}_i\varphi \to \mathsf{K}_i\mathsf{K}_i\varphi$ |
| 5($\mathsf{K}_i$) | $\neg\mathsf{K}_i\varphi \to \mathsf{K}_i\neg\mathsf{K}_i\varphi$ |

Up to today, S5 is considered to be 'the' logic of knowledge in theoretical computer science, artificial intelligence, and dynamic epistemic logics. However, in the formal epistemology literature one can find strong arguments that S5's negative introspection axiom makes it too strong a logic of knowledge. We go into the details in the next section.

### 5.2 Negative introspection is undesirable!

Among others, Lenzen and Voorbraak have provided formal arguments that Axiom 5($\mathsf{K}_i$) of negative introspection is not a desirable property of a knowledge operator [43,72,73,119]. Perhaps the most striking argument involves its interaction with a modality of strong belief $\mathsf{B}_i$, called conviction by Lenzen, where the formula $\mathsf{B}_i\varphi$ reads "agent $i$ (strongly) believes that $\varphi$". The only reasoning principles that we need are $\mathsf{K}_i\varphi \to \mathsf{B}_i\varphi$ (knowledge implies belief), and $\neg\mathsf{B}_i\bot$ (belief is consistent). The latter is the modal D($\mathsf{B}_i$) axiom and can be replaced by $(\mathsf{B}_i\varphi \wedge \mathsf{B}_i\neg\varphi) \to \bot$. Below we give a slightly modified version of the argument.

Suppose $\neg p \wedge \mathsf{B}_i\mathsf{K}_i p$ is the case: $i$ wrongly believes to know that $p$. Let us prove that such a hypothesis leads to inconsistency:

1. $\neg p \to \neg\mathsf{K}_i p$ (contraposition of Axiom T($\mathsf{K}_i$))
2. $\neg\mathsf{K}_i p \to \mathsf{K}_i\neg\mathsf{K}_i p$ (Axiom 5($\mathsf{K}_i$))
3. $\mathsf{K}_i\neg\mathsf{K}_i p \to \mathsf{B}_i\neg\mathsf{K}_i p$ ('knowledge implies belief')
4. $\neg p \to \mathsf{B}_i\neg\mathsf{K}_i p$ (from 1, 2, 3 by propositional logic)
5. $(\neg p \wedge \mathsf{B}_i\mathsf{K}_i p) \to (\mathsf{B}_i\mathsf{K}_i p \wedge \mathsf{B}_i\neg\mathsf{K}_i p)$ (from 4 by propositional logic)
6. $(\mathsf{B}_i\mathsf{K}_i p \wedge \mathsf{B}_i\neg\mathsf{K}_i p) \to \bot$ (Axiom D($\mathsf{B}_i$))
7. $(\neg p \wedge \mathsf{B}_i\mathsf{K}_i p) \to \bot$ (from 5, 6 by propositional logic)

The last line says that no agent can wrongly believe to know $p$. This is against our intuitions: one may believe to know that $p$, and even strongly so, but may nevertheless be wrong.

Lenzen proposed that Axiom 5($\mathsf{K}_i$) should be replaced by the weaker $\neg\mathsf{K}_i\neg\mathsf{K}_i\varphi \to \mathsf{K}_i\neg\mathsf{K}_i\neg\varphi$, which is perhaps better readable as $\hat{\mathsf{K}}_i\,\mathsf{K}_i\varphi \to \mathsf{K}_i\hat{\mathsf{K}}_i\,\varphi$ and is called Axiom .2($\mathsf{K}_i$) in the modal logic literature [72]. The resulting modal logic is called S4.2 and is Lenzen's 'official' logic of knowledge. Meyer and van der Hoek proposed the stronger logic S4.3 [83]. We will see in the next section that the move to S4.2 creates some new difficulties when we integrate dynamic operators.

## 6 Individual knowledge, nonstrategic actions: PAL

Starting with [86], many authors studied the interplay of knowledge and action. We here concentrate on the interplay of knowledge and *epistemic events*, which only change the agents' knowledge while leaving the facts unchanged. (The term 'epistemic action' became standard in the literature, however, 'event' is more appropriate because it is supposed that the

**Table 7** Axiomatisation of PAL

| | |
|---|---|
| EL | Some axiomatics of epistemic logic (e.g. S5 of Table 6) |
| RE($\langle\varphi!\rangle$) | $\dfrac{\psi \leftrightarrow \psi'}{\langle\varphi!\rangle\psi \leftrightarrow \langle\varphi!\rangle\psi'}$ |
| Dual($\langle\varphi!\rangle$) | $[\varphi!]\psi \leftrightarrow \neg\langle\varphi!\rangle\neg\psi$ |
| Red(p) | $\langle\varphi!\rangle p \leftrightarrow \varphi\wedge p$, for $p$ atomic |
| Red($\neg$) | $\langle\varphi!\rangle\neg\psi \leftrightarrow \varphi\wedge\neg\langle\varphi!\rangle\psi$ |
| Red($\wedge$) | $\langle\varphi!\rangle(\psi_1\wedge\psi_2) \leftrightarrow \langle\varphi!\rangle\psi_1\wedge\langle\varphi!\rangle\psi_2$ |
| Red($\mathsf{K}_\mathtt{i}$) | $\langle\varphi!\rangle\mathsf{K}_i\psi \leftrightarrow \varphi\wedge\mathsf{K}_i[\varphi!]\psi$ |

announcement is made either by the environment or by some agent that is not part of what is being modelled.) They behave just as expansion operations in Alchourrón, Gärdenfors and Makinson's AGM belief revision theory [3]. The resulting dynamic epistemic logics were intensely studied in the last decade. We here present the simplest dynamic epistemic logic: PAL.

### 6.1 Public Announcement Logic (PAL)

Public Announcement Logic (PAL) extends epistemic logic by dynamic operators $\langle\varphi!\rangle$ of public announcement of the truth of $\varphi$. The formula $\langle\varphi!\rangle\psi$ reads "the truthful public announcement of $\varphi$ can be made and $\psi$ will be true afterwards". The dual $[\varphi!]\psi$ is read "if the truthful public announcement of $\varphi$ is made then $\psi$ will be true afterwards". Announcements only change the agents' knowledge, but not the facts.

The axiomatics of PAL is in Table 7. It follows the presentation in [13]. Axiom Red(p) says that announcements leave the facts unchanged. The equivalences Red($\neg$), Red($\wedge$), and Red($\mathsf{K}_\mathtt{i}$) are reduction axioms: together with Rule RE($\langle\varphi!\rangle$) they allow to push dynamic operators inwards until they face atomic formulas. Then Red(p) allows to entirely eliminate the dynamic operators. Observe that the reduction axioms $[\varphi!]\mathsf{K}_i\psi \leftrightarrow \varphi\rightarrow\mathsf{K}_i[\varphi!]\psi$ and $[\varphi!]p \leftrightarrow \psi\rightarrow p$ can be respectively proved from Red($\mathsf{K}_\mathtt{i}$) and Red($\neg$) and from Red(p) and Red($\neg$).

Here is a much debated theorem of PAL: $[p\wedge\neg\mathsf{K}_i p!]\mathsf{K}_i p$. Indeed, by the reduction axioms it is successively equivalent to $(p\wedge\neg\mathsf{K}_i p) \rightarrow \mathsf{K}_i[p\wedge\neg\mathsf{K}_i p!]p$ and to $(p\wedge\neg\mathsf{K}_i p) \rightarrow \mathsf{K}_i\big((p\wedge\neg\mathsf{K}_i p)\rightarrow p\big)$, and the latter is a theorem of any normal modal logic because the right hand side of the implication is so. So after the announcement of $p\wedge\neg\mathsf{K}_i p$ it is true that $\mathsf{K}_i p$: the formula $p\wedge\neg\mathsf{K}_i p$ becomes false by the announcement. This so-called *Moore sentence* illustrates that even if what is announced is true before the announcement, it may become false by its mere announcement. Such formulas have been called unsuccessful. In contrast, non-epistemic formulas are always successful. More generally, if $\varphi$ is a formula without any negative occurrence of epistemic operators then $[\varphi!]\varphi$ is a theorem of PAL.

The complexity of satisfiability of PAL formulas is the same as that of the underlying epistemic logic EL [81]. It was also shown there that if EL equals K then PAL is more succinct than EL. This was generalised by French et al. to other epistemic logics, including S5 [38].

### 6.2 The problem of closure under updates in PAL

When it comes to choosing the epistemic logic underlying PAL, almost all authors either opt for S5 or for the basic modal logic K. This 'works' because both S5-based PAL and K-based PAL are conservative extensions of S5 and K: their theorems without the announcement operator coincide. Basically, the reason is that the axioms for the announcement operator do

not interfere with the axioms of the epistemic logic. This fails to hold for other epistemic logics: PAL extensions of many logics that are intermediate between K and S5 fail to be conservative extensions. This applies in particular to the logic of belief KD45 and to Lenzen's official logic of knowledge S4.2: their extensions by the reduction axioms of PAL of Table 7 turn out to augment the theorems of the underlying epistemic logic [13]. Let us demonstrate this by showing that in the PAL extension of the logic of belief KD45 one can prove the knowledge axiom $\mathtt{T}(\mathsf{K}_{\mathtt{i}})$.

1. $\bot \leftrightarrow \mathsf{K}_i\bot$ (KD45)
2. $\langle\neg\varphi!\rangle\bot \leftrightarrow \langle\neg\varphi!\rangle\mathsf{K}_i\bot$ (from (1) by $\mathtt{RE}(\langle\varphi!\rangle)$)
3. $\neg\langle\neg\varphi!\rangle\bot$ (by $\mathtt{Red}(\wedge)$, $\mathtt{Red}(\neg)$, $\mathtt{Red}(\mathtt{p})$)
4. $\neg\langle\neg\varphi!\rangle\mathsf{K}_i\bot$ (from (2), (3) by propositional logic)
5. $\neg\big(\neg\varphi \wedge \mathsf{K}_i[\neg\varphi!]\bot\big)$ (from (4) by $\mathtt{Red}(\mathsf{K}_{\mathtt{i}})$)
6. $[\neg\varphi!]\bot \leftrightarrow \varphi$ (by $\mathtt{Red}(\wedge)$, $\mathtt{Red}(\neg)$, $\mathtt{Red}(\mathtt{p})$)
7. $\neg\big(\neg\varphi \wedge \mathsf{K}_i\varphi\big)$ (from (5), (6) by substitution of equivalents)
8. $\mathsf{K}_i\varphi \rightarrow \varphi$ (from (7) by propositional logic)

This disqualifies KD45-based PAL as a logic of belief dynamics. It is currently not clear how this problem has to be addressed

### 6.3 Variants of PAL

Public Announcement Logic (PAL) turned out to be extremely fruitful. We briefly mention here some of its extensions in order to be able to refer to them in the sequel.

PAL being about epistemic actions only, it was extended by assignments in several papers [69,106,114,115,117]. It was shown in the latter two papers that the epistemic extension of Reiter's solution to the frame problem [95,96] can be captured in the extension of DL-PA by epistemic operators and public announcements.

While PAL announcements are public, starting with [42] several authors undertook to relax that hypothesis. Most prominently, dynamic epistemic logics [14] have *event models* which model that agents perceive events only incompletely.

Group Announcement Logic (GAL) is an extension of PAL by group announcements [2]. It actually blends PAL's announcements with an ATL-like ability operator involving a group of agents. The formula $\langle J\rangle\psi$ reads "$J$ can achieve $\psi$ by announcing some known formulas".

Arbitrary Public Announcement Logic (APAL) is an extension of PAL by arbitrary announcements [10]. Formulas involving the latter take the form $\langle!\rangle\psi$ and are read "there is a $\varphi$ such that $\langle\varphi!\rangle\psi$". APAL provides a formal framework for the analysis of the knowability paradox, also known as the Fitch paradox. The paradox—or rather, the problem—is that if we add to APAL the *prima facie* reasonable principle that all truths are knowable, formally $\varphi \rightarrow \langle!\rangle\mathsf{K}_i\varphi$, then $\varphi \rightarrow \mathsf{K}_i\varphi$ can be proved. Indeed, $(\varphi \wedge \neg\mathsf{K}_i\varphi) \rightarrow \langle!\rangle\mathsf{K}_i(\varphi \wedge \neg\mathsf{K}_i\varphi)$ is an instance of the knowability principle whose right hand side is inconsistent because $\mathsf{K}_i(\varphi \wedge \neg\mathsf{K}_i\varphi)$ is so in S5. In other words, while APAL is a conservative extension of PAL, the extension of APAL by the knowability principle fails to be a conservative extension of PAL and also of the underlying epistemic logic (unless the latter has $\varphi \rightarrow \mathsf{K}_k\varphi$ as a theorem). The satisfiability problem of APAL is undecidable [37]. In the next section we show that one cannot reason about uniform choices in APAL.

### 6.4 The problem of uniform choice in APAL

The problem of uniform choice was up to now only studied in logics with a strategic interpretation of actions. It however already occurs with nonstrategic action operators as

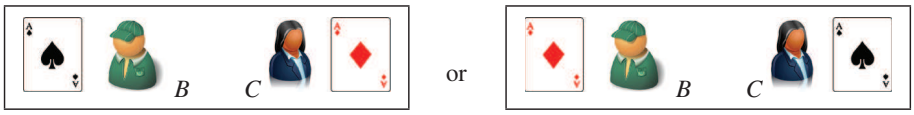


**Fig. 1** Two possible alternative situations for agent $Y$

soon as they quantify existentially over action choices. We call this the problem of uniform choice: the problem of how to express in the logical language that an agent knows how to achieve some $\varphi$. This implies that the agent also knows that there is a choice of his achieving $\varphi$, but not the other way round. Here is an example illustrating this in APAL.

Suppose you ($Y$) know that there are only two cards, the ace of spades and the ace of diamonds. Bob ($B$) and Cath ($C$) each hold one of these two. You don't see $B$'s and $C$'s cards, and they only see their own cards; so if Bob holds the ace of spades then he does not know that the other card is an ace of diamonds, and so on. There are two alternative situations for you: one where Bob holds spades and Cath holds diamonds, and one where Bob holds diamonds and Cath holds spades; they are depicted in Fig. 1.

Suppose you want $B$ to know $C$'s card, but not the other way round. Is there a public announcement doing the job?

We start by describing the initial situation and the goal situation:

$$\begin{aligned} Init &= \mathsf{K}_Y Spades \wedge \mathsf{K}_Y Diamonds \wedge \mathsf{K}_Y\big((\mathsf{K}_B Spades \wedge \neg\mathsf{K}_C Spades) \vee \\ &\qquad (\mathsf{K}_B Diamonds \wedge \neg\mathsf{K}_C Diamonds)\big) \\ Goal &= \mathsf{K}_B(Spades \wedge Diamonds) \wedge \neg\mathsf{K}_C(Spades \wedge Diamonds) \end{aligned}$$

where *Spades* means that one of agents $B$ and $C$ holds the spades and similarly for *Diamonds*. The following are theorems of PAL:

$$\begin{aligned} \vdash_{\mathtt{PAL}} (\mathsf{K}_B Spades \wedge \neg\mathsf{K}_C Spades) &\rightarrow \langle Spades{\rightarrow}Diamonds!\rangle Goal \\ \vdash_{\mathtt{PAL}} (\mathsf{K}_B Diamonds \wedge \neg\mathsf{K}_C Diamonds) &\rightarrow \langle Diamonds{\rightarrow}Spades!\rangle Goal \end{aligned}$$

So if $B$ holds the ace of spades then the announcement $Spades{\rightarrow}Diamonds!$ does the job; and if $B$ holds the ace of diamonds then the announcement $Diamonds{\rightarrow}Spades!$ does the job. It follows that $\vdash_{\mathtt{APAL}} Init \rightarrow \mathsf{K}_Y\langle\exists!\rangle Goal$: given the initial situation *Init*, you know that *there is* an announcement achieving the goal. However, you don't know which announcement you should make. Our example is a typical situation where you know that there is a strategy, but you don't know which action to choose. There seems to be no way to express in APAL that there is no uniform choice guaranteeing the goal.

The problem can be solved in GAL, where one can write both $\mathsf{K}_Y\langle\{Y\}\rangle\varphi$ and $\langle\{Y\}\rangle\mathsf{K}_Y\varphi$. The former says that you know that you have a (possibly non-uniform) choice achieving $\varphi$, while the latter says that you have a (necessarily uniform) choice of which you know that it achieves $\varphi$.

## 7 Individual knowledge, strategic actions: ATEL

We now briefly turn towards a paradigmatic logic of knowledge and strategic action: Alternating-time Temporal Epistemic Logic (ATEL) [109]. ATEL combines ATL with epistemic logic. It has formulas of the form $\langle\!\langle J\rangle\!\rangle\psi$ where $\psi$ is a formula governed by a temporal

operator just as in ATL (cf. Sect. 4.1), read "coalition $J$ can achieve $\psi$ (whatever its opponents do)", and $\mathsf{K}_i\varphi$, read "agent $i \in \mathbb{I}$ knows that $\varphi$".

It was pointed out by Schobbens that in ATEL one cannot represent uniform strategies [98]. The problem is basically the same as APAL's problem of uniform choice that we have exposed in Sect. 6.4, viz. to design a logical language allowing to express that an agent knows a strategy to enforce some property, which is stronger a statement than that the agent knows that *there is* a strategy; in the latter case he does not necessarily know what strategy to select. Indeed, it is irrelevant whether there is a quantification over the opponents' actions or not: the problem rather comes from the existential quantification occurring in the modal operators of ability. It may be exemplified in the language of ATEL by the formula $\mathsf{K}_i\langle\!\langle i\rangle\!\rangle\mathsf{X}\mathit{SafeOpen}$ which expresses that agent $i$ knows that there is an action of his opening the safe: the formula may be true without agent $i$ knowing the combination of the safe.

Following Schobbens's paper, several amendments of ATEL where proposed in the literature in order to allow to express uniformity of strategies. It seems fair to say that the accounts they propose are fairly complex: Constructive Strategic Logic [66] evaluates formulas not at state but at a set of states, leading to a non-standard negation operator; Alternating-time Temporal Observation Logic (ATOL) [67] has formulas with indexed path quantifiers $\langle\!\langle J\rangle\!\rangle_{Obs(i)}\psi$ "group $J$ has a strategy to enforce $\psi$, and agent $i$ can see the strategy".

Perhaps the most promising solutions are provided by extensions of ATL framework with strategic contexts. In the language of ATEL with Explicit Actions ATELEA [61], which is the epistemic extension of logic ATLEA of Sect. 4.2, one might write for example $\mathsf{K}_i\langle\!\langle i\rangle\!\rangle_{i:dial_{1234}}\mathsf{X}SafeOpen$ in order to express that $i$ knows the combination of the safe. A further step in this direction is Epistemic Strategy Logic ESL [18] where Strategy Logic SL is extended with epistemic operators.

It has also been proposed to solve the problem of uniform strategies by moving to an epistemic extension of the logic of agency STIT [23,25,57]. The language of that logic has four kinds of modal operators: the epistemic operator $\mathsf{K}_i$, the historic possibility operator $\Diamond$, the agency operator $\mathsf{Stit}_i$, and temporal operators $\mathsf{X}$ etc. As we have said in Sect. 4.3, the sequence $\Diamond\mathsf{Stit}_i\mathsf{X}$ captures the ATL operator $\langle\!\langle i\rangle\!\rangle\mathsf{X}$. The difference is that $\langle\!\langle i\rangle\!\rangle\mathsf{X}$ is a single, fused operator while in STIT it is decomposed into its three constituents. This provides a greater linguistic flexibility because the epistemic operator $\mathsf{K}_i$ can be inserted anywhere in the sequence. First, the formula $\mathsf{K}_i\Diamond\mathsf{Stit}_i\mathsf{X}\varphi$ is what can be said in ATEL about agent $i$'s *de dicto* knowledge, while the formula $\Diamond\mathsf{K}_i\mathsf{Stit}_i\mathsf{X}\varphi$ is about $i$'s *de re* knowledge; it cannot be formulated in ATEL. In our example, the conjunction $\mathsf{K}_i\Diamond\mathsf{Stit}_i\mathsf{X}\,\mathit{SafeOpen}\wedge\neg\Diamond\mathsf{K}_i\mathsf{Stit}_i\mathsf{X}\,\mathit{SafeOpen}$ adequately expresses that $i$ knows that there is an action opening the safe, but does not know which action opens it: there no choice of an action for which $i$ knows that $i$ sees to it that the safe is open next. In the case of multiple agents the picture can be further refined by distinguishing $i$'s *ex ante* knowledge—what $i$ knows before choosing his action—, $i$'s intermediate knowledge—what $i$ knows given that he has chosen his action but does not know the other agents' choices—, and $i$'s *ex post* knowledge—what $i$ knows once all the agents' choices are known to him [80].

## 8 Group knowledge, no actions: $\mathsf{S5}^{\mathsf{CK}}$

The concept of common knowledge is central in several disciplines. In the social sciences it is the basis of conventions [74]. In semantics it provides the notion of common ground in

**Table 8** Axiomatisation of epistemic logics with common knowledge $\mathtt{EL}^{\mathsf{CK}}$

| | |
|---|---|
| $\mathtt{EL}$ | Some axiomatics of epistemic logic |
| $\mathtt{FP}(\mathsf{CK}_{\mathtt{J}})$ | $\mathsf{EK}_J\mathsf{CK}_J\varphi \leftrightarrow \mathsf{EK}_J(\varphi \wedge \mathsf{CK}_J\varphi)$ |
| $\mathtt{GFP}(\mathsf{CK}_{\mathtt{J}})$ | $\big(\mathsf{EK}_J\varphi \wedge \mathsf{CK}_J(\varphi \to \mathsf{EK}_J\varphi)\big) \to \mathsf{CK}_J\varphi$ |

conversation [30]. In game theory, the hypothesis of common knowledge of rationality was proved to be fundamental for equilibria [9]. In theoretical computer science it is used to study communication in distributed systems [35].

The language has common knowledge operators $\mathsf{CK}_J$, one per group $J \subseteq \mathbb{I}$. The formula $\mathsf{CK}_J\varphi$ is read "it is common knowledge in group $J$ that $\varphi$".

As we have said in Sect. 2.1, common knowledge of $\varphi$, $\mathsf{CK}_J\varphi$, is typically explained in terms of the modal operator of shared knowledge $\mathsf{EK}_J$ as being the infinite conjunction $\mathsf{EK}_J\varphi \wedge \mathsf{EK}_J\mathsf{EK}_J\varphi \wedge \mathsf{EK}_J\mathsf{EK}_J\mathsf{EK}_J\varphi \wedge \cdots$. However, due to this infiniteness $\mathsf{CK}_J$ cannot be syntactically reduced to $\mathsf{EK}_J$ and requires a proper axiomatisation. Most of the axiomatisations in the literature are built on top of the epistemic logic $\mathtt{S5}$. Given that we have stressed the inappropriateness of the latter as a logic of knowledge in Sect. 5.2, the axiomatics we give in Table 8 is based on a non-specified normal modal logic of individual knowledge $\mathtt{EL}$ that may be any modal logic between $\mathtt{K}$ and $\mathtt{S5}$. It is obtained by adding two axioms to the logic of individual knowledge $\mathtt{EL}$: the Fixpoint Axiom $\mathtt{FP}(\mathsf{CK}_{\mathtt{J}})$ and the Greatest Fixpoint Axiom $\mathtt{GFP}(\mathsf{CK}_{\mathtt{J}})$. Both involve shared knowledge $\mathsf{EK}_J\varphi$, which as we have mentioned in Sect. 2.1 abbreviates $\bigwedge_{i\in J} \mathsf{K}_i\varphi$. Axiom $\mathtt{GFP}(\mathsf{CK}_{\mathtt{J}})$ can be replaced by the following inference rule:

It can be proved that $\mathsf{CK}_J$ is a normal modal operator. We observe that the formula

$$\mathsf{CK}_{\{1,2\}}\Big((\varphi_1 \leftrightarrow \varphi_2) \wedge \bigwedge_{i\in\{1,2\}} (\varphi_i \to \mathsf{K}_i\varphi_i)\Big) \to \mathsf{CK}_{\{1,2\}}\Big((\varphi_1 \wedge \varphi_2) \to \mathsf{EK}_{\{1,2\}}(\varphi_1 \wedge \varphi_2)\Big)$$

is a theorem of $\mathtt{EL}^{\mathsf{CK}}$ if $\mathtt{EL}$ is a normal modal logic. It says that if it is common knowledge that $\varphi_1$ and $\varphi_2$ have the same truth value and that each agent $i$ knows the truth value of 'his' $\varphi_i$ then it is common knowledge that if both are true then everybody knows this. Here is a proof.

1. $\Big(\bigwedge_{i\in\{1,2\}} (\varphi_i \to \mathsf{K}_i\varphi_i)\Big) \to \Big((\varphi_1 \wedge \varphi_2) \to \mathsf{EK}_{\{1,2\}}(\varphi_1 \vee \varphi_2)\Big)$ (by $\mathtt{EL}$)
2. $\Big(\mathsf{CK}_{\{1,2\}} \bigwedge_{i\in\{1,2\}} (\varphi_i \to \mathsf{K}_i\varphi_i)\Big) \to \mathsf{CK}_{\{1,2\}}\Big((\varphi_1 \wedge \varphi_2) \to \mathsf{EK}_{\{1,2\}}(\varphi_1 \vee \varphi_2)\Big)$
   (from (1) with $\mathsf{CK}_J$ normal modal operator)
3. $\mathsf{CK}_{\{1,2\}}(\varphi_1 \leftrightarrow \varphi_2) \to \mathsf{CK}_{\{1,2\}}\mathsf{EK}_{\{1,2\}}\big((\varphi_1 \vee \varphi_2) \to (\varphi_1 \wedge \varphi_2)\big)$
   ($\mathtt{FP}(\mathsf{CK}_{\mathtt{J}})$ with $\mathsf{CK}_J$ normal modal operator)
4. $\mathsf{CK}_{\{1,2\}}\Big((\varphi_1 \leftrightarrow \varphi_2) \wedge \bigwedge_{i\in\{1,2\}} (\varphi_i \to \mathsf{K}_i\varphi_i)\Big) \to \mathsf{CK}_{\{1,2\}}\Big((\varphi_1 \wedge \varphi_2) \to \mathsf{EK}_{\{1,2\}}(\varphi_1 \wedge \varphi_2)\Big)$
   (from (2) and (3) with $\mathsf{CK}_J$ normal modal operator)

We are going to use that theorem in the next section where we are going to discuss a property of Axiom $\mathtt{GFP}(\mathsf{CK}_{\mathtt{J}})$ that we find debatable.

## 9 Group knowledge, nonstrategic actions: $\mathtt{PAL}^{\mathsf{CK}}$

We now study the extension of PAL by the common knowledge operator as a prototypical logic involving both group knowledge and nonstrategic actions. The dynamic dimension enables us to discuss what we believe to be a surprising and debatable consequence of the Greatest Fixpoint Axiom $\mathtt{GFP}(\mathsf{CK}_{\mathbb{J}})$ of Table 8.

### 9.1 Public announcement logic with common knowledge $\mathtt{PAL}^{\mathsf{CK}}$

Public Announcement Logic with Common Knowledge $\mathtt{PAL}^{\mathsf{CK}}$ is an extension of the logic of common knowledge by public announcements. It may as well be viewed as an extension of PAL by common knowledge. An axiomatisation of $\mathtt{PAL}^{\mathsf{CK}}$ can be found in [14, Fig. 9]. It extends the axiomatics of PAL of Table 7 by the following rule:

$$\frac{\chi \to \langle\varphi!\rangle\psi \qquad (\chi \wedge \varphi) \to \mathsf{EK}_{\mathbb{I}}\psi}{\chi \to \langle\varphi!\rangle\mathsf{CK}_{\mathbb{I}}\psi}$$

The common knowledge version $\langle\psi!\rangle\mathsf{CK}_J\varphi \leftrightarrow \psi\wedge\mathsf{CK}_J[\psi!]\varphi$ of the reduction axiom $\mathtt{Red}(\mathsf{K}_{\mathtt{i}})$ cannot be proved in $\mathtt{PAL}^{\mathsf{CK}}$: while the right-to-left direction is a theorem, the left-to-right direction is not. So common knowledge may 'pop up' as a consequence of a public announcement in a way that is unforeseeable for the group $J$. In contrast, the reduction axiom $\mathtt{Red}(\mathsf{K}_{\mathtt{i}})$ says that an agent can foresee what he will know after the announcement of $\psi$, namely what he currently knows to be true after the announcement of $\psi$. From a more general perspective, this means that the logical properties of the common knowledge operator differ from that of the individual knowledge operator, while these two operators obey exactly the same principles in the underlying epistemic logic: when the principles for $\mathsf{K}_i$ are those of S5 then the principles for $\mathsf{CK}_J$ are those of S5, too. Beyond such formal considerations, the sequel provides an example casting some doubt on the suitability of $\mathtt{PAL}^{\mathsf{CK}}$ as a logic of common knowledge and public announcements.

### 9.2 The ignorant compatriots example

We take over an example from [77]. Suppose agents $B$ and $C$ are both Italian, expressed by the conjunction $it_B \wedge it_C$, and don't know each other. They accidentally meet in the coffee break of some conference and start to talk in English. It is commonsense knowledge—and by that also common knowledge of $A$ and $B$—that each of them knows whether he/she is Italian or not. This is described by the following formula:

$$Init \;=\; it_B \wedge it_C \wedge \bigwedge_{i\in\{B,C\}} \mathsf{CK}_{\{B,C\}}\big((it_i \to \mathsf{K}_i\, it_i) \wedge (\neg it_i \to \mathsf{K}_i \neg it_i)\big).$$

However, $B$ does not know that $C$ is Italian and $C$ does not know that $B$ is Italian.

We consider two continuations of our story.

In the first scenario, a third agent who happens to know both of them overhears their conversation and truthfully declares: "Hey, you are both Italian!" By the announcement it clearly becomes common knowledge that both are Italian: we have

$$\vdash_{\mathtt{PAL}^{\mathsf{CK}}} Init \to \big\langle it_B \wedge it_C!\big\rangle\, \mathsf{CK}_{\{B,C\}}\big(it_B \wedge it_C\big).$$

In the second scenario, the third agent truthfully says: "Hey, you are compatriots!" Surprisingly, the principles of $\mathtt{PAL}^{\mathsf{CK}}$ tell us that $it_B \wedge it_C$ immediately becomes common knowledge, too:

$$\vdash_{\mathtt{PAL}^{\mathsf{CK}}} Init \to \big\langle it_B \leftrightarrow it_C!\big\rangle\, \mathsf{CK}_{\{B,C\}}(it_B \wedge it_C).$$

To see this, remember from Sect. 8 that $it_B \leftrightarrow it_C$ being boolean, we have

$$\vdash_{\mathtt{PAL}^{\mathsf{CK}}} \big[it_B \leftrightarrow it_C!\big]\, \mathsf{CK}_{\{B,C\}}(it_B \leftrightarrow it_C),$$

and that we have moreover

$$\vdash_{\mathtt{EL}^{\mathsf{CK}}} \mathsf{CK}_{\{B,C\}}\Big(\big(it_B \leftrightarrow it_C\big) \wedge \bigwedge_{i \in \{B,C\}} \big(it_i \rightarrow \mathsf{K}_i it_i\big)\Big) \rightarrow \mathsf{CK}_{\{B,C\}} \Big(\big(it_B \wedge it_C\big) \rightarrow \mathsf{EK}_{B,C}\big(it_B \wedge it_C\big)\Big).$$

Together with the instance

$$\vdash_{\mathtt{PAL}^{\mathsf{CK}}} \Big(\big(it_B \wedge it_C\big) \wedge \mathsf{CK}_{\{B,C\}}\Big(\big(it_B \wedge it_C\big) \rightarrow \mathsf{EK}_{\{B,C\}}(it_B \wedge it_C)\Big)\Big) \rightarrow \mathsf{CK}_{\{B,C\}}\big(it_B \wedge it_C\big)$$

of the Greatest Fixpoint Axiom $\mathtt{GFP}(\mathsf{CK}_{\mathrm{J}})$ it follows that

$$\vdash_{\mathtt{PAL}^{\mathsf{CK}}} \mathit{Init} \rightarrow \mathsf{CK}_{\{B,C\}}\big(it_B \wedge it_C\big).$$

This inference is however at least questionable: in real life conversations, agent $B$ is able to infer from his knowledge that $\mathsf{CK}_J\big(it_C \rightarrow \mathsf{K}_C it_C\big)$ and the truthful public announcement $it_B \leftrightarrow it_C!$ that $it_B$, $\mathsf{K}_C it_B$, $\mathsf{K}_C\mathsf{K}_B it_B$, $\mathsf{K}_C\mathsf{K}_B\mathsf{K}_C it_B$, etc. are all true, up to any level of nesting; however, it seems that the announcement of $it_B \leftrightarrow it_C$ never makes $it_B \wedge it_C$ part of the common ground of the conversation.[5] One may argue that human agents may fail to infer all logical consequences in such a situation, just as they fail to infer the right conclusion in the case of other logical fallacies. However, we find it puzzling that the formal model of common knowledge does not match human reasoning in a simple situation with two agents and two propositions at stake.

It was proposed in [77] that one should distinguish two forms of common knowledge: implicit common knowledge and explicit common knowledge, with modal operators $\mathsf{CK}_J$ and $\mathsf{ECK}_J$. Implicit common knowledge is $\mathtt{PAL}^{\mathsf{CK}}$ common knowledge: it has the least fixpoint axiom and there is no reduction axiom for it. In contrast, explicit common knowledge $\mathsf{ECK}_J$ does not have the least fixpoint axiom (it is *some* fixpoint, but not necessarily the least) and has the following reduction axiom:

$$[\psi!]\mathsf{ECK}_J\varphi \leftrightarrow \neg\psi \vee \mathsf{ECK}_J[\psi!]\varphi.$$

Let $\mathtt{PAL}^{\mathsf{CK},\mathsf{ECK}}$ be the resulting logic (for some underlying normal modal logic of individual knowledge $\mathtt{EL}$). In the second scenario of our compatriots example we get the following two theorems:

$$\vdash_{\mathtt{PAL}^{\mathsf{CK},\mathsf{ECK}}} \mathit{Init} \rightarrow \langle it_B \wedge it_C!\rangle\Big(\mathsf{CK}_{\{B,C\}}(it_B \wedge it_C) \wedge \mathsf{ECK}_{\{B,C\}}(it_B \wedge it_C)\Big),$$

$$\vdash_{\mathtt{PAL}^{\mathsf{CK},\mathsf{ECK}}} \mathit{Init} \rightarrow \langle it_B \leftrightarrow it_C!\rangle\Big(\mathsf{CK}_{\{B,C\}}(it_B \wedge it_C) \wedge \neg\mathsf{ECK}_{\{B,C\}}(it_B \wedge it_C)\Big).$$

So the two operators lead to the same outcome in the first scenario, while we obtain two different outcomes in the second scenario, depending on the kind of common knowledge under consideration. It therefore seems that the modal operator $\mathsf{ECK}_J$ might be more appropriate for example for reasoning about the common ground in conversation.

We contend that the above criticism is not as stringent as the others we have put forward up to now, which had a more formal nature. However, we believe that the standard account of common knowledge in terms of the least fixpoint axiom deserves to be questioned. Actually

[5] Thanks are due to Nicholas Asher for a confirmation of that point in conversation.

the issue brought up by our second scenario can be related to the omniscience problem: it illustrates that the inference of common knowledge that is delivered by the least fixpoint axiom is by no means guaranteed if agents have limited reasoning capabilities. The least fixpoint axiom has therefore to be considered as a further omniscience principle, to be added to those that we have listed for the individual knowledge operator in Sect. 5.1. The systematic failure of humans to infer common knowledge in our simple scenario indicates that it should actually be the first epistemic principle to be questioned when investigating the omniscience problem.

### 9.3 A remark on distributed knowledge

The extension of PAL by distributed knowledge has not been studied much, with the exception of [50,122]. We do not discuss the issue here and refer the reader to the above papers.

This is however the right moment to point out the shortcoming of our informal definition of distributed knowledge as 'knowledge achievable by communication'. This works for boolean formulas as well as for positive epistemic formulas: formulas without action operators and without epistemic operators in the scope of negation. It is inappropriate for Moore sentences such as $p \wedge \neg \mathsf{K}_i\, p$. To see this consider the extension of Arbitrary Public Announcement Logic APAL of Sect. 6.3 by the shared knowledge operator and define "$\varphi$ is distributed knowledge in the group of all agents $\mathbb{I}$" as the formula $\langle ! \rangle \mathsf{EK}_{\mathbb{I}} \varphi$. Suppose $\mathbb{I}$ contains at least two agents $i$ and $j$, and suppose $j$ knows that $p \wedge \neg \mathsf{K}_i\, p$. As individual knowledge implies distributed knowledge (cf. Sect. 2.1), $p \wedge \neg \mathsf{K}_i\, p$ must be distributed knowledge in $\mathbb{I}$. However, $p \wedge \neg \mathsf{K}_i\, p$ cannot become knowledge of $i$, i.e., $\mathsf{K}_i\,(p \wedge \neg \mathsf{K}_i\, p)$ is inconsistent in S5 (it is actually already so in the weaker logic KD4). It follows that $p \wedge \neg \mathsf{K}_i\, p$ cannot become shared knowledge of any group involving $i$ and, *a fortiori*, not of $\mathbb{I}$: the formula $\mathsf{EK}_{\mathbb{I}}(p \wedge \neg \mathsf{K}_i\, p)$ is inconsistent in $\mathtt{S5}^{\mathsf{CK}}$.

Similarly, in our scenario the formula $p \wedge \neg \mathsf{K}_i\, p$ cannot become common knowledge of $\mathbb{I}$. So a more demanding definition of distributed knowledge as achievable common knowledge leads to a problem, too. The problem actually already shows up for a less demanding scenario where $j$ knows that $p \wedge \neg \mathsf{CK}_{\mathbb{I}}\, p$.

This calls for further investigation of the interplay between announcements and distributed knowledge.

## 10 Group knowledge, strategic actions: $\mathtt{ATEL}^{\mathsf{CK}}$

The *grande finale* of our presentation is the class of logics having both group knowledge and strategic actions. We focus on the extension of the logic ATEL of Sect. 7 by common knowledge, appropriately baptised $\mathtt{ATEL}^{\mathsf{CK}}$.

The logic $\mathtt{ATEL}^{\mathsf{CK}}$ inherits all the problems of its fragments: as to the epistemic dimension, we have seen that we get in trouble if we choose S5 as the logic of knowledge and should choose S4.2 instead (cf. Sect. 5.2), which in turn leads to technical problems (cf. Sect. 6.2) and we have seen that the Greatest Fixpoint Axiom $\mathtt{GFP}(\mathsf{CK}_{\mathtt{J}})$ might be too strong for common knowledge (cf. Sect. 9.2); as to the dynamic dimension, we have seen that ATL strategies lack commitments (cf. Sect. 4.2) and that ATEL does not enable us to reason about uniform choices and uniform strategies (cf. Sect. 7).

We do not want to discuss these problems again here. Instead, we briefly mention a further difficulty: when we want to say that a group $J$ knows how to achieve some goal $\varphi$, i.e., that $J$ knows that $J$ has a strategy to achieve $\varphi$, then it is not clear which kind of group knowledge is

required. Sometimes distributed knowledge $\mathsf{DK}_J\langle\!\langle J\rangle\!\rangle\mathsf{F}\varphi$ is enough; sometimes shared knowledge $\mathsf{EK}_J\langle\!\langle J\rangle\!\rangle\mathsf{F}\varphi$ is enough; and sometimes common knowledge $\mathsf{CK}_J\langle\!\langle J\rangle\!\rangle\mathsf{F}\varphi$ is required. For example, the Byzantine agreement problem where communication is asynchronous and where messages may get lost requires common knowledge of a uniform strategy. In contrast, when public announcements can be made then distributed knowledge is enough. We believe that this is an interesting topic for further investigation.

## 11 Summary and research challenges

We have categorised logics for MAS in terms of the knowledge dimension and the action dimension, and for each category we have revisited and criticised its most prominent logics. As to logics with nonstrategic action operators, we have advocated DL-PA as an interesting logic that solves the frame problem and that embeds logics of nonstrategic ability. We believe it to be an appropriate basis for formal approaches to knowledge representation in general and to MAS in particular. Beyond the embedding of CL-PC that we have recalled in Sect. 3.5, the central role of DL-PA is confirmed by DL-PA accounts of update and revision operations [54], belief merging operations [63], abstract argumentation theories and their modification [33], and answer-set programs and their modification [36]; moreover, its epistemic extension, public announcement logic with assignments, solves the epistemic frame problem [114,115]. As to logics with epistemic operators, we have recalled that S5 is inadequate as a logic of knowledge and should be replaced by S4.2; furthermore, we have shown that the definition of common knowledge by means of the Greatest Fixpoint Axiom GFP($\mathsf{CK}_J$) is at least questionable. Finally, as to logics with strategic action operators, we have shown that ATL has to be augmented by some notion of commitment (to an action or to a strategy) and that just as arbitrary public announcement logic (APAL), its epistemic extension is too weak to cover reasoning about uniform strategies.

Let us terminate by sketching and discussing some challenges for future research.

### 11.1 From logics of knowledge to logics of belief (including belief revision)

In Sect. 5.2 we have criticised the choice of S5 as the logic of knowledge. One may however hope that negative introspection is justifiable at least in some contexts. For example, the interpreted systems from which knowledge is defined in [35] naturally validate that axiom. It would be interesting to investigate under which circumstances the axiom is a reasonable assumption. As far as we know this has not been investigated up to now.

The epistemic concepts of individual knowledge and group knowledge have doxastic counterparts, viz. the concepts of individual belief and group belief. Belief is actually more primitive a concept than knowledge, and many have tried to define the latter from the former. However, there is no consensus about their precise relationship apart from simple principles such as 'knowledge implies belief' that we have mentioned in Sect. 6.

The move towards beliefs comes with some new problems. The first difficulty is that while knowledge is binary—either you know or you don't—, belief typically comes in degrees. How this should be modeled and how degrees relate to (subjective) probability theory is an old and much debated subject of research. The second and main difficulty is that contrarily to knowledge, beliefs sometimes have to be revised: while knowledge is always true (because knowledge implies truth), an agent's beliefs may be false. In the context of PAL, an agent may wrongly believe that some announcement $\psi$ cannot be made because he wrongly believes $\psi$ to be false. In such circumstances only the left-to-right direction $\langle\psi!\rangle\mathsf{B}_i\varphi \rightarrow (\psi \wedge \mathsf{B}_i[\psi!]\varphi)$

of the (belief version of the) reduction axiom $\mathtt{Red}(\mathsf{K}_i)$ of Table 7 holds, and trivially so, while the right-to-left direction $(\psi \wedge \mathsf{B}_i[\psi!]\varphi) \rightarrow \langle\psi!\rangle\mathsf{B}_i\varphi$ does not. A satisfactory doxastic version of PAL clearly has to integrate some notion of belief revision. How this should be done is still very much an open issue, despite numerous recent efforts [6,7,102,107,113]. In our view, a further problem that has not been addressed up to now in a satisfactory manner is that almost all belief revision theories presuppose some kind of preference information: in many cases it is not clear at all where this information comes from. In other words, the challenge is to find simpler, more modest theories of revision that can be easily and smoothly integrated into logics of belief and action. First steps in that direction can be found in recent papers, but much remains to be done [15–17].

Past artificial intelligence and MAS research on logics of knowledge and belief focussed a lot on the omniscience problem that we have mentioned in Sect. 5.1, probably because it was believed to be central in the implementation of BDI agents. We think the problem might have been overestimated a bit, while problems related to belief update and the analysis of plans were rather neglected in the implementation of BDI agents; see [4] for a similar point of view. Moreover, limited reasoning has often to be disregarded when we want to reason about opponents, for example when we model an attacker in a cryptographic protocol. Let us finally note that limited reasoning about common knowledge was not considered up to now in the literature on the omniscience problem. Our 'compatriots' example against $\mathtt{PAL}^{\mathsf{CK}}$ might be viewed as evidence that failure of application of the Greatest Fixpoint Axiom $\mathtt{GFP}(\mathsf{CK}_{\mathbb{J}})$ should be the first thing to model in a satisfactory account of the multiagent omniscience problem.

### 11.2 Group belief and acceptance

Beyond individual belief, the investigation of different kinds of group belief is a highly interesting field of research. Examples are group acceptance and we-belief [105]. Acceptance is contextual: a lawyer accepts that his client is innocent in the context of the court, while he may privately believe that the client is guilty. As our law court example illustrates, neither does acceptance imply belief nor the other way round. In contrast, belief and we-belief are both non-contextual. We-belief differs from the standard notions of shared and common belief (whose knowledge counterparts were presented in Sects. 2.1 and 8) in that the latter two imply individual belief of the members of the group: for shared belief, the formula $\mathsf{EB}_J\varphi \rightarrow \bigwedge_{i\in J}\mathsf{B}_i\varphi$ is a theorem, and for common belief $\mathsf{CB}_J\varphi \rightarrow \bigwedge_{i\in J}\mathsf{B}_i\varphi$ is so. In contrast and just as group acceptance, we-belief does not imply individual belief: there may be a group belief in $J$ that $\varphi$ while some of the members of $J$ don't believe that $\varphi$. For example, in the former Soviet Union there was a we-belief that the capitalist countries would perish soon, while only very little citizens individually believed this [105].

Failure to imply individual belief makes group acceptance and group belief relevant for the analysis of concepts such as reputation: an agent $i$ may have a good reputation in group $J$ although some members of $J$ do not adhere to the group view. This was explored in [58] in order to provide a logical analysis of reputation. The logics of different kinds of group belief and their interrelations were recently investigated by several authors [40,46,79]. One of the issues that are essentially left open in these approaches is the relation between different contexts: in the existing proposals they are unanalysed objects.

### 11.3 Logic of action

As its name indicates, PAL presupposes that announcements are perceived correctly and completely by all agents. We have mentioned in Sect. 6.3 that Baltag et al.'s event models provide

an elegant generalisation where the agents' perception of the action is modelled by means of a Kripke model whose points are possible events and where the possible events $a$ and $b$ are related by the accessibility relation for agent $i$ if and only if $a$ and $b$ are indistinguishable for $i$ [14]. This gave rise to a flourishing research avenue that was baptised dynamic epistemic logics. We believe that its version with assignments as designed in [117] provides a promising basis for a unified logic allowing for reasoning about non-strategic action in MAS.

While being theoretically elegant, event models are rather intricate to handle when modelling applications where there are few agents who perceive an action while most of the agents don't, which is a typical situation for example in cryptographic protocols. This is due to the occurrence of event models in the language of the logic as such: the language provides no other means to talk about actions. This is at odds with the idea of syntactic knowledge representation that we have followed in our presentations. It is also not clear how an appropriate language not incorporating such a semantical devices can be designed. First steps are in [8,116] which propose syntactical constructions that are interpreted in event models. The recent paper by Hales et al. [47] establishes that such a language can be built from Gerbrandy's logic of private announcements [42] together with the PDL program operators.

We have mentioned in Sect. 4.2 that there is currently a lot of activity around strategic reasoning, with a dedicated workshop series that started in 2013. A major challenge is to tame complexity. Beyond extensions and fragments of strategy logic, a further interesting research avenue is to provide such results for stit logics that were mentioned in Sect. 7 as a framework for reasoning about uniform strategies. Some decidability results are in [99].

### 11.4 Logic of goals (including group goals)

In the introduction we have mentioned three fundamental concepts in MAS: knowledge, actions and goals. However, in the present paper we did not talk at all about logics of goals, nor about logics for the closely related concept of intention. One of the reasons for that is that there is far less consensus here than about logics of knowledge and action. It is a major challenge to provide logics improving that state of affairs. Just as in the case of group beliefs, one of the most important issues is to logically analyse the relation between group goals and individual goals. More generally, what remains to be done is an integration of so-called belief-desire-intention (BDI) logics [31,32] with strategic logics such as ATL. First steps are in [123].

### 11.5 The relation with game theory

Game theory is seen by many as the most prominent theory of interaction between rational agents. Some of the logics that we have presented have strong links with game theory: ATL, CL and STIT have semantic structures that were taken over from game theory. A challenge for future research is to integrate into these logics the notion of utility. First steps in this direction were undertaken recently by several researchers, typically encoding utilities by means of special propositional variables [26,76,78,107]. A more principled approach should involve an analysis of the notion of goal, which brings us back to the preceding challenge.

We note that there is also a sort of mismatch that remains to be investigated: while Bratman, Cohen & Levesque and others have forcefully argued that logics of belief and goal should be augmented by the concept of future-oriented intention—leading to BDI logics [31,32])—, game theory in its current state does not include such a concept. It therefore remains to be investigated whether the concept of intention could enrich game theory or not.

## 11.6 A general technical problem

Let us terminate on a more general note. As is always the case when modelling real-world problems, there is a delicate balance between conceptual richness and mathematical complexity. Each of the logics of knowledge and action that we have presented is a *multi-dimensional modal logic*: a product logic or a semi-product [39]. Such logics are often undecidable, and when they are decidable then they are in general very complex. This is not a problem as long as one only wants to *model* some application. However, as soon as one aims at the mechanical verification of MAS specifications then decidability and complexity become relevant. We believe that one should then start from a very simple logic of action and a very simple logic of knowledge. In this paper we have advocated `DL-PA` as a simple yet powerful logic of action. Perhaps models of knowledge where knowledge is based on an underlying relation of 'propositional perception' where an agent only sees the truth value of some propositional variables [111] is a good starting point for a simple logic of knowledge.

**Acknowledgments** The ideas in this paper were presented in an invited talk at the 2013 International Joint Conference on Artificial Intelligence (IJCAI 2013) that was entitled "Logics for multiagent systems: a critical overview". It is based on joint work with several colleagues that I would like to thank here for our smooth and fruitful collaboration: Philippe Balbiani (U. Toulouse, CNRS), Jan Broersen (U. Utrecht), Hans van Ditmarsch (U. Nancy, CNRS), Benoit Gaudou (U. Toulouse), Tiago de Lima (U. Artois, CNRS), Emiliano Lorini (U. Toulouse, CNRS), Faustine Maffre (U. Toulouse), Frédédric Moisan (U. Toulouse), François Schwarzentruber (U. Rennes, ENS), Nicolas Troquard (CNR, Trento), Dirk Walther (U. Politecnica de Madrid/U. Dresden). The overview presented in this paper is part of the efforts of the SINTELNET network (www.sintelnet.eu) to reexamine the logical foundations of concepts that are commonly used in artificial intelligence and the social sciences.

## References

1. Ågotnes, T., Goranko, V., & Jamroga, W. (2007). Alternating-time temporal logics with irrevocable strategies. In: D. Samet (Ed.), *Proceedings of the 11th Conference on TARK* (pp. 15–24).
2. Ågotnes, T., Balbiani, P., van Ditmarsch, H. P., & Seban, P. (2010). Group announcement logic. *Journal of Applied Logic*, *8*, 62–81.
3. Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, *50*, 510–530.
4. Alechina, N. (2013). Logic and agent programming languages. In L. Libkin, U. Kohlenbach, & R. J. G. B. de Queiroz (Eds.), *WoLLIC. Lecture notes in computer science* (Vol. 8071, pp. 1–10). Berlin: Springer.
5. Alur, R., Henzinger, T. A., & Kupferman, O. (2002). Alternating-time temporal logic. *Journal of the ACM*, *49*(5), 672–713.
6. Aucher, G. (2004). A combined system for update logic and belief revision. In M. Barley & N. K. Kasabov (Eds.), *PRIMA. Lecture notes in computer science* (Vol. 3371, pp. 1–17). Berlin: Springer.
7. Aucher, G. (2008). Internal models and private multi-agent belief revision. In L. Padgham, D. C. Parkes, J. P. Müller, & S. Parsons (Eds.), *Proceedings of the 7th International Joint Conference AAMAS* (Vol. 2, pp. 721–727). IFAAMAS.
8. Aucher, G. (2008). *Perspectives on belief and change*. PhD thesis, University of Otago and University Paul Sabatier.
9. Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, *4*(6), 1236–1239.
10. Balbiani, P., Baltag, A., van Ditmarsch, H. P., Herzig, A., Hoshi, T., & de Lima, T. (2008). Knowable'as 'known after an announcement. *Review of Symbolic Logic*, *1*(3), 305–334.
11. Balbiani, P., Herzig, A., & Troquard, N. (2008). Alternative axiomatics and complexity of deliberative STIT theories. *Journal of Philosophical Logic*, *37*(4), 387–406.
12. Balbiani, P., Herzig, A., &Troquard, N. (2013). Dynamic logic of propositional assignments: A well-behaved variant of PDL. In O. Kupferman (Ed.), *Logic in computer science (LICS)* (pp. 143–152), June 25–28, 2013. New Orleans, LA: ACM/IEEE. Retrieved from http://www.ieee.org/.

13. Balbiani, P., van Ditmarsch, H. P., Herzig, A., & de Lima, T. (2012). Some truths are best left unsaid. In T. Bolander, T. Braüner, S. Ghilardi, & L. S. Moss (Eds.), *Advances in modal logic* (pp. 36–54). London: College Publications.
14. Baltag, A., & Moss, L. S. (2004). Logics for epistemic programs. *Synthese*, *139*(2), 165–224.
15. Baltag, A., & Smets, S. (2009). Group belief dynamics under iterated revision: Fixed points and cycles of joint upgrades. In A. Heifetz (Ed.), *Proceedings of the 12th Conference on TARK* (pp. 41–50).
16. Baltag, A., Gierasimczuk, N., & Smets, S. (2011). Belief revision as a truth-tracking process. In K. R. Apt (Ed.), *Proceedings of the 13th Conference on TARK* (pp. 187–190). ACM.
17. Baltag, A., Renne, B., & Smets, S. (2014). The logic of justified belief, explicit knowledge, and conclusive evidence. *Annals of Pure and Applied Logic*, *165*(1), 49–81.
18. Belardinelli, F. (2014). Reasoning about knowledge and strategies: Epistemic strategy logic. In F. Mogavero, A. Murano, & M. Y. Vardi (Eds.) *SR, EPTCS* (Vol. 146, pp. 27–33).
19. Belnap, N., Perloff, M., & Xu, M. (2001). *Facing the future: Agents and choices in our indeterminist world*. Oxford: Oxford University Press.
20. Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic. Cambridge tracts in theoretical computer science*. Cambridge: Cambridge University Press.
21. Bonzon, E., Lagasquie-Schiex, M. C., Lang, J., & Zanuttini, B. (2006). Boolean games revisited. In G. Brewka, S. Coradeschi, A. Perini, & P. Traverso (Eds.), *ECAI. Frontiers in artificial intelligence and applications* (Vol. 141, pp. 265–269). Amsterdam: IOS Press.
22. Brihaye, T., da Costa Lopes, A., Laroussinie, F., & Markey, N. (2009). ATL with strategy contexts and bounded memory. *Proceedings of LFSC'09. LNCS* (Vol. 5407, pp. 92–106). Berlin: Springer.
23. Broersen, J., & Herzig, A. (to appear). Using STIT theory to talk about strategies. In J. van Benthem, S. Ghosh, & R. Verbrugge (Eds.), *Modeling strategic reasoning, texts in logic and games*. Springer.
24. Broersen, J., Herzig, A., & Troquard, N. (2006). Embedding alternating-time temporal logic in strategic STIT logic of agency. *Journal of Logic and Computation*, *16*(5), 559–578.
25. Broersen, J., Herzig, A., & Troquard, N. (2009). What groups do, can do, and know they can do: An analysis in normal modal logics. *Journal of Applied Non-Classical Logics*, *19*(3), 261–289.
26. Bulling, N., & Goranko, V. (2013). How to be both rich and happy: Combining quantitative and qualitative strategic reasoning about multi-player games (extended abstract). In F. Mogavero, A. Murano, M. Y. Vardi (Eds.), *Proceedings 1st International Workshop on Strategic Reasoning, SR 2013, EPTCS* (Vol. 112, pp. 33–41), March 16–17, 2013, Rome, Italy.
27. Bulling, N., Jamroga, W., & Dix, J. (2008). Reasoning about temporal properties of rational play. *Annals of Mathematics and Artificial Intelligence*, *53*(1–4), 51–114.
28. Chareton, C., Brunel, J., & Chemouil, D. (2013). Towards an updatable strategy logic. In F. Mogavero, A. Murano, M. Y. Vardi (Eds.), *Proceedings 1st International Workshop on Strategic Reasoning, SR 2013, EPTCS* (Vol. 112, pp. 91–98), March 16–17, 2013, Rome, Italy.
29. Chellas, B. (1980). *Modal logic: An introduction*. Cambridge, MA: Cambridge University Press.
30. Clark, H. H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding*. Cambridge, MA: Cambridge University Press.
31. Cohen, P. R., & Levesque, H. J. (1990). Intention is choice with commitment. *Artificial Intelligence*, *42*(2–3), 213–261.
32. Cohen, P. R., & Levesque, H. J. (1990). Persistence, intentions, and commitment. *Intentions in communication, chap 3* (pp. 33–69). Cambridge, MA: MIT Press.
33. Doutre, S., Herzig, A., & Perrussel, L. (2014). A dynamic logic framework for abstract argumentation. In C. Baral, G. De Giacomo (Eds.), *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR 2014)*. AAAI Press.
34. Elgesem, D. (1997). The modal logic of agency. *Nordic Journal of Philosophical Logic*, *2*(2), 1–46.
35. Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. Cambridge, MA: MIT Press.
36. Fariñas del Cerro, L., Herzig, A., & Su, E. I. (2013). Combining equilibrium logic and dynamic logic. *Logic programming and nonmonotonic reasoning (LPNMR). LNAI* (Vol. 8148, pp. 304–316). Berlin: Springer.
37. French, T., & van Ditmarsch, H. P. (2008). Undecidability for arbitrary public announcement logic. In C. Areces & R. Goldblatt (Eds.), *Advances in modal logic* (pp. 23–42). London: College Publications.
38. French, T., van der Hoek, W., Iliev, P., & Kooi, B. P. (2011). Succinctness of epistemic languages. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence* (pp. 881–886).
39. Gabbay, D. M., Kurucz, A., Wolter, F., & Zakharyaschev, M. (2003). *Many-dimensional modal logics: Theory and applications. Studies in logic and the foundations of mathematics* (Vol. 148). North-Holland: Elsevier.

40. Gaudou, B., Herzig, A., & Longin, D. (2007). Grounding and the expression of belief. In P. Doherty, J. Mylopoulos, & C.Welty (Eds.), *Proceedings of the 10th International Conference on Principles of Knowledge Representation and Reasoning (KR'06)* (pp. 221–229). AAAI Press.
41. Gelfond, M., & Lifschitz, V. (1998). Action languages. *Electronic Transactions on Artificial Intelligence*, *2*, 193–210.
42. Gerbrandy, J., & Groeneveld, W. (1997). Reasoning about information change. *Journal of Logic, Language, and Information*, *6*, 147–196.
43. Gochet, P., & Gribomont, P. (2003). Epistemic logic. In D. Gabbay & J. Woods (Eds.), *Handbook of the history and philosophy of logic* (Vol. 4, pp. 463–502). Amsterdam: Elsevier Science.
44. Goranko, V. (2001). Coalition games and alternating temporal logics. *TARK'01: Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge* (pp. 259–272). San Francisco, CA: Morgan Kaufmann Publishers Inc.
45. Goranko, V., & van Drimmelen, G. (2006). Complete axiomatization and decidability of alternating-time temporal logic. *Theoretical Computer Science*, *353*(1–3), 93–117.
46. Hakli, R. (2006). Group beliefs and the distinction between belief and acceptance. *Cognitive Systems Research*, *7*(2–3), 286–297.
47. Hales, J., Tay, E., & French, T. (2014). A composable language for action models. In R. Goré & A. Kurucs (Eds.), *Advances in modal logic*. London: College Publications.
48. Halpern, J. Y., & Shore, R. A. (2004). Reasoning about common knowledge with infinitely many agents. *Information and Computation*, *191*(1), 1–40.
49. Halpern, J. Y. & Vardi, M. Y. (1991). Model checking vs. theorem proving: A manifesto. In J. F. Allen, R. Fikes, & E. Sandewall (Eds.) *KR, Morgan Kaufmann* (pp. 325–334).
50. Hansen, J. U. (2011). A hybrid public announcement logic with distributed knowledge. *Electronic Notes in Theoretical Computer Science*, *273*, 33–50.
51. Harel, D. (1984). Dynamic logic. In D. M. Gabbay & F. Günthner (Eds.), *Handbook of philosophical logic* (Vol. II, pp. 497–604). Dordrecht: D. Reidel.
52. Harel, D., Kozen, D., & Tiuryn, J. (2000). *Dynamic logic*. Cambridge, MA: MIT Press.
53. Harrenstein, P., van der Hoek, W., Meyer, J. J., & Witteveen, C. (2001). Boolean games. In *Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge, TARK '01* (pp. 287–298) San Francisco, CA: Morgan Kaufmann Publishers Inc.
54. Herzig, A. (2014). Belief change operations: A short history of nearly everything, told in dynamic logic of propositional assignments. In C. Baral, G. De Giacomo (Eds.) *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning, KR 2014*. AAAI Press.
55. Herzig, A., & Lorini, E. (2010). A dynamic logic of agency I: STIT, abilities and powers. *Journal of Logic, Language and Information*, *19*, 89–121.
56. Herzig, A., & Schwarzentruber, F. (2008). Properties of logics of individual and group agency. In C. Areces & R. Goldblatt (Eds.), *Advances in modal logic (AiML)* (pp. 133–149). Nancy: College Publications.
57. Herzig, A., & Troquard, N. (2006). Knowing how to play: Uniform choices in logics of agency. In H. Nakashima, M. P. Wellman, G. Weiss, & P. Stone (Eds.), *Proceedings of the Fifth International Joint conference on AAMAS* (pp. 209–216). ACM.
58. Herzig, A., Lorini, E., Hübner, J. F., & Vercouter, L. (2010). A logic of trust and reputation. *Logic Journal of the IGPL*, *18*(1), 214–244.
59. Herzig, A., Lorini, E., Moisan, F., & Troquard, N. (2011). A dynamic logic of normative systems. In T. Walsh (Ed.), *International Joint Conference on Artificial Intelligence (IJCAI), IJCAI/AAAI, Barcelona* (pp. 228–233).
60. Herzig, A., Lorini, E., & Troquard, N. (2014). Action theories. In S. O. Hansson & V. F. Hendricks (Eds.), *Handbook of formal philosophy*. Dordrecht: Springer.
61. Herzig, A., Lorini, E., & Walther, D. (2013). Alternating-time temporal logic with explicit actions. In D. Grossi (Ed.), *LORI 2013* (pp. 228–233). Berlin: Springer Verlag.
62. Herzig, A., Menezes, V., Nunes De Barros, L., & Wassermann, R. (2014). On the revision of planning tasks. In T. Schaub (Ed.), *European Conference on Artificial Intelligence (ECAI)*. IOS Press.
63. Herzig, A., Pozos Parra, P., & Schwarzentruber, F. (2014). Belief merging in dynamic logic of propositional assignments. In C. Beierle & C. Meghini (Eds.), *International Symposium on Foundations of Information and Knowledge Systems (FoIKS) (FolKS)*, Bordeaux, March 3–7, 2014. Springer.
64. Horty, J. F. (2001). *Agency and deontic logic*. Oxford: Oxford University Press.
65. Hustadt, U., Dixon, C., Schmidt, R. A., Fisher, M., Meyer, J. J. C., & van der Hoek, W. (2001). Reasoning about agents in the karo framework. In C. Bettini & A. Montanari (Eds.), *Proceedings of the Eighth International Symposium on Temporal Representation and Reasoning (TIME-01)* (pp. 206–213). Los Alamitos, CA: IEEE Press.

66. Jamroga, W., & Ågotnes, T. (2007). Constructive knowledge: what agents can achieve under imperfect information. *Journal of Applied Non-Classical Logics*, *17*(4), 423–475.
67. Jamroga, W., & van der Hoek, W. (2004). Agents that know how to play. *Fundamenta Informaticae*, *63*(2–3), 185–219.
68. Jamroga, W., van der Hoek, W., & Wooldridge, M. (2005). Intentions and strategies in game-like scenarios. In C. Bento, A. Cardoso, & G. Dias (Eds.), *EPIA. Lecture notes in computer science* (Vol. 3808, pp. 512–523). Berlin: Springer.
69. Kooi, B. (2007). Expressivity and completeness for public update logic via reduction axioms. *Journal of Applied Non-Classical Logics*, *17*(2), 231–253.
70. Lakemeyer, G., & Levesque, H. J. (2011). A semantic characterization of a useful fragment of the situation calculus with knowledge. *Artificial Intelligence*, *175*(1), 142–164.
71. Lee, J., Lifschitz, V., & Yang, F. (2013). Action language BC: Preliminary report. In F. Rossi (Ed.), *Proceedings of the Twenty-Third International Joint Conference on IJCAI*. IJCAI/AAAI.
72. Lenzen, W. (1978). *Recent work in epistemic logic*. Amsterdam: North Holland Publishing Company.
73. Lenzen, W. (1995). On the semantics and pragmatics of epistemic attitudes. In A. Laux & H. Wansing (Eds.), *Knowledge and belief in philosophy and AI* (pp. 181–197). Berlin: Akademie Verlag.
74. Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
75. Lomuscio, A., & Raimondi, F. (2006). Model checking knowledge, strategies, and games in multi-agent systems. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (pp. 161–168). ACM.
76. Lorini, E. (2013). On the epistemic foundation for iterated weak dominance: An analysis in a logic of individual and collective attitudes. *Journal of Philosophical Logic*, *42*(6), 863–904.
77. Lorini, E., & Herzig, A. (2013). Direct and indirect common belief. In A. Konzelmann Ziv, H.B. Schmid, & U. Schmid (Eds.), *Collective intentionality. Studies in the philosophy of sociality* (Vol. VII). Springer.
78. Lorini, E., & Schwarzentruber, F. (2010). A modal logic of epistemic games. *Games*, *1*(4), 478–526.
79. Lorini, E., Longin, D., Gaudou, B., & Herzig, A. (2009). The logic of acceptance: Grounding institutions on agents' attitudes. *Journal of Logic and Computation*, *19*, 901–940. doi:10.1093/logcom/exn103.
80. Lorini, E., Longin, D., & Mayor, E. (2014). A logical analysis of responsibility attribution: Emotions, individuals and collectives. *Journal of Logic and Computation*. doi:10.1093/logcom/ext072.
81. Lutz, C. (2006). Complexity and succintness of public announcement logic. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)* (pp. 137–144).
82. McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Mitchie (Eds.), *Machine intelligence* (Vol. 4, pp. 463–502). Edinburgh: Edinburgh University Press.
83. Meyer, J. J. C., & van der Hoek, W. (1995). *Epistemic logic for AI and computer science*. Cambridge, MA: Cambridge University Press.
84. Meyer, J. J. C., van der Hoek, W., & van Linder, B. (1999). A logical approach to the dynamics of commitments. *Artificial Intelligence*, *113*(1–2), 1–40.
85. Mogavero, F., Murano, A., & Vardi, M. Y. (2010). Reasoning about strategies. In *Proceedings of FSTTCS'10* (pp. 133–144).
86. Moore, R. C. (1985). A formal theory of knowledge and action. In J. Hobbs & R. Moore (Eds.), *Formal theories of the commonsense world* (pp. 319–358). Norwood, NJ: Ablex.
87. O'Hearn, P. W., Reynolds, J. C., & Yang, H. (2001). Local reasoning about programs that alter data structures. In L. Fribourg (Ed.), *CSL. Lecture notes in computer science* (Vol. 2142, pp. 1–19). Berlin: Springer.
88. Pauly, M. (2002). A modal logic for coalitional power in games. *Journal of Logic and Computation*, *12*(1), 149–166.
89. Penczek, W., & Lomuscio, A. (2003). Verifying epistemic properties of multi-agent systems via bounded model checking. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems* (pp. 209–216). ACM.
90. Pörn, I. (1977). *Action theory and social science: Some formal models. Synthese library* (Vol. 120). Dordrecht: D. Reidel.
91. Pratt, V. (1980). A near-optimal method for reasoning about action. *Journal of Computer and System Sciences*, *20*, 231–254.
92. Reiter, R. (1991). The frame problem in the situation calculus: A simple solution (sometimes) and a completeness result for goal regression. In V. Lifschitz (Ed.), *Artificial intelligence and mathematical theory of computation: Papers in honor of John McCarthy* (pp. 359–380). San Diego, CA: Academic Press.

93. Reiter, R. (2001). *Knowledge in action: Logical foundations for specifying and implementing dynamical systems*. Cambridge, MA: MIT Press.
94. Roelofsen, F. (2007). Distributed knowledge. *Journal of Applied Non-Classical Logics*, *17*(2), 255–273.
95. Scherl, R., & Levesque, H. J. (1993). The frame problem and knowledge producing actions. In *Proceedings of the Eleventh National Conference on AI (AAAI'93)* (pp. 689–695). AAAI Press.
96. Scherl, R., & Levesque, H. J. (2003). The frame problem and knowledge producing actions. *Artificial Intelligence*, *144*(1–2), 1–39.
97. Schewe, S. (2008). ATL* satisfiability is 2EXPTIME-complete. In L. Aceto, I. Damgård, L. A. Goldberg, M. M. Halldórsson, A. Ingólfsdóttir, & I. Walukiewicz (Eds.), *ICALP (2). Lecture notes in computer science* (Vol. 5126, pp. 373–385). Berlin: Springer.
98. Schobbens, P. Y. (2004). Alternating-time logic with imperfect recall. *Electronic Notes in Theoretical Computer Science*, *85*(2), 82–93.
99. Schwarzentruber, F. (2012). Complexity results of STIT fragments. *Studia Logica*, *100*(5), 1001–1045.
100. Segerberg, K. (Ed.). (1992). *"Logic of Action": Special issue of Studia Logica* (Vol. 51, pp. 3–4). Netherlands: Springer.
101. Shanahan, M. (1997). *Solving the frame problem: A mathematical investigation of the common sense law of inertia*. Cambridge, MA: MIT Press.
102. Shapiro, S., Pagnucco, M., Lespérance, Y., Levesque, H.J. (2000). Iterated belief change in the situation calculus. In *Principles of Knowledge Report and Reasoning: Proceedings of the 7th International Conference* (pp. 527–538).
103. Thielscher, M. (2005). *Reasoning robots—The art and science of programming robotic agents. Applied logic* (Vol. 33). Dordrecht: Springer.
104. Troquard, N., & Walther, D. (2012). On satisfiability in ATL with strategy contexts. *Proceedings of JELIA'12. LNCS* (Vol. 7519, pp. 398–410). Berlin: Springer.
105. Tuomela, R. (1995). *The importance of us: A philosophical study of basic social notions. Stanford series in philosophy*. Stanford, CA: Stanford University Press.
106. van Benthem, J. (2006). One is a lonely number: On the logic of communication. In Z. Chatzidakis, P. Koepke, & W. Pohlers (Eds.), *Logic Colloquium'02 Tech Report PP-2002-27* (pp. 96–129), ILLC Amsterdam (2002). Wellesley, MA: ASL & A.K. Peters.
107. van Benthem, J. (2007). Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, *17*(2), 129–155.
108. van Benthem, J. (2013). Reasoning about strategies. In B. Coecke, L. Ong, & P. Panangaden (Eds.), *Computation, logic, games, and quantum foundations. Lecture notes in computer science* (Vol. 7860, pp. 336–347). Berlin: Springer.
109. van der Hoek, W., & Wooldridge, M. (2003). Cooperation, knowledge, and time: Alternating-time temporal epistemic logic and its applications. *Studia Logica*, *75*, 125–157.
110. van der Hoek, W., & Wooldridge, M. (2005). On the logic of cooperation and propositional control. *Artificial Intelligence*, *164*(1–2), 81–119.
111. van der Hoek, W., Troquard, N., & Wooldridge, M. (2011). Knowledge and control. In L. Sonenberg, P. Stone, K. Tumer, & P. Yolum (Eds.), *The 10th International Conference on AAMAS* (pp. 719–726). IFAAMAS.
112. van der Hoek, W., Walther, D., & Wooldridge, M. (2010). On the logic of cooperation and the transfer of control. *Journal of AI Research (JAIR)*, *37*, 437–477.
113. van Ditmarsch, H. P. (2005). Prolegomena to dynamic logic for belief revision. *Synthese*, *147*(2), 229–275.
114. van Ditmarsch, H. P., Herzig, A., & de Lima, T. (2011). From situation calculus to dynamic logic. *Journal of Logic and Computation*, *21*(2), 179–204.
115. van Ditmarsch, H. P., Herzig, A., & De Lima, T. (2012). Public announcements, public assignments and the complexity of their logic. *Journal of Applied Non-Classical Logics*, *22*(3), 249–273.
116. van Ditmarsch, H. P., Herzig, A., Lorini, E., & Schwarzentruber, F. (2013). Listen to me! Public announcements to agents that pay attention—or not. In D. Grossi (Ed.), *LORI 2013* (pp. 228–233). Berlin: Springer Verlag.
117. van Ditmarsch, H. P., van der Hoek, W., & Kooi, B. P. (2005). Dynamic epistemic logic with assignment. In F. Dignum, V. Dignum, S. Koenig, S. Kraus, M. P. Singh, & M. Wooldridge (Eds.), *Proceedings of the Fourth International Joint Conference on AAMAS* (pp. 141–148). ACM.
118. Vassos, S., Lakemeyer, G., & Levesque, H. J. (2008). First-order strong progression for local-effect basic action theories. In G. Brewka & J. Lang (Ed.), *KR* (pp. 662–672). AAAI Press.
119. Voorbraak, F. (1993). *As far as I know: Epistemic logic and uncertainty*. PhD thesis, Universiteit Utrecht.
120. Walther, D., Lutz, C., Wolter, F., & Wooldridge, M. (2006). ATL satisfiability is indeed EXPTIME-complete. *Journal of Logic and Computation*, *16*(6), 765–787.

121. Walther, D., van der Hoek, W., & Wooldridge, M. (2007). Alternating-time temporal logic with explicit strategies. In *Proceedings of TARK'07* (pp. 269–278). ACM.
122. Wáng, Y. N., & Ågotnes, T. (2013). Public announcement logic with distributed knowledge: Expressivity, completeness and complexity. *Synthese*, *190*(18), 135–162.
123. Yadav, N., & Sardiña, S. (2012). Reasoning about agent programs using ATL-like logics. In L. Fariñas del Cerro, A. Herzig, & J. Mengin (Eds.), *JELIA. Lecture notes in computer science* (Vol. 7519, pp. 437–449). Berlin: Springer.