Knowledge-based strategies for multi-agent teams playing against Nature

Dilian Gurov [a,*], Valentin Goranko [b,c,1], Edvin Lundberg [d,2]

[a] *KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science, Lindstedtsvägen 3, 100 44 Stockholm, Sweden*
[b] *Stockholm University, Sweden*
[c] *Institute for Intelligent Systems, University of Johannesburg, South Africa (visiting professorship)*
[d] *Rocker AB, Sweden*

A R T I C L E   I N F O

A B S T R A C T

We study teams of agents that play against Nature towards achieving a common objective. The agents are assumed to have imperfect information due to partial observability, and have no communication during the play of the game. We propose a natural notion of *higher-order knowledge* of agents. Based on this notion, we define a class of knowledge-based strategies, and consider the problem of synthesis of strategies of this class. We introduce a multi-agent extension, MKBSC, of the well-known *knowledge-based subset construction* applied to such games. Its iterative applications turn out to compute higher-order knowledge of the agents. We show how the MKBSC can be used for the design of knowledge-based strategy profiles, and investigate the transfer of existence of such strategies between the original game and in the iterated applications of the MKBSC, under some natural assumptions. We also relate and compare the "intensional" view on knowledge-based strategies based on explicit knowledge representation and update, with the "extensional" view on finite memory strategies based on finite transducers and show that, in a certain sense, these are equivalent.

1. Introduction

In this work we explore the strategy synthesis problem for teams (or coalitions) of agents that have to accomplish a given common objective, while acting under imperfect information and under various other natural assumptions. In particular, we are interested in the notion of knowledge of agents in that context and how it affects the strategic abilities of a team.

When attempting to achieve an objective, intelligent agents act upon their knowledge: about the structure of the game itself, the history of the play so far, the other agents' strategies, observations, and actions. Knowledge has various aspects, but in the context of the present study the term refers to information that is structured and represented in a suitable way to be used by an agent for deciding on its actions towards achieving an objective. This knowledge can be "static", i.e., about the game structure, or "dynamic", i.e., about the play of the game. While the static knowledge can be assumed as "built" into the agents' brain or design, the dynamic knowledge is re-computed and, if necessary, stored on-the-fly during the play.

---

* Corresponding author. ORCID id: 0000-0002-0074-8786.
  E-mail addresses: dilian@kth.se (D. Gurov), valentin.goranko@philosophy.su.se (V. Goranko), edvin_lundberg@msn.com (E. Lundberg).
[1] ORCID id: 0000-0002-0157-1644.
[2] ORCID id: 0000-0003-2109-2525.

Motivation   Strategy[3] synthesis for teams of agents is a complex problem. In general, if no bound is put on the size of the memory of the agents, the strategy synthesis problem is undecidable for coalitions of two or more agents in the presence of imperfect information, even for some basic classes of objectives (see, e.g., [1]).

When information is imperfect, agents typically need to maintain and use a finite abstraction of the history in order to be able to achieve the objective. We refer to this information, suitably structured for use, as "(dynamic) knowledge". By "knowledge states" for an agent, in our context we mean sets of locations which the agent considers currently possible to be the actual location. We call strategies that are directly based on knowledge, i.e., strategies that map knowledge states to actions and update the knowledge state during play, "knowledge-based strategies". Such strategies can be attractive, since they are convenient for play, and are natural to explain to humans.

To achieve certain objectives, agents may even have to maintain "higher-order" knowledge (i.e., knowledge about each other's knowledge). Intuitively, the higher the order (or nesting depth) of knowledge, the higher the strategic abilities of the coalition.

For a bounded order of knowledge, the space of potential knowledge states is finite and the synthesis problem of knowledge-based strategies becomes decidable. It is this class of strategies and their synthesis that we investigate here.

Approach   We study the above problem in the context of *multi-agent games with imperfect information against Nature* (or MAGIIAN for short). We make the following assumptions on the games and assume that they are common knowledge amongst all agents:

1. the game arena is discrete, finite, and known to the agents,
2. certain game states are indistinguishable for certain agents, thus modelling the "imperfect information",
3. the agents cooperate, i.e., they are all in one team playing against Nature,
4. the agents may or may not see each others' actions,
5. the agents cannot explicitly communicate with each other,
6. the agents may or may not know each others' strategies.

We argue that the less the agents know or observe, i.e., the higher their uncertainty is about the current state-of-affairs, the higher the impact is of maintaining higher-order knowledge. Thus, the case where agents cannot observe each others' actions and do not know each others' strategies is a natural starting point for studying also the less restricted cases, which we will discuss below.

As explained above, we only consider knowledge representations with bounded memory (since the unbounded case gives rise to undecidability results). However, within that class there is no a priori best choice of knowledge representation. One may choose, for instance, to use the memory to remember the last n observed game locations; but most generally, the memory is used to compute and maintain some abstraction over the observed history of locations.

Inspired by a subset construction on single-agent games against Nature, namely the Knowledge-Based Subset Construction (or KBSC for short), which reduces games with imperfect information in a strategy-preserving fashion to "expanded" games of perfect information (see, e.g., [2,3]), we choose a representation based on sets of game locations. The semantic interpretation of such a set is "the best estimate the agent can make about the current state-of-affairs". In the single-agent case this representation turns out to be sufficient for the class of parity objectives, as shown in [3]. Then, we represent higher-order knowledge by nesting recursively such sets of locations in a suitable fashion.

Also inspired by the KBSC, we investigate the correspondence between knowledge-based strategies and memoryless, observation-based strategies in expanded games resulting from applying a generalised, multi-agent version of the KBSC, which we introduce here and call the MKBSC. The locations of the expanded games are conceptually joint knowledge states of the agents. We call these two views on strategies the "intensional" and the "extensional" view, respectively. The MKBSC can be iterated, essentially computing higher-order knowledge (i.e., incrementing the order of knowledge with each iteration).

The correspondence between the two views is useful in several ways. First, one can reduce the synthesis problem of knowledge-based strategies to the synthesis problem of memoryless, observation-based strategies. Furthermore, by virtue of the MKBSC construction, the individual observation-based memoryless strategies in the expanded games are simultaneously memoryless strategies in single-agent games of perfect information that are intermediate games produced while computing the expansions. This can serve as the basis for the design of efficient knowledge-based strategy synthesis algorithms, since algorithmic strategy synthesis for the latter class is well-studied (see e.g. [4]). Second, while strategy synthesis is more conveniently performed on the expanded games, once synthesised, the strategies can be presented to the agents as knowledge-based strategies, without the need for storing the expanded games, but by recomputing the knowledge in the course of the play (i.e., on-the-fly). And third, there is a phenomenon that manifests itself much more explicitly in the extensional view: for some games, the iterated MKBSC "stabilises", producing isomorphic games from some iteration on. In the intensional view, stabilisation corresponds to the existence of a finite knowledge representation that contains the higher-order knowledge of the agents of *any* order.

---

[3] Also called "policy" in the literature on planning and Dec-POMDP.

Contributions   In this work we offer the following results and contributions. First, we propose a formal notion of higher-order knowledge with a representation and semantic interpretation, and a notion of knowledge update. Based on this, we provide a formal notion of knowledge-based strategies. Next, we develop a generalisation of the KBSC to the case of multiple agents, as a scheme that can be reused for other similar "expansions". The construction in effect computes knowledge, and its iteration computes higher-order knowledge. For this construction, we show a strategy preservation result for perfect recall strategies with respect to reachability and safety objectives. The proof of this result is constructive, and also reveals how to preserve finite-memory strategies. We then establish a formal relationship between memoryless observation-based strategies in the expanded games, knowledge-based strategies in the original games, and the corresponding finite-memory strategies in the original games, and the equivalence of the latter two. With this we also exhibit formally the duality between the intensional and the extensional views. From this correspondence, we obtain a reduction of the synthesis problem of knowledge-based strategies to that of memoryless observation-based strategies in expanded games. Then, we sketch a heuristic for strategy synthesis, exploiting that the individual observation-based memoryless strategies in the expanded multi-agent games of imperfect information are simultaneously strategies in single-agent games of perfect information, which are intermediate games produced while computing the expansions. Further, we give a formal meaning to the statement that the higher the order of knowledge, the higher the strategic abilities of the team, and argue that this indeed is the case here. (However, this increase is not strict, as will be explained further.) Finally, we establish that for some games the iterated MKBSC stabilises, in the sense that from some iteration on it results in isomorphic games. One implication of this is that for stabilising games, the problem of existence of a winning knowledge-based strategy without a given bound on the knowledge nesting depth, is decidable.

Related work   The present work is in the intersection of several major research areas, including decentralised cooperative decision making (often modelled by decentralised POMDPs), multi-agent planning, knowledge-based programs, games with imperfect information and strategy synthesis in them, etc. There is a huge body of more or less related literature, which we cannot possibly survey in any reasonable degree of detail here. So we only mention and briefly discuss some of the conceptually and technically closest works to ours and provide extensive, yet inevitably incomplete, lists of relevant references for these in Section 7.

Structure   The paper is organised as follows. Section 2 presents the strategy synthesis problem studied here. In particular, we show a motivating example where one needs knowledge of at least second-order in order to achieve the given objective. In Section 3 we define formally MAGIIAN, the formal object of our study, and recall some standard notions from the theory of games over finite graphs. Section 4 is the central section of this paper, in which we define the MKBSC expansion, study its properties with respect to the preservation of certain classes of objectives, and describe how the construction can be used for the synthesis of first-order knowledge-based strategies (the intensional view), or alternatively, of finite-memory strategies in the form of transducers (the dual, extensional view). In Section 5 we study the iterated MKBSC construction and how it can be used for the synthesis of higher-order knowledge-based strategies. In Section 6 we discuss the phenomenon of possible stabilisation in the iterated MKBSC construction, its implications on the strategy synthesis problem, and some limitations of the construction. Section 7 discusses in some detail related work, while in Section 8 we summarise our conclusions from the current work and provide directions for future work.

2. Synthesis of knowledge-based strategies

In this section, we offer an informal discussion on knowledge-based strategies and their synthesis. We then describe the concrete strategy profile synthesis problems studied in the paper.

2.1. Knowledge-based strategies

By a **knowledge-based strategy** we mean a strategy that uses suitably structured knowledge to determine the agent's course of action. That notion is conceptually very close to *knowledge-based programs*, cf. [5,6], but here it is used in the context of multi-agent games against Nature, defined in Section 4. More precisely, a **knowledge-based strategy** consists of:

1. a **knowledge representation** (especially, for the dynamic knowledge) by a suitable data structure;
2. a **knowledge update** function that computes, after every transition in the game, the new knowledge state of the agent, from the old one, the action taken, and the observation made during and upon the transition;
3. an **action mapping**, from knowledge states to prescribed actions of the agent.

The simplest knowledge-based strategies are the *memoryless observation-based strategies*, cf. Section 3.2, where the only knowledge used is the immediate observation of the agent on the current location. Most generally, however, the agent's knowledge is represented by its full observation history, or by some finite abstraction of it. Thus, the most general and abstract case of knowledge-based strategies are the *memory-based strategies*, where the used knowledge is not explicitly represented and structured but implicitly processed in the course of the play, by the strategy-computing device (e.g.,
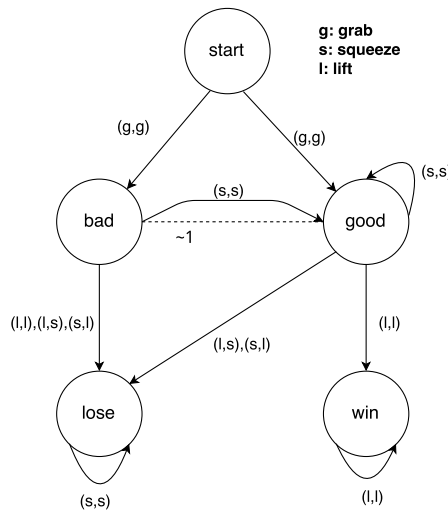
Fig. 1. The two-robot cup-lifting game.

the transducer,[4] in the case of finite-memory strategies). We call this approach to knowledge-based strategies "extensional". Alternatively, there is an "intensional" view, where the knowledge states do have structure, representing the dynamic knowledge of the agents during the course of a play. Several structures for knowledge representation, suitable for strategy design (though, some of them developed for the purpose of epistemic model checking, not strategy synthesis), have been studied, including: *multi-agent epistemic models* [6], *knowledge structures* [7,6], *k-trees* [8], and *epistemic unfolding* [9,10]. Some of the important questions arising here are: *what knowledge is sufficient* to achieve a given objective, and *what is the minimal knowledge needed* for the purpose? We note two further related issues.

First, structured knowledge in general requires memory to be stored and processed. Our idea of using structures for knowledge representation for strategy synthesis is to encode that knowledge in the states of the suitably expanded multi-agent games studied here, where memory-based strategies can be replaced by memoryless ones. The implication of this is that one can use as "knowledge" a data structure that is simply a set of game locations, with the predefined interpretation that it designates "the most precise estimate an agent can make about the current location, based on its initial knowledge about the game and the history of all actions and observations made hitherto" (and this is in effect what the expansion computes as locations of the expanded game). Such knowledge can be updated after each action and observation. Thus, we can interpret memoryless strategies in the expanded games as knowledge-based strategies in the original games, with this particular representation and interpretation of knowledge.

Second, when designing joint strategies of a team of agents acting towards a common goal, it can be essential to take into account their *higher-order knowledge* about the other agents' knowledge. Intuitively, the reason for this is that a given agent from the coalition is not trying to achieve the objective on its own (in which case it would have made sense for the agent to model the other agents as "nature"), but is collaborating with the rest. Therefore, a representation – within an agent's knowledge – of the estimates about the current location, possibly made by the other agents, can offer higher strategic ability for achieving joint objectives. The depth of such (nested) knowledge can increase without bound, and that generates a hierarchy of knowledge-representing structures and a respective hierarchy of knowledge-based strategies. Because that hierarchy may grow strictly, the search for a knowledge-based strategy for a given objective may never terminate, especially if such does not exist. This suggests that the strategy synthesis problem may generally be undecidable, and that is, indeed, the case [11–13].

Example 1. Here is a running example, adapted from [14] and used further in the paper. Consider a scenario where two robots, henceforth referred to as robot 0 and robot 1, must cooperate to lift a cup of acid. Both robots must first grab the cup. Grabbing the cup may non-deterministically result in a good overall grip or a bad one; the grip, however, can always be improved by simultaneously squeezing the cup. Then, both robots must simultaneously lift the cup; otherwise the cup will spill (and the game will be lost). In our scenario, robot 0 has a sensor that detects whether the overall grip is good or not, while robot 1 does not have such a sensor (still, robot 1 can in some situations deduce from its actions and observations that the grip can only be a good one at this point).

The scenario is modelled as the game depicted on Fig. 1. While the formal notion of game will only be defined later, in Section 3, the concrete game model should be intuitively clear. The game is always in some location (i.e., node of the graph), which changes as the result of the joint actions of the robots, represented as action-pairs labelling the edges. Only

---

4 Also called "(local) controller" in the literature on planning and Dec-POMDP.

the available actions (at the respective location) and transitions that they enable are given in the figure. The uncertainty of robot 1 about the grip is modelled as a so-called observation; the corresponding indistinguishability equivalence $\sim_1$ is depicted with a dotted line. In terms of the game graph, the objective of the two robots is to coordinate their actions so as to reach location win.

It should be easy to see that there cannot be an observation-based memoryless[5] strategy that is winning. For a strategy to be observation-based, it has to respect the indistinguishability equivalences (i.e., observations) of the agents. Thus, robot 1 has to perform the same action in locations bad and good. But for the team of robots to win, robot 1 needs to squeeze in location bad and lift in location good.

The situation is similar when we consider first-order knowledge and the corresponding class of strategies. First-order knowledge of an agent is represented as a set of locations. The first-order knowledge of robot 0 will always be a singleton set, representing the current location. The first-order knowledge of robot 1 will be, after the initialisation, represented as the set {bad, good}. Then, if after squeezing robot 1 makes the same observation (i.e., {bad, good}), it can now deduce that it only can be in {good}. (Thus, knowledge is a refinement of what robots observe, using the power of deduction.) However, no first-order knowledge-based strategy can win the game, since knowing {good} is insufficient for robot 0 to decide whether to squeeze or to lift; to win the game, this must depend on whether robot 1 knows {bad, good} or {good}. But robot 0 has no knowledge of this when using first-order knowledge.

This is remedied when using second-order knowledge: if robot 0 not only knows {good}, but also knows whether the first-order knowledge of robot 1 is {bad, good} or {good}, it can make the correct decision whether to squeeze or to lift. As to robot 1, it will squeeze when knowing {bad, good}, and will lift when knowing {good}.

We will use the above scenario as a running example and will elaborate on it throughout the paper.

Now that we have seen that higher-order knowledge can be necessary for achieving objectives, the question arises of how such knowledge can be computed and used for strategy synthesis. As already indicated, our approach is to apply suitable expansions that compute the higher-order knowledge of the agents, and then to search for memoryless strategies in the expanded games.

2.2. Synthesis of knowledge-based strategies

We study the problem of synthesis of knowledge-based strategies in the context of *multi-agent games with imperfect information against Nature* (MAGIIAN, defined in Section 3): given such a game **G** and an **objective** Γ (most generally, a given set of "winning plays"), the task of a central authority – the **team supervisor** – is to design a "winning" knowledge-based strategy profile (consisting of an individual knowledge-based strategy for each agent) for the team Agt that guarantees the achievement of Γ regardless of the behaviour of the environment (Nature). We assume that the game structure is known by the supervisor, hereafter also called the **(strategy) designer**, and is also common knowledge amongst all agents.

Once the strategy profile is designed, each agent is assigned its strategy from the profile and the play begins. It is assumed that there is *no explicit communication* between the agents during the play, that is, the only possible communication is by means of signalling through the agents actions in the model, but not by explicit communicative actions, such as public or private announcements.

Then, four natural cases arise regarding the agents' knowledge and mutual observability during the play, that should be taken into account by the strategy designer *in advance* and used for the design of their strategies:

1. **Case (NN):** *no strategy knowledge and no action observability*. The agents do not know the others' strategies[6] and cannot observe each others' actions during the play, but only their own. This will be our basic case of consideration, as we regard it as the most interesting one.
2. **Case (NY):** *no strategy knowledge but action observability*. The agents do not know the others' strategies, but can observe each others' actions executed during the play.
3. **Case (YN):** *strategy knowledge but no action observability*. The agents know each others' strategies, i.e., the full strategy profile, but cannot observe the others' actions during the play. Note that, because of the partial observability, the agents generally do not know the other agents' observations, hence do not know exactly what actions they execute, so the case remains a priori non-trivial.
4. **Case (YY):** *strategy knowledge and action observability*. The agents know each others' strategies and can observe each others' actions during the play. For deterministic systems, this case is essentially reducible to a single-agent's play in a suitably modified model, as formally proved in [15]. However, such reduction does not work explicitly in the case of non-deterministic games against Nature of the type we consider. Intuitively, this is because after the first transition neither of the agents may know the current state in the game, so even when knowing each others' strategies and

---

[5] Note that all memoryless strategies are also stationary.

[6] It is not trivial to formalise such knowledge, but what we intuitively mean by "agent **a** knowing the strategy of agent **b**" is that at every observation history for **a**, that agent would only consider as possible those actions of **b** that are prescribed by the strategy of **b** known by **a** on some observation history that **a** considers possible for **b** to have observed.

observing each others' actions, they would not be able to fully coordinate their actions so as to be mergeable into a single agent. However, that can still be done in a suitably expanded model constructed as the product of the individual views by the different agents, by applying the general construction presented in Section 4.

Each of these cases has its own justification. For example, the supervisor may inform all agents about the full strategy profile, or may decide not to do that, for reasons of security or privacy. It is important to emphasize that the designer keeps in mind the specific case when designing the strategy profile, because these assumptions may make a difference for the existence of a winning strategy profile. It is generally clear that the more knowledge and observability the agents will have during the play, the greater the chance for existence of a winning strategy profile. What is not obvious, is whether all four cases are strictly different in that respect. To begin with, note that the actions observation ability generally helps agents to obtain more precise knowledge of the current location, and that can be used by the designer to construct a synchronised strategy profile. An illustrating example is given below.

Example 2. The figure below describes a simple turn-based MAGIIAN game with 2 agents a_1 and a_2, whose collective objective is to reach one of the W-states.
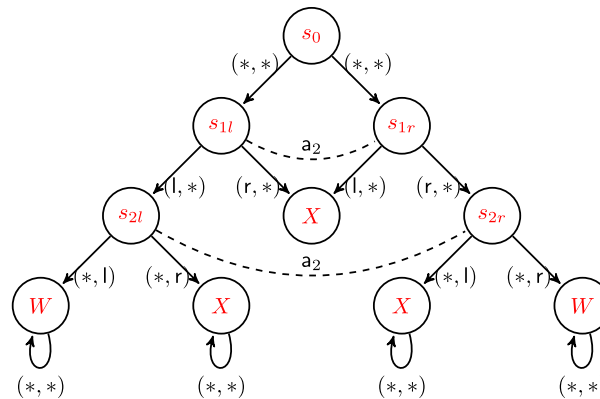
The game goes as follows. First, at s_0 each agent has only one idling action, $*$, and Nature decides to go left, to s_{1l}, or right, to s_{1r}. These successor states are only distinguishable by a_1 but not by a_2 (indicated by a dotted line in the diagram), who has only one action, $*$, at each of these, whereas agent a_1 gets to choose to go left or right. If he does not match Nature's choice the game ends in a bad state, denoted by X, from which no W-state is reachable. Otherwise, the game goes respectively in state s_{2l} or in s_{2r}. These are, again, indistinguishable by a_2, who is to make a left-right choice at each of them there, whereas a_1 has no choice (only one action, $*$).

The choice of a_2 will be successful if and only if a_2 matches the choice of a_1. Clearly, if a_2 could observe that action, she would easily succeed. However, if a_2 cannot observe a_1's action, there is no way that a synchronised action profile can be pre-designed, because the correct action of a_1 cannot be decided in advance but only after Nature has moved. This analysis applies regardless of whether the agents will know each other's strategy at play time.

On the other hand, while knowing the other agents' strategies can be of importance for the knowledge update function of a knowledge-based strategy, it appears that it does not affect the *existence* of a knowledge-based strategy profile achieving the team objective.[7] We consider this claim not intuitively obvious, but the reason, informally, is that the designer can use all the benefit from a common knowledge of the strategy profile at design time in order to synchronise all strategies, without having to provide the individual agents with that knowledge, as noted e.g. in [16].

Likewise, in the **(YY)** case the designer has an apparently stronger power to synthesise a winning strategy profile than in each of the preceding cases. Still, the presumed common knowledge of the strategy profile at the play time is inessential for the existence and design of such a strategy profile, while the observability of the other agents' actions is (the above example also distinguishes the cases **(YN)** and **(YY)**); see also [16].

We are interested in synthesising knowledge-based strategies for each of the cases described above, but in this work we hereafter will focus mainly on the most challenging case **(NN)**.

3. Preliminaries on multi-agent games with imperfect information against Nature

We consider a fixed **team** of n **agents (players)**, $\mathsf{Agt} = \{a_1, \ldots, a_n\}$, which aims to achieve a common goal.

---

[7] This observation was first made to us by Dietmar Berwanger in a private communication.

Definition 3. A **multi-agent game with imperfect information against Nature (MAGIIAN)** is a tuple $\mathbf{G} = (\mathrm{Agt}, \mathrm{Loc}, l_{\mathrm{init}}, \mathrm{Act}, \Delta, \mathrm{Obs})$, where:

(i) Loc is a set of **locations**, usually assumed finite.
(ii) $l_{\mathrm{init}} \in \mathrm{Loc}$ is the **initial location**.
(iii) For each $i \in \mathrm{Agt}$, Act_i is a finite set of **possible actions of agent** i (see remark 3 below).
(iv) $\mathrm{Act} = \mathrm{Act}_{a_1} \times \ldots \times \mathrm{Act}_{a_n}$ are the **possible action profiles** (or joint actions) of the team (see remark 3 below).
(v) $\Delta \subseteq \mathrm{Loc} \times \mathrm{Act} \times \mathrm{Loc}$ is a **transition relation** between locations, with transitions labelled by action profiles.
(vi) For each $i \in \mathrm{Agt}$, Obs_i is a partition of Loc, the blocks of which are the **possible observations of agent** i. Given any location l, the unique observation for i containing l is denoted by $\mathrm{obs}_i(l)$. We denote with $\sim_i$ the equivalence relation on locations induced by the partition.
(vii) $\mathrm{Obs} = \mathrm{Obs}_{a_1} \times \ldots \times \mathrm{Obs}_{a_n}$ is the set of all **observation profiles** (or joint observations) of the team Agt. An observation profile $o \in \mathrm{Obs}$ is **possible** iff $\cap_{i \in \mathrm{Agt}} o(i) \neq \varnothing$. We denote by Obs^p the set of possible observation profiles.

We already saw an example of a MAGIIAN in Example 1 in Section 2.1 above. Some essential remarks are due here:

1. The transition relation is assumed non-deterministic, in general, because the game is played against an unpredictable (and possibly stochastic) **environment**, or **Nature**, the possible behaviours of which are modelled through that non-determinism.
2. We study games of *imperfect information*, where, in general, the agents can only partly observe the current location. This is modelled by *observational equivalence relations between locations* for each agent. Each such relation partitions the set of locations into blocks of indistinguishable locations, which are the possible observations of that agent. The particular case of perfect information is when all observations are singletons.
3. We implicitly assume that all actions of any given agent are available at every location. That is generally not justified, and we only assume it for the sake of technical convenience. Instead, we capture action availability via Δ: if an action profile act contains an action that is not available for the respective agent at the given location l, then no transition is enabled[8] by that action profile from l, i.e., there is no l' such that $(l, \mathrm{act}, l') \in \Delta$. Furthermore, we assume that non-available actions will never be included in the designed strategy profiles.
4. Lastly, a terminological remark: a MAGIIAN model is a variant of a "factored model for a Qualitative Decentralised Partially Observable Markov Decision Problem (QDec-POMDP)", as defined in [17], which is itself a variation of the "QDec-POMDP model" defined in [18] (see further comments on these in Section 7). The (not very essential) differences of our models and QDec-POMDP models are that:
   (a) we assume the agents' observations to be determined by the locations, rather than given by non-deterministic observation functions;
   (b) the goal is not fixed in the model (e.g., as a reachability goal, as in the QDec-POMDP models) but is exogenously specified and can be reachability, safety, or more general, e.g., an LTL-definable objective;
   (c) no horizon is explicitly specified in our models, and we implicitly assume it to be unbounded.
   For a more detailed discussion on the relevant works on QDec-POMDPs and the relation of this study to them, see Section 7.3.
   To avoid possible confusion, not to clutter further the terminology, and to emphasize the importance of MAGIIAN models on their own, we will not use the Dec-POMDP-based terminology but will adopt the acronym MAGIIAN throughout the paper.

3.1. Plays and objectives

The game on $\mathbf{G}$ is played by the agents for infinitely many rounds (in general), starting from the initial location l_{init}. In each round, given the current location $l \in \mathrm{Loc}$, each agent i chooses an action $a_i \in \mathrm{Act}_i$ that is available to i at l, giving rise to an action profile $\mathrm{act} \in \mathrm{Act}$. Then, Nature resolves the non-determinism by choosing the next location $l' \in \mathrm{Loc}$ so that $(l, \mathrm{act}, l') \in \Delta$.

A **full play** in a MAGIIAN $\mathbf{G}$ is an infinite sequence $\pi = l_0 \sigma_1 l_1 \sigma_2 l_2 \ldots$ of alternating locations and action profiles such that $l_0 = l_{\mathrm{init}}$ and $\sigma_j \in \mathrm{Act}$ and $(l_j, \sigma_{j+1}, l_{j+1}) \in \Delta$ for all $j \geq 0$. A **full history** is a finite prefix $\pi(j) = l_0 \sigma_1 l_1 \sigma_2 \ldots l_j$ of a full play π. A **play** is the reduction of a full play to the subsequence of locations, $\pi = l_0 l_1 l_2 \ldots$. Respectively, a **history** is the reduction of a full history to the subsequence of locations, $\pi(j) = l_0 l_1 \ldots l_j$. The last location on a history h is denoted by l(h).

From the perspective of any given agent, a play, resp. history, is a sequence of *observations*, not of locations. Thus, for every agent i, a play $\pi = l_0 l_1 l_2 \ldots$ generates an **observation trace** of that play, which is the sequence of respective

---

[8] In general, this design choice can be problematic when there are adversarial agents, as they can deliberately block a transition from a given state and "sabotage" the evolution of the system by applying an unavailable action at that stage. However, this is not the case in our framework.

observations $\mathsf{obs_i}(l_0)\,\mathsf{obs_i}(l_1)\,\mathsf{obs_i}(l_2)\ldots$ for that agent. Likewise, for any history h we define the **observation history** for the given agent, being the respective finite prefix of the observation trace generated by the play containing the history.

An **objective** for the team Agt in the MAGIIAN **G** is, most generally, a set of plays, declared as **winning plays** for Agt. Often, an objective is expressed by a linear-time temporal logic formula Γ, in the sense that the winning plays are precisely those satisfying that formula, where the atomic propositions in Γ are assumed to have fixed interpretations in **G**. Here are the most common types of objectives that we consider in this work:

- A **reachability objective** can be defined by a non-empty set of locations $\mathsf{R} \subseteq \mathsf{Loc}$. A play $\pi = l_0 l_1 l_2 \ldots$ is winning if it visits some location in R, i.e., if $l_i \in \mathsf{R}$ for some $i \geq 0$.
  A reachability objective is **observable for an agent** i, if it is a union of observations for i, and can therefore be defined alternatively as a set $\mathsf{R} \subseteq \mathsf{Obs_i}$ of observations for i; the objective is **observable (for the team)** if it is observable, at some point in time, for at least one agent in the team, and thus $\mathsf{R} \subseteq \cup_{i \in \mathsf{Agt}} \mathsf{Obs_i}$. The latter notion of observability for the team may not always be justified, and indeed alternative formulations are also possible. However, our results on strategy preservation (see Section 4.2 below) are for the notion stated here.
- A **safety objective** is defined by a non-empty set of locations $\mathsf{S} \subseteq \mathsf{Loc}$. A play $\pi = l_0 l_1 l_2 \ldots$ is winning if it only visits locations in S, i.e., if $l_i \in \mathsf{S}$ for all $i \geq 0$.
  Observable safety objectives are defined similarly to observable reachability objectives. Thus, an observable (for the team) safety objective is a set $\mathsf{S} \subseteq \cup_{i \in \mathsf{Agt}} \mathsf{Obs_i}$, and to win, at every point in time at least one agent must observe the objective. Again, alternative formulations are possible, but our results are for the given one.

In this work we will be concerned with reachability and safety objectives that are observable for the team.

3.2. Observation-based strategies

In games with imperfect information the simplest type of agents' strategies are based on agents' observations. These are also the simplest type of *knowledge-based strategies*, more generally discussed in Section 2.1.

Given a MAGIIAN **G**, a **(deterministic) perfect-recall observation-based strategy** for an agent i is a mapping $\alpha_i : \mathsf{Obs_i}^+ \to \mathsf{Act_i}$ prescribing for every observation history h for the agent i an action $\alpha_i(\mathsf{h})$ that is available for i at every state in l(h).

An observation-based strategy for i is called **memoryless** (or **positional**) if it only takes into account the *current* observation (the last one of the observation history). Such a strategy can be simply presented as a mapping of the type $\mathsf{Obs_i} \to \mathsf{Act_i}$.

A **finite-memory observation-based strategy** is commonly modelled as a finite-state **transducer**, or **Moore machine**, reading game histories and mapping them to actions by using memory states, and is formally defined as follows.

Definition 4 (*Finite-memory strategy*). A **finite-memory observation-based strategy** for agent i in a MAGIIAN **G** = (Agt, Loc, l_{init}, Act, Δ, Obs) is a structure $\mathsf{M_i} = (M, m_0, \mathsf{Obs_i}, \mathsf{Act_i}, \tau, \gamma)$, where:

(i) M is a finite set of **memory states**;
(ii) $m_0 \in M$ is the **initial memory state**;
(iii) $\mathsf{Obs_i}$ and $\mathsf{Act_i}$ are as in **G**;
(iv) $\tau : M \times \mathsf{Obs_i} \rightharpoonup M$ is a (partial) transition function;
(v) $\gamma : M \to \mathsf{Act_i}$ is a mapping from memory states to actions for i.

Agent i follows the strategy encoded by $\mathsf{M_i}$ as follows. In each round, i selects as its next action $\gamma(m)$, where m is the current memory state of $\mathsf{M_i}$ (initially m_0). After the team has applied its action profile and Nature has chosen the next location l, agent i makes the corresponding observation $\mathsf{obs_i}(l)$ and updates its memory state to $\tau(m, \mathsf{obs_i}(l))$.

Note that τ can be partial, since, in the context of a given MAGIIAN **G**, some combinations of memory states and observations might never occur during play.

A **perfect-recall** (resp., **positional**, or **finite-memory**) **observation-based strategy profile** is a strategy profile consisting of perfect-recall (resp., positional, or finite-memory) observation-based strategies.

An **outcome** of an agent's (observation-based) strategy is any play in which the agent chooses its actions according to that strategy. Likewise we define an **outcome of a strategy profile**. Note that, because of potential non-determinism, such outcomes are generally not unique. A strategy profile is **winning for an objective** Γ if all of its outcomes belong to Γ.

It should be noted that the restriction to observable objectives can be partly overcome by using the following technical trick.[9] Given a MAGIIAN **G**, one can transform **G** to another game **G**′ by adding to the set of agents a new, "dummy" agent whose observations are singletons, and who has in all its locations a single idling action at its disposal that is appended to all existing action profiles. Clearly, **G** and **G**′ have the same strategies; furthermore, according to our definition, all objectives in **G**′ are observable, even the ones that correspond to objectives in **G** that are not observable in **G**. Then, if there is no

---

[9] This was suggested to us by a reviewer.

winning strategy in $\mathbf{G}'$ for a given objective, there is no winning strategy in $\mathbf{G}$ either. On the other hand, if there is a winning strategy in $\mathbf{G}'$, it may not be directly usable if the objective is only observed by the dummy agent (recall that we assume that an objective is observable for the team if it is observable for at least one agent in the team), since the latter does not really exist in the actual game $\mathbf{G}$. That is why, the trick with the dummy agent only works partly here, but the latter situation can be easily detected, so no essential problem arises.

3.3. Knowledge-based strategies

We now present a framework for defining knowledge-based strategies of an agent in a MAGIIAN game. It consists of two parts:

I. An **information update module for the agent** i **in the game G**, where:

- $Kn(i)$ represents the a priori knowledge (information) of the agent i about the game, the other agents, their strategies, etc. (to be specified).
- $KS(i)$ is the set of possible knowledge states of the agent i.
- Act_i is the set of actions that i can take.
- Obs_i is the set of observations that i can make (about locations and actions).

The information update module is defined as a mapping:

$$update(\mathbf{G}, i, Kn(i)) : KS(i) \times Act_i \times Obs_i \to KS(i)$$

II. An **information (or, knowledge) based strategy** for i: a mapping:

$$str(i) : KS(i) \to Act_i$$

Observe that finite-memory observation-based strategies as defined in Definition 4 are an instance of the above framework, with M for $KS(i)$, γ for $str(i)$, and τ for $update(\mathbf{G}, i, Kn(i))$, where the latter is restricted to actions as prescribed by $str(i)$.

The rationale behind the above formulation is the following. The possible knowledge states are abstractions over the sequences of actions and observations of the agents, and are updated upon each action and observation. In a knowledge-based strategy, the next action of an agent is completely determined by its current knowledge state.

4. A multi-agent knowledge-based subset construction

Here we introduce and study a new construction, which generalises to the multi-agent case the well-known *knowledge-based subset construction* (KBSC) [2]. The KBSC transforms single-agent games with imperfect information to (expanded) single-agent games with perfect information. The transformation is strategy-preserving for the large class of parity objectives, cf. [3]. We do not present here the KBSC on its own, but as a component of the generalised construction.

Note that the results of this section concern first-order knowledge only, whereas higher-order knowledge will be studied in the following section, in the context of the iterated construction.

4.1. Generalising the KBSC

To connect knowledge-based strategies in multi-agent games to observation-based memoryless strategies in expanded games, we propose a generic scheme for expansion, which is independent of the concrete knowledge representation and the concrete assumptions on what the agents can know and observe.

Our generic scheme for extending the KBSC to the multi-agent setting consists of four stages:

1. **Projection**: for each agent $i \in Agt$, compute the individual views of the input game $\mathbf{G}$, based on what the agent knows, does and sees. This stage results in n single-agent games with imperfect information.
2. **Expansion**: expand each of the individual views with the KBSC. The results are n single-agent games with perfect information.
3. **Composition**: combine the individual expansions by using a product construction, resulting in a single multi-agent game with perfect information.
4. **Partition**: define each agent's observations as induced by the composition product, reflecting their local knowledge. The final result is a multi-agent game with imperfect information.

A concrete instantiation of that scheme is the **Multi-Agent Knowledge-Based Subset Construction (MKBSC)** for the case **(NN)**, defined below. An implementation of the MKBSC as a tool[10] is described in [19]. The game graphs in the rest of the paper have been produced and visualised with this tool.

---

[10] Available from github.com/helmernylen/mkbsc.
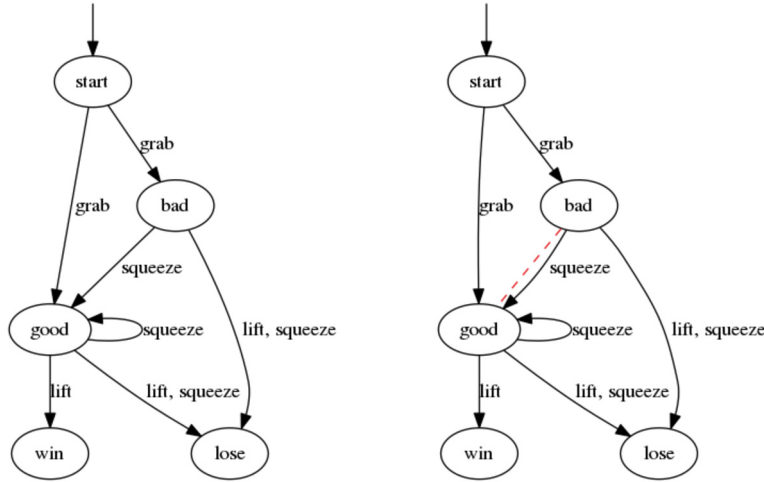
Fig. 2. The individual projections $\mathbf{G}|_0$ and $\mathbf{G}|_1$.

Definition 5 *(MKBSC).* Let $\mathbf{G} = (\mathsf{Agt}, \mathsf{Loc}, l_{\mathsf{init}}, \mathsf{Act}, \Delta, \mathsf{Obs})$ be a MAGIIAN.

1. **Projection**: Given an agent $i \in \mathsf{Agt}$, we define the projection of $\mathbf{G}$ onto i as the single-agent game with imperfect information:

$$\mathbf{G}|_i \stackrel{\mathsf{def}}{=} (\mathsf{Loc}, l_{\mathsf{init}}, \mathsf{Act}_i, \Delta_i, \mathsf{Obs}_i),$$

where $(l, \mathsf{act}_i, l') \in \Delta_i$ iff there exists $\mathsf{act} \in \mathsf{Act}$ such that $\mathsf{act}(i) = \mathsf{act}_i$ and $(l, \mathsf{act}, l') \in \Delta$.

2. **Expansion**: Given $\mathbf{G}|_i$ as above, we define its KBSC expansion, following [3], as the single-agent game with perfect information:

$$(\mathbf{G}|_i)^{\mathsf{K}} \stackrel{\mathsf{def}}{=} (S_i, s_{I,i}, \mathsf{Act}_i, \Delta_i^{\mathsf{K}}),$$

where $S_i \stackrel{\mathsf{def}}{=} \left\{ s \in 2^{\mathsf{Loc}} \setminus \{\varnothing\} \mid \exists o_i \in \mathsf{Obs}_i. \ s \subseteq o_i \right\}$ is the set of possible knowledge states of agent i, $s_{I,i} \stackrel{\mathsf{def}}{=} \{l_{\mathsf{init}}\}$ is its initial knowledge state, and $\Delta_i^{\mathsf{K}} \stackrel{\mathsf{def}}{=} \left\{ (s, \mathsf{act}_i, s') \in S_i \times \mathsf{Act}_i \times S_i \mid \exists o_i \in \mathsf{Obs}_i. \ s' = \left\{ l' \in o_i \mid \exists l \in s. \ (l, \mathsf{act}_i, l') \in \Delta_i \right\} \right\}$.

3. **Composition**: Given $(\mathbf{G}|_i)^{\mathsf{K}}$ as above, for all $i \in \mathsf{Agt}$, we construct their synchronous product, with *joint knowledge states* $S \stackrel{\mathsf{def}}{=} \times_{i \in \mathsf{Agt}} S_i$ and transitions Δ^{K} labelled by joint actions Act. The initial knowledge state $s_I \in S$ is the tuple $(s_{I,i})_{i \in \mathsf{Agt}}$. We *prune* the product by removing inconsistent knowledge states s, i.e., tuples of sets of locations, the intersection $\cap_{i \in \mathsf{Agt}} s(i)$ of which is empty, and unrealisable transitions, i.e., transitions $s \xrightarrow{\mathsf{act}} s'$ for which there is no transition $l \xrightarrow{\mathsf{act}} l'$ in Δ such that $l \in \cap_{i \in \mathsf{Agt}} s(i)$ and $l' \in \cap_{i \in \mathsf{Agt}} s'(i)$.

4. **Partition**: We define observations $\mathsf{Obs}_i^{\mathsf{K}}$ of every agent $i \in \mathsf{Agt}$ by means of the epistemic relations inducing Obs_i^K in terms of the agent's local knowledge:

$$s_1 \sim_i^K s_2 \stackrel{\mathsf{def}}{\Longleftrightarrow} s_1(i) = s_2(i).$$

The observations thus represent indistinguishability with respect to knowledge rather than locations.

The result is the **MKBSC expansion**:

$$\mathbf{G}^{\mathsf{K}} = (\mathsf{Agt}, S, s_I, \mathsf{Act}, \Delta^{\mathsf{K}}, \mathsf{Obs}^{\mathsf{K}})$$

of $\mathbf{G}$, which is a MAGIIAN. Since only the part reachable from s_I is of interest, the rest is disregarded.

A different, but equivalent formulation was originally proposed in [14] by the third co-author. The formulation given here makes explicit how the resulting game is composed of individual expansions, which is the basis for several of our results below.

Example 6. We illustrate the MKBSC construction on our running Example 1, here to be denoted as $\mathbf{G}$. The individual projections $\mathbf{G}|_0$ and $\mathbf{G}|_1$ are shown in Fig. 2, while the individual expansions $(\mathbf{G}|_0)^{\mathsf{K}}$ and $(\mathbf{G}|_1)^{\mathsf{K}}$ of $\mathbf{G}|_0$ and $\mathbf{G}|_1$, respectively, are shown in Fig. 3. And finally, the pruned product $\mathbf{G}^{\mathsf{K}}$ of $(\mathbf{G}|_0)^{\mathsf{K}}$ and $(\mathbf{G}|_1)^{\mathsf{K}}$ is shown in Fig. 4.

Due to the pruning, there only are consistent knowledge states in $\mathbf{G}^{\mathsf{K}}$. There is an edge labelled (squeeze, squeeze) from vertex $(\{\mathsf{bad}\}, \{\mathsf{bad}, \mathsf{good}\})$ to vertex $(\{\mathsf{good}\}, \{\mathsf{good}\})$ in $\mathbf{G}^{\mathsf{K}}$ because there is an edge labelled squeeze from vertex
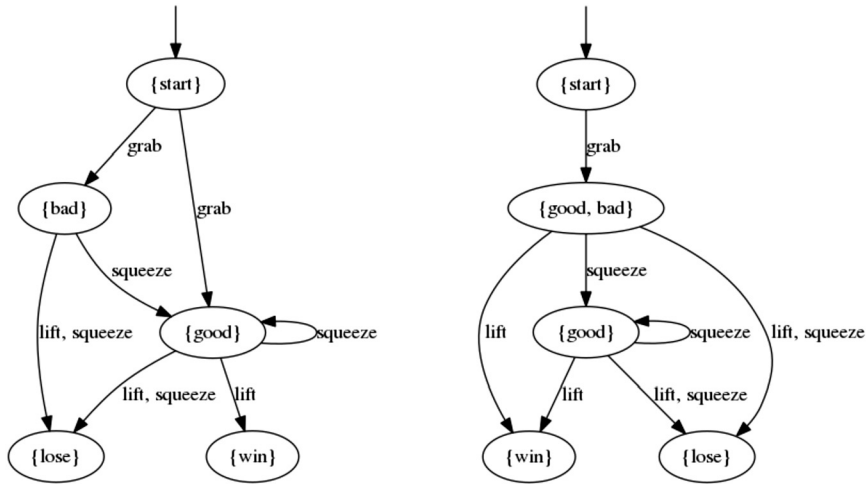
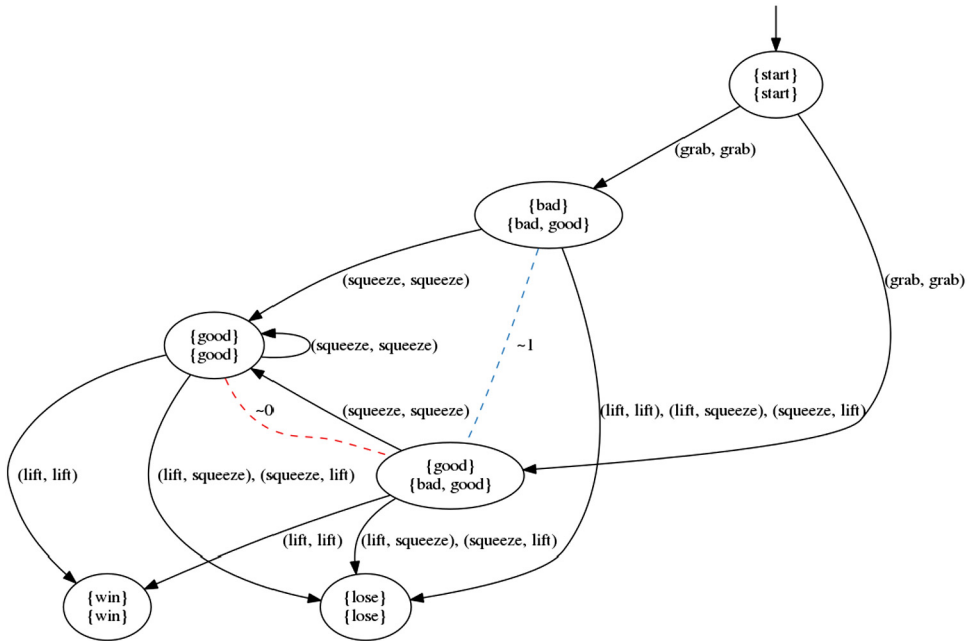**Fig. 3.** The individual expansions $(\mathbf{G}|_0)^K$ and $(\mathbf{G}|_1)^K$.

Fig. 4. The pruned product $\mathbf{G}^K$.

{bad} to vertex {good} in $(\mathbf{G}|_0)^K$, an edge labelled squeeze from {bad, good} to {good} in $(\mathbf{G}|_1)^K$, and an edge labelled (squeeze, squeeze) from location bad to location good in $\mathbf{G}$. Note that if the latter edge had not been present in $\mathbf{G}$, the discussed edge in $\mathbf{G}^K$ would have been unrealisable, and would have been pruned out.

While in this paper we do not focus on the algorithmic aspects of the MKBSC construction, a naive analysis of its space complexity reveals that: (a) projection preserves the locations of $\mathbf{G}$, (b) expansion (being a subset construction) is worst-case exponential in the number of locations of $\mathbf{G}$, and (c) composition results in the product of the numbers of locations of the individual expansions. The space complexity of the construction can thus be upper-bounded by $O(2^{|\mathsf{Loc}| \cdot |\mathsf{Agt}|})$.

For the **(NY)** case, the MKBSC construction can be adapted as follows:

1. The projections do not filter out the complementary actions of the other agents, but keep the full joint actions, and only abstract from the observations of the other agents; this results in the games $\mathbf{G}|_i \stackrel{\mathsf{def}}{=} (\mathsf{Loc}, l_{\mathsf{init}}, \mathsf{Act}, \Delta, \mathsf{Obs}_i)$.
2. The individual expansion stage remains unchanged.
3. The synchronous product has now to synchronise on common action profiles.
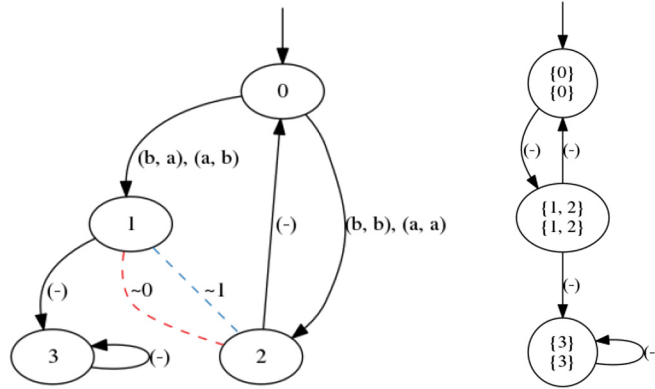4. The partition stage remains unchanged.

Fig. 5. A game **G** and its expansion **G**^{K}.

For example, consider the game **G** shown in Fig. 5 left. In the **(NY)** case, its expansion **G**^{K} is isomorphic to the original game **G**, but has knowledge states $(\{l\}, \{l\})$ for the corresponding locations l in **G**.

In the sequel, unless otherwise specified, we only refer to the **(NN)** case.

4.2. Strategy preservation

In this section we present results on the preservation of *observation-based perfect recall strategies* for observable reachability objectives. Note that every observable reachability objective $R \subseteq \cup_{i \in Agt} Obs_i$ in a MAGIIAN **G** with observations Obs translates uniformly to an observable reachability objective R^K in **G**^{K}, as follows:

$$R^K \overset{def}{=} \{s \in S \mid \exists i \in Agt. \; \exists o \in R \cap Obs_i. \; s(i) \subseteq o\}$$

Likewise, every observable safety objective S in **G** translates uniformly to an observable safety objective S^K in **G**^{K}.

The synchronous product of the MKBSC construction is a form of *existential abstraction*, and can give rise to "spurious" plays in **G**^{K} that are not present in **G**. Such spurious plays can give rise to spurious outcomes of a given strategy, and can thus prevent a strategy from achieving a given objective.

Example 7. Consider the two-agent game **G** and its expansion **G**^{K}, shown in Fig. 5 (where the symbol "(-)" is used to denote any joint action), and let $R \overset{def}{=} \{\{3\}\}$ be our observable reachability objective. The game is easily won in **G** with the profile of observation-based memoryless strategies, where the first agent always does action a, while the second agent always does b. In the game **G**^{K}, however, the (corresponding) strategy is thwarted by the outcome $(\{0\}, \{0\})$ $(\{1, 2\}, \{1, 2\})$ $(\{0\}, \{0\})$, which is spurious: there is no corresponding outcome $0\,1\,0$ or $0\,2\,0$ in **G**.

Thus, it is not possible to preserve arbitrary winning strategies from **G** to **G**^{K}. However, under the additional condition that for every joint knowledge state s that is reachable from the initial one in **G**^{K}, $\cap_{i \in Agt} s(i)$ is a singleton set, no spurious plays exist. This condition can be stated as **perfect distributed knowledge** (or PDK for short).[11]

For instance, the game **G**^{K} from Fig. 4 satisfies the PDK, while **G**^{K} from Fig. 5 does not. An obvious sufficient condition on **G** to guarantee that **G**^{K} fulfills the PDK condition is that no two distinct locations of **G** are indistinguishable by all agents (or equivalently, that any two distinct locations of **G** are distinguishable by at least one agent).[12]

When game **G**^{K} satisfies the PDK, one can view **G**^{K} as a *refinement* of the original game **G**, in the sense that the locations in **G**^{K} are obtained from "splitting" locations of **G**. Since any two locations of **G**^{K} that derive from the same location in **G** will always be distinguishable by at least one agent, one can say that, while the MKBSC does not necessarily eliminate imperfect information (as it does in the single-agent case), it in some sense decreases the degree of imperfectness.

The following result generalises Lemma 3.1 from [3].

Lemma 8. *Let* **G** $= (Agt, Loc, l_{init}, Act, \Delta, Obs)$ *be a MAGIIAN for a set of agents* Agt, *and let* **G**$^K = (Agt, S, s_I, Act, \Delta^K, Obs^K)$ *be its MKBSC expansion. Further, let* $s \in S$, $act \in Act$ *and* $o \in Obs^p$. *Define the set:*

$$X \overset{def}{=} \left\{ l' \in \cap_{i \in Agt} o(i) \;\middle|\; \exists l \in \cap_{i \in Agt} s(i). \; (l, act, l') \in \Delta \right\}$$

---

[11] In the literature on Dec-POMDP this condition is also called "joint observability" [20], "collective observability" [21] and "decentralised full observability" [22], and the models satisfying it are called "decentralised Markov decision processes (Dec-MDP)".

[12] Therefore, the technical trick of adding a dummy agent described at the end of Section 3.2 enforces the PDK condition.

Then, X is non-empty if and only if there is $s' \in S$ such that $(s, \mathsf{act}, s') \in \Delta^K$ and for all $\mathsf{i} \in \mathsf{Agt}$, $s'(\mathsf{i}) \subseteq o(\mathsf{i})$. This $s' \in S$ is then unique and we have $X \subseteq \cap_{\mathsf{i} \in \mathsf{Agt}} s'(\mathsf{i})$, and if $\mathbf{G}^K$ fulfills the PDK condition, we have $X = \cap_{\mathsf{i} \in \mathsf{Agt}} s'(\mathsf{i})$.

Proof. Let $s \in S$, $\mathsf{act} \in \mathsf{Act}$ and $o \in \mathsf{Obs}^p$. The stated equivalence is a direct consequence of the expansion and composition steps of Definition 5; in particular, when there is $s' \in S$ with the stated properties, the set X cannot be empty since otherwise the transition (s, act, s') would be pruned out (as being unrealisable) in the composition step. Furthermore, if $\mathbf{G}^K$ fulfills the PDK condition, the set $\cap_{\mathsf{i} \in \mathsf{Agt}} s'(\mathsf{i})$ must be a singleton, and the two sets must therefore be equal. $\quad\square$

Analogues of this result can be proved likewise for stochastic, rather than non-deterministic models, of the type studied in the literature on Dec-POMDP.

The result lifts naturally to observation histories: every sequence π of joint actions and joint observations of $\mathbf{G}$ (where π is not necessarily a path in $\mathbf{G}$) gives rise to at most one path in $\mathbf{G}^K$ such that at each corresponding step of the two sequences, $s(\mathsf{i}) \subseteq o(\mathsf{i})$ holds for all $\mathsf{i} \in \mathsf{Agt}$. Furthermore, every full play π in $\mathbf{G}$ gives rise to exactly one full play in $\mathbf{G}^K$ that is consistent with the actions and the observations of the agents.

We obtain the following result on strategy preservation under the PDK condition.

Theorem 9 (*Strategy preservation*). *Let $\mathbf{G}$ be a MAGIIAN, and let $\mathbf{G}^K$ be its MKBSC expansion. Assume that $\mathbf{G}^K$ fulfills the PDK condition. Let R be an observable reachability objective in $\mathbf{G}$, and R^K be its translation for $\mathbf{G}^K$. If there is a winning profile of observation-based perfect recall strategies in $\mathbf{G}$ for R, then there is also one in G^K for R^K.*

Proof. Informally, given a profile $\{\alpha_\mathsf{i}\}_{\mathsf{i} \in \mathsf{Agt}}$ of observation-based perfect recall strategies in $\mathbf{G}$, every agent $\mathsf{i} \in \mathsf{Agt}$ plays in $\mathbf{G}^K$ by:

1. recording the history of individual observations in $\mathbf{G}^K$ it has made so far during the play,
2. converting this sequence to the corresponding sequence of observations of the agent in $\mathbf{G}$, and
3. taking the action prescribed by α_i for that sequence of observations.

Formally, let $\{\alpha_\mathsf{i}\}_{\mathsf{i} \in \mathsf{Agt}}$, where $\alpha_\mathsf{i} : \mathsf{Obs}_\mathsf{i}^+ \to \mathsf{Act}_\mathsf{i}$ for all $\mathsf{i} \in \mathsf{Agt}$, be a winning profile of observation-based perfect recall strategies in $\mathbf{G}$ for the observable reachability objective R. For all $\mathsf{i} \in \mathsf{Agt}$, define the functions $\alpha_\mathsf{i}^K : (\mathsf{Obs}_\mathsf{i}^K)^+ \to \mathsf{Act}_\mathsf{i}$ as the function compositions $\alpha_\mathsf{i}^K \stackrel{\mathsf{def}}{=} \alpha_\mathsf{i} \circ \mathsf{ob}_\mathsf{i}$, where for every $o_\mathsf{i}^K \in \mathsf{Obs}_\mathsf{i}^K$, $\mathsf{ob}_\mathsf{i}(o_\mathsf{i}^K)$ denotes the unique observation $o_\mathsf{i} \in \mathsf{Obs}_\mathsf{i}$ in $\mathbf{G}$ such that $A \subseteq o_\mathsf{i}$ holds for the common i'th component A of the tuples comprising o_i^K, and where ob_i is then lifted to sequences. Thus the functions define a profile $\{\alpha_\mathsf{i}^K\}_{\mathsf{i} \in \mathsf{Agt}}$ of observation-based perfect recall strategies in G^K. We show that this profile is winning in G^K for the objective R^K whenever G^K satisfies the PDK condition.

Let $\pi^K = s_0 \sigma_0 s_1 \sigma_1 \ldots$ be an arbitrary outcome of $\{\alpha_\mathsf{i}^K\}_{\mathsf{i} \in \mathsf{Agt}}$ in $\mathbf{G}^K$. Then, by Lemma 8 and since G^K satisfies the PDK condition, the sequence $\pi = l_0 \sigma_0 l_1 \sigma_1 \ldots$ such that $\{l_k\} = \cap_{\mathsf{i} \in \mathsf{Agt}} s_k(\mathsf{i})$ for all $k \geq 0$, must be a full play in $\mathbf{G}$. But then $s_k(\mathsf{i}) \subseteq \mathsf{obs}_\mathsf{i}(l_k)$ for all $\mathsf{i} \in \mathsf{Agt}$ and $k \geq 0$, and thus, by the definitions of α_i^K, π must be an outcome of $\{\alpha_\mathsf{i}\}_{\mathsf{i} \in \mathsf{Agt}}$ in $\mathbf{G}$, and must be winning for R, i.e., there is an agent $\mathsf{i} \in \mathsf{Agt}$ and an index $k \geq 0$ such that $\mathsf{obs}_\mathsf{i}(l_k) \in \mathsf{R}$. But then there is also an agent $\mathsf{i} \in \mathsf{Agt}$, an index $k \geq 0$ and an observation $o \in \mathsf{R} \cap \mathsf{Obs}_\mathsf{i}$ such that $s_k(\mathsf{i}) \subseteq o$, and hence π^K is winning for R^K. Since π^K is arbitrary, the profile $\{\alpha_\mathsf{i}^K\}_{\mathsf{i} \in \mathsf{Agt}}$ must be winning in G^K for the objective R^K. $\quad\square$

It is easy to see from the proof that winning profiles of observation-based perfect recall strategies for *observable safety objectives* $\mathsf{S} \subseteq \cup_{\mathsf{i} \in \mathsf{Agt}} \mathsf{Obs}_\mathsf{i}$ are also preserved, by a slightly modified argument: in the proof, simply replace "there is an index $k \geq 0$" with "for all indices $k \geq 0$".

It is important to observe that the above proof is constructive and shows how observation-based *memoryless* strategies $\alpha_\mathsf{i} : \mathsf{Obs}_\mathsf{i} \to \mathsf{Act}_\mathsf{i}$ in $\mathbf{G}$ are mapped to observation-based memoryless strategies $\alpha_\mathsf{i}^K \stackrel{\mathsf{def}}{=} \alpha_\mathsf{i} \circ \mathsf{ob}_\mathsf{i}$ in $\mathbf{G}^K$.

The following result establishes an important property of expanded games. Let $\mathbf{G}^{2K}$ denote $(\mathbf{G}^K)^K$.

Lemma 10. *Let $\mathbf{G}$ be a MAGIIAN, $\mathbf{G}^K$ be its MKBSC expansion, and $\mathbf{G}^{2K}$ be the MKBSC expansion of $\mathbf{G}^K$. Then, $\mathbf{G}^{2K}$ fulfills the PDK condition.*

Proof. As pointed out above, a sufficient condition on a game to guarantee that its MKBSC expansion fulfills the PDK condition is that no two distinct locations of the game are indistinguishable by all agents. This sufficient condition is enforced by the partition step of Definition 5, since two knowledge states of the expansion can only be indistinguishable by all agents if they are equal.

Formally, the proof proceeds by contradiction. Assume that $\mathbf{G}^{2K}$ does not fulfill the PDK condition. By the definition of the PDK condition, there must then be a knowledge state $\mathbf{s}$ of $\mathbf{G}^{2K}$ such that $\cap_{\mathsf{i} \in \mathsf{Agt}} \mathbf{s}(\mathsf{i})$ is not a singleton set. And since the latter set, due to the pruning in the composition step of Definition 5, also cannot be empty, there must be (at least) two *distinct* knowledge states s_1 and s_2 in $\mathbf{G}^K$ such that $\{s_1, s_2\} \subseteq \mathbf{s}(\mathsf{i})$ for all $\mathsf{i} \in \mathsf{Agt}$. Hence, by the expansion step of Definition 5,

s_1 and s_2 must be indistinguishable in $\mathbf{G}^K$ for all $i \in \mathsf{Agt}$, and therefore, by the partition step of Definition 5, $s_1(i) = s_2(i)$ for all $i \in \mathsf{Agt}$. But then $s_1 = s_2$, and we arrive at a contradiction. $\quad\square$

This result will be important for the properties of the iterated construction studied in Section 5.

Strategy preservation in the reverse direction, from $\mathbf{G}^K$ to $\mathbf{G}$, does *not* depend on the PDK condition.

Theorem 11 (*Reverse strategy preservation*). *Let $\mathbf{G}$ be a MAGIIAN, and let $\mathbf{G}^K$ be its MKBSC expansion. Let R be an observable reachability objective in $\mathbf{G}$, and R^K be its translation for $\mathbf{G}^K$. If there is a winning profile of observation-based perfect recall strategies in G^K for R^K, then there is also one in $\mathbf{G}$ for R.*

Proof. Informally, given a strategy profile $\{\alpha_i^K\}$ in $\mathbf{G}^K$, every agent $i \in \mathsf{Agt}$ plays in $\mathbf{G}$ by:

1. recording the sequence of actions it has taken and observations it has made so far during the play,
2. following the unique path in $\mathbf{G}^K$ that corresponds to this sequence, and
3. for the corresponding sequence of observations in $\mathbf{G}^K$, taking the action as prescribed by α_i^K for that sequence.

Formally, let $\{\alpha_i^K\}_{i \in \mathsf{Agt}}$, where $\alpha_i^K : (\mathsf{Obs}_i^K)^+ \to \mathsf{Act}_i$ for all $i \in \mathsf{Agt}$, be a winning profile of observation-based perfect recall strategies in G^K for the observable reachability objective R^K. For all $i \in \mathsf{Agt}$, we define the functions $\alpha_i : \mathsf{Obs}_i^+ \to \mathsf{Act}_i$ by induction on the length of observation sequences. In the base case, define $\alpha_i(\{l_{\mathsf{init}}\}) \stackrel{\mathsf{def}}{=} \alpha_i^K(\{s_I\})$. Let $\mu = o_0 o_1 \ldots o_m \in \mathsf{Obs}_i^+$ be an observation sequence, where $o_0 = \{l_{\mathsf{init}}\}$, and assume that α is defined for all its prefixes (induction hypothesis). Then, the actions $\sigma_k \stackrel{\mathsf{def}}{=} \alpha_i(\mu(k))$ are also defined for all $k : 0 \le k \le m$. Let $o_{m+1} \in \mathsf{Obs}_i$. By Lemma 8, the observation history $o_0 \sigma_0 o_1 \sigma_1 \ldots o_m \sigma_m o_{m+1}$ defines at most one path $s_0 \sigma_0 s_1 \sigma_1 \ldots s_m \sigma_m s_{m+1}$ in $\mathbf{G}^K$. Define $\alpha_i(\mu \cdot o_{m+1}) \stackrel{\mathsf{def}}{=} \alpha_i^K(s_0 s_1 \ldots s_{m+1})$ if such a path exists (and otherwise its choice is immaterial). Thus the functions define a profile $\{\alpha_i\}_{i \in \mathsf{Agt}}$ of observation-based perfect recall strategies in $\mathbf{G}$. We show that this profile is winning in $\mathbf{G}$ for the objective R.

Let $\pi = l_0 \sigma_0 l_1 \sigma_1 \ldots$ be an arbitrary outcome of $\{\alpha_i\}_{i \in \mathsf{Agt}}$ in $\mathbf{G}$. Then, by Lemma 8, there is a play $\pi^K = s_0 \sigma_0 s_1 \sigma_1 \ldots$ in G^K such that $l_k \in \cap_{i \in \mathsf{Agt}} s_k(i)$ for all $k \ge 0$. By the definitions of α_i, π^K must be an outcome of $\{\alpha_i^K\}$ in $\mathbf{G}^K$, and must thus be winning for R^K, i.e., there is an agent $i \in \mathsf{Agt}$, an index $k \ge 0$ and an observation $o \in \mathsf{R} \cap \mathsf{Obs}_i$ such that $s_k(i) \subseteq o$. But then there is also an agent $i \in \mathsf{Agt}$ and an index $k \ge 0$ such that $\mathsf{obs}_i(l_k) \in \mathsf{R}$, and hence π is winning for R. Since π is arbitrary, the profile $\{\alpha_i\}_{i \in \mathsf{Agt}}$ must be winning in $\mathbf{G}$ for the objective R. $\quad\square$

Again, it is easy to see from the proof that winning profiles of observation-based perfect recall strategies for observable safety objectives are also preserved, by replacing in the proof "there is an index $k \ge 0$" with "for all indices $k \ge 0$".

Further, note that the proof is constructive and reveals how observation-based memoryless strategies in $\mathbf{G}^K$ can be mapped to observation-based finite-memory strategies (i.e., transducers) in $\mathbf{G}$ (which may, in some cases, "degenerate" to memoryless strategies in $\mathbf{G}$). We will make use of this in Section 4.3.1.

4.3. Strategy translation

While the results of the preceding subsection concern profiles of observation-based perfect recall strategies, in this work we focus on searching for profiles of observation-based *memoryless* strategies in the game $\mathbf{G}^K$. For this class of strategies the synthesis problem has already been studied (see e.g. [23]). If such a strategy profile can be found, it needs to be converted to an observation-based strategy profile for play in the original game structure $\mathbf{G}$. For this, we will offer here two solutions:

(*i*) the *extensional solution*, where we convert each individual strategy to an individual observation-based finite-memory strategy (i.e., transducer), and

(*ii*) the *intensional solution*, where we interpret the individual strategies as knowledge-based strategies, based on a (common) knowledge representation and individual update functions.

The two solutions will be shown to be equivalent, i.e., to give rise to the same sets of outcomes.

4.3.1. Translation to transducers

We start with the important observation that, by virtue of how the observations in $\mathbf{G}^K$ are defined in Definition 5, every observation-based memoryless strategy for agent i in $\mathbf{G}^K$ is simultaneously a memoryless strategy in the game with perfect information $(\mathbf{G}|_i)^K$. This observation motivates the following construction, which essentially combines each game $(\mathbf{G}|_i)^K$ and individual memoryless strategy α_i^K into a transducer $A_i(\alpha_i^K)$ for agent i for play in $\mathbf{G}$.

Definition 12 (*Induced transducer*). Let $\mathbf{G} = (\mathsf{Agt}, \mathsf{Loc}, l_{\mathsf{init}}, \mathsf{Act}, \Delta, \mathsf{Obs})$ and for any $i \in \mathsf{Agt}$, let $(\mathbf{G}|_i)^K = (S_i, s_{I,i}, \mathsf{Act}_i, \Delta_i^K)$. Let also $\alpha_i^K : S_i \to \mathsf{Act}_i$ be a memoryless strategy in $(\mathbf{G}|_i)^K$. We define the following α_i^K-**induced transducer**:

$$A_i(\alpha_i^K) \stackrel{\text{def}}{=} (S_i, s_{I,i}, \mathsf{Obs}_i, \mathsf{Act}_i, \tau_i, \alpha_i^K)$$

where $\tau_i(s, o_i)$ is defined for $s \in S_i$ and $o_i \in \mathsf{Obs}_i$ as the unique $s' \in S_i$ such that $s' \subseteq o_i$ and $(s, \alpha_i^K(s), s') \in \Delta_i^K$, if such an s' exists, and is undefined otherwise.

Uniqueness of s' in the definition is guaranteed by Lemma 8.

The transducer $A_i(\alpha_i^K)$ is, by Definition 4, an observation-based finite-memory strategy for agent i in **G**. The transducer can be *pruned* by removing, from each memory state s, the outgoing edges for actions other than $\alpha_i^K(s)$, then by removing the unreachable memory states, and finally by abstracting away the structure of s (since only the identity of the memory states is relevant).

Theorem 13 (*Strategy correspondence*). *Let **G** be a MAGIIAN for a set of agents* Agt, *and let **G**^{K} be its MKBSC expansion. Let* R *be an observable reachability objective in* **G**, *and* R^K *be its translation in* **G**^{K}. *Finally, let* $\{\alpha_i^K\}_{i \in \mathsf{Agt}}$ *be a profile of observation-based memoryless strategies in* **G**^{K}, *and* $\{A_i(\alpha_i^K)\}_{i \in \mathsf{Agt}}$ *be the corresponding profile of induced transducers for* **G**.

(*i*) *If* $\{A_i(\alpha_i^K)\}_{i \in \mathsf{Agt}}$ *is winning for* R *in* **G**, *and* **G**^{K} *fulfills the PDK condition, then* $\{\alpha_i^K\}_{i \in \mathsf{Agt}}$ *is winning for* R^K *in* **G**^{K}.
(*ii*) *If* $\{\alpha_i^K\}_{i \in \mathsf{Agt}}$ *is winning for* R^K *in* **G**^{K}, *then* $\{A_i(\alpha_i^K)\}_{i \in \mathsf{Agt}}$ *is winning for* R *in* **G**.

Proof. (*i*) The proof adapts the strategy construction used in the proof of Theorem 9. Let $\{A_i(\alpha_i^K)\}_{i \in \mathsf{Agt}}$ be winning for R in **G**, and let **G**^{K} fulfill the PDK condition. Let $\pi^K = s_0\sigma_0 s_1\sigma_1 \dots$ be an arbitrary outcome of $\{\alpha_i^K\}_{i \in \mathsf{Agt}}$ in **G**^{K}. Then, by Lemma 8 and since G^K fulfills the PDK condition, the sequence $\pi = l_0\sigma_0 l_1\sigma_1 \dots$ such that $\{l_k\} = \cap_{i \in \mathsf{Agt}} s_k(i)$ for all $k \geq 0$, must be a full play in **G**. Now, by Definition 12, π must be an outcome of $\{A_i(\alpha_i^K)\}_{i \in \mathsf{Agt}}$ in **G**. Since $\{A_i(\alpha_i^K)\}_{i \in \mathsf{Agt}}$ is winning for R in **G**, π must be winning for R in **G**, and hence, by the definition of R^K, π^K must be winning for R^K in **G**^{K}. But π^K is arbitrary, and therefore $\{\alpha_i^K\}_{i \in \mathsf{Agt}}$ must be winning for R^K in **G**^{K}.

(*ii*) The proof adapts the strategy construction used in the proof of Theorem 11, using the observation that observation-based memoryless strategies for agent i in **G**^{K} correspond to memoryless strategies in $(\mathbf{G}|_i)^K$.

Let $\{\alpha_i^K\}_{i \in \mathsf{Agt}}$ be winning for R^K in **G**^{K}. Let $\pi = l_0 l_1 l_2 \dots$ be an arbitrary outcome of $\{A_i(\alpha_i^K)\}_{i \in \mathsf{Agt}}$ in **G**. This outcome induces a corresponding sequence of joint observations, from which, using $\{A_i(\alpha_i^K)\}_{i \in \mathsf{Agt}}$, one can recover the corresponding sequence of joint actions. By Lemma 8, these two sequences (of joint actions and joint observations) give rise to a unique play π^K in **G**^{K}, the individual knowledge states of which are subsets of the corresponding individual observations. Now, by Definition 12, π^K must be an outcome of $\{\alpha_i^K\}_{i \in \mathsf{Agt}}$ in **G**^{K}. Since $\{\alpha_i^K\}_{i \in \mathsf{Agt}}$ is winning for R^K in **G**^{K}, π^K must be winning for R^K in **G**^{K}, and hence, by the definition of R^K, π must be winning for R in **G**. But π is arbitrary, and therefore $\{A_i(\alpha_i^K)\}_{i \in \mathsf{Agt}}$ must be winning for R in **G**. $\square$

The above results suggest a *method to synthesise observation-based finite-memory strategies* for reachability objectives R in a MAGIIAN **G**, based on:

(*i*) computing the MKBSC expansion **G**^{K} of the game,
(*ii*) searching for a winning profile of observation-based memoryless strategies (for the translated objective R^K) there, and if such a profile $\{\alpha_i^K\}_{i \in \mathsf{Agt}}$ is found,
(*iii*) translating the latter back in the form of the transducers $\{A_i(\alpha_i^K)\}_{i \in \mathsf{Agt}}$.

4.3.2. Translation to knowledge-based strategies

To be able to interpret the individual strategies α_i^K of the agents in **G**^{K} as individual knowledge-based strategies in **G**, following the framework outlined in Section 3.3, we need to define a knowledge representation and individual knowledge update functions. As a **first-order knowledge representation** structure we will use non-empty sets $A \subseteq \mathsf{Loc}$ of locations in **G**. For a given agent, the intended interpretation of such a set is as the agent's *most precise estimate of the actual location of the game* and, thus, represents the agent's exact uncertainty about the actual state-of-affairs. Given this knowledge representation, each memoryless strategy $\alpha_i^K : S_i \to \mathsf{Act}_i$ in $(\mathbf{G}|_i)^K$ can simultaneously be viewed as an individual **first-order knowledge-based strategy** for agent i in **G**.

Definition 14 (*Knowledge update*). For $s \in 2^{\mathsf{Loc}} \setminus \{\varnothing\}$, $\mathsf{act}_i \in \mathsf{Act}_i$ and $o_i \in \mathsf{Obs}_i$, the **knowledge update** function of agent $i \in \mathsf{Agt}$ is defined as follows:

$$\delta_i(s, \mathsf{act}_i, o_i) \stackrel{\text{def}}{=} \{l' \in o_i \mid \exists \mathsf{act} \in \mathsf{Act}. \, (\mathsf{act}(i) = \mathsf{act}_i \land \exists l \in s. \, ((l, \mathsf{act}, l') \in \Delta))\}$$

if the set is non-empty, and is undefined otherwise.

The set is the most precise estimate of agent i of the new actual location upon taking the action act_i and making the new observation o_i. Note that the update functions δ_i do *not* depend on α_i^K (which may be the case in the **(YN)** and **(YY)** cases discussed in Section 2.2). The **initial knowledge** of each agent i is $\{l_{init}\}$.

As the following result states, the first-order knowledge-based strategies defined in this way agree with the finite-memory ones from Definition 12.

Theorem 15 (*Strategy equivalence*). *Let* **G** *be a MAGIIAN,* $\mathbf{G}^K$ *be its MKBSC expansion,* $\left\{\alpha_i^K\right\}_{i \in \text{Agt}}$ *be a profile of observation-based memoryless strategies in* $\mathbf{G}^K$, *and* $\left\{A_i(\alpha_i^K)\right\}_{i \in \text{Agt}}$ *be the corresponding profile of induced transducers for* **G**. *Then, the strategy profile* $\left\{A_i(\alpha_i^K)\right\}_{i \in \text{Agt}}$, *and the profile of first-order knowledge-based strategies based on* $\left\{\alpha_i^K\right\}_{i \in \text{Agt}}$ *and* $\{\delta_i\}_{i \in \text{Agt}}$, *give rise to the same set of outcomes in* **G**.

Proof. We show that $\tau_i(s, o_i) = \delta_i(s, \alpha_i^K(s), o_i)$ for all $s \in 2^{\text{Loc}} \setminus \{\varnothing\}$ and $o_i \in \text{Obs}_i$. The result then follows from Definition 12, Definition 4, and the definition of first-order knowledge-based strategies.

Let $s \in 2^{\text{Loc}} \setminus \{\varnothing\}$ and $o_i \in \text{Obs}_i$. We have:

$$\tau_i(s, o_i)$$
$$= \text{the unique } s' \in S_i \text{ such that } s' \subseteq o_i \text{ and } (s, \alpha_i^K(s), s') \in \Delta_i^K \qquad \{\text{Definition 12}\}$$
$$= \left\{ l' \in o_i \mid \exists l \in s.\ (l, \alpha_i^K(s), l') \in \Delta_i \right\} \qquad \{\text{Definition 5.2}\}$$
$$= \left\{ l' \in o_i \mid \exists \text{act} \in \text{Act}.\ (\text{act}(i) = \alpha_i^K(s) \wedge \exists l \in s.\ (l, \text{act}, l') \in \Delta) \right\} \qquad \{\text{Definition 5.1}\}$$
$$= \delta_i(s, \alpha_i^K(s), o_i) \qquad \{\text{Definition 14}\}$$

if such an s' exists; otherwise, by Lemma 8, also $\delta_i(s, \alpha_i^K(s), o_i)$ is undefined. $\square$

As a corollary of Theorems 13 and 15, we have that:

(*i*) if $\left\{\alpha_i^K\right\}_{i \in \text{Agt}}$ with $\{\delta_i\}_{i \in \text{Agt}}$ is winning for R in **G**, and $\mathbf{G}^K$ fulfills the PDK condition, then $\left\{\alpha_i^K\right\}_{i \in \text{Agt}}$ is winning for R^K in $\mathbf{G}^K$, and

(*ii*) if $\left\{\alpha_i^K\right\}_{i \in \text{Agt}}$ is winning for R^K in $\mathbf{G}^K$, then $\left\{\alpha_i^K\right\}_{i \in \text{Agt}}$ with $\{\delta_i\}_{i \in \text{Agt}}$ is winning for R in **G**.

Every profile of first-order knowledge-based strategies in **G** is at the same time a profile of observation-based memoryless strategies in $\mathbf{G}^K$, and vice versa. Then, for a given MAGIIAN **G** and observable reachability objective R, if $\mathbf{G}^K$ fulfills the PDK condition, a winning profile of first-order knowledge-based strategies exists if and only if a winning profile of observation-based memoryless strategies exists in $\mathbf{G}^K$ for R^K.

Our strategy synthesis method is therefore *complete*, under the PDK condition, for the class of first-order knowledge-based strategies with respect to observable reachability objectives: if a winning profile of first-order knowledge-based strategies exists, it will be found with our method.

Example 16. Consider again our running Example 1, and note that the PDK condition holds for the MKBSC expansion of this game. Let the joint observable objective (in **G**) be to reach location good (and recall that it suffices for just one of the robots to observe this). Here is a winning profile of first-order knowledge-based strategies for play in **G**, where robot 0 follows the strategy defined in the left two columns of the table below, and updates its knowledge according to δ_0, partially shown in the right two columns:

Knowledge state	Action	On observing {bad}	On observing {good}
{start}	grab	{bad}	{good}
{bad}	squeeze	NA	{good}
{good}	squeeze	NA	{good}

while robot 1 follows the strategy defined in the left two columns of the table below, and updates its knowledge according to δ_1, partially shown in the right column:

Knowledge state	Action	On observing {good, bad}
{start}	grab	{good, bad}
{good, bad}	squeeze	{good}

If, however, the objective is to reach location win, then there is *no winning profile of first-order knowledge-based strategies*. Intuitively, the reason for this is that robot 0 is obliged to take the same action in both locations of $\mathbf{G}^K$ where it knows {good}, but win can only be reached by taking different actions. This problem will be resolved below with the help of second-order knowledge-based strategies.

The *duality* between the intensional and the extensional views exhibited above is not surprising. Having a knowledge-based strategy in the form of α_i^K and δ_i, agent i can reconstruct the transducer $A_i(\alpha_i^K)$, as evidenced by the proof of Theorem 15 (and one can thus view the execution of an individual knowledge-based strategy by an agent as constructing the corresponding pruned transducer *on-the-fly*). On the other hand, the reverse direction is far less obvious, and we see our approach as a way to provide an explanation for the otherwise uninterpreted strategies represented by transducers.

4.4. Strategy synthesis

In the beginning of Section 4.3.1 we noted that every observation-based memoryless strategy α_i^K for agent i in $\mathbf{G}^K$ is also a memoryless strategy in the single-agent game with perfect information $(\mathbf{G}|_i)^K$. This fact can also be useful for the *synthesis* of profiles of observation-based memoryless strategies in $\mathbf{G}^K$, since the synthesis of memoryless strategies in single-agent games of perfect information is well-studied. For instance, in the context of reachability objectives the standard synthesis technique is based on the notion of *controllable predecessors* (see e.g. [4]).

Another useful fact is that if a profile $\{\alpha_i^K\}_{i\in\text{Agt}}$ of observation-based memoryless strategies in $\mathbf{G}^K$ is winning for a (translated) reachability objective R^K, then, due to the consistency of the joint knowledge states of $\mathbf{G}^K$, every individual memoryless strategy α_i^K has an outcome $\pi_i = s_0 s_1 s_2 \dots$ in $(\mathbf{G}|_i)^K$ such that $\exists r \geq 0.\ \exists o \in \mathsf{R}.\ s_r \cap o \neq \varnothing$. This suggests the following *simple heuristic* for synthesis of profiles of observation-based memoryless strategies in $\mathbf{G}^K$:

1. For each agent $i \in \text{Agt}$, find a memoryless strategy α_i^K in $(\mathbf{G}|_i)^K$ that has a winning outcome for the reachability objective $\mathsf{R}_i^K = \{s \in S_i \mid \exists o \in \mathsf{R}.\ s \cap o \neq \varnothing\}$.
2. Check whether the profile $\{\alpha_i^K\}_{i\in\text{Agt}}$ is winning for R^K. If it is not, backtrack to step 1.

To implement the first step of the heuristic, one can adapt the notion of controllable predecessors to finding strategies where *some* outcome is winning (rather than *all* outcomes, as the standard formulation achieves). For the second step, it suffices to simulate the profile $\{\alpha_i^K\}_{i\in\text{Agt}}$ on the game $\mathbf{G}^K$, following each outcome up to the first knowledge state which either belongs to R^K, or else has already been visited by the outcome. In the latter case, the profile is not winning. Since the set of knowledge states is finite, this check terminates.

5. The iterated MKBSC construction

Applying the MKBSC to a MAGIIAN $\mathbf{G}$ does not necessarily result in a game with perfect information, but in general produces another MAGIIAN. Thus, as it was first observed in [14] by the third co-author, the construction can always be applied again, iteratively, producing an infinite (but possibly collapsing) hierarchy of expansions $\mathbf{G}^K$, $\mathbf{G}^{2K}$, $\mathbf{G}^{3K}$, ... We show here that such repeated application produces a hierarchy of higher-order knowledge representation structures for the agents in the team, and using these can increase its strategic abilities.

Example 17. We apply below the MKBSC construction on the game $\mathbf{G}^K$ from Fig. 4 to produce $\mathbf{G}^{2K}$. The individual projections $\mathbf{G}^K|_0$ and $\mathbf{G}^K|_1$ are shown in Fig. 6, with corresponding expansions $(\mathbf{G}^K|_0)^K$ and $(\mathbf{G}^K|_1)^K$, as shown in Fig. 7. The pruned product $\mathbf{G}^{2K}$ is shown in Fig. 8.

Now, for the robot team to reach location win in $\mathbf{G}$, robot 0 can follow the second-order knowledge-based strategy (extracted from the above games as described below):

Knowledge state	Action
$\{(\{\text{start}\}, \{\text{start}\})\}$	grab
$\{(\{\text{good}\}, \{\text{bad}, \text{good}\})\}$	squeeze
$\{(\{\text{bad}\}, \{\text{bad}, \text{good}\})\}$	squeeze
$\{(\{\text{good}\}, \{\text{good}\})\}$	lift

while the strategy of robot 1 follows the table:

Knowledge state	Action
$\{(\{\text{start}\}, \{\text{start}\})\}$	grab
$\{(\{\text{bad}\}, \{\text{bad}, \text{good}\}), (\{\text{good}\}, \{\text{bad}, \text{good}\})\}$	squeeze
$\{(\{\text{good}\}, \{\text{good}\})\}$	lift

This means, for instance, that robot 0 will squeeze whenever it *knows* that robot 1 (the one without a grip sensor) is uncertain about whether the grip is good or not.

It is well-known that coordinated action in multi-agent games requires *common knowledge* (see, e.g., [26]). In the context of the above example, it is worth noting that the knowledge state:
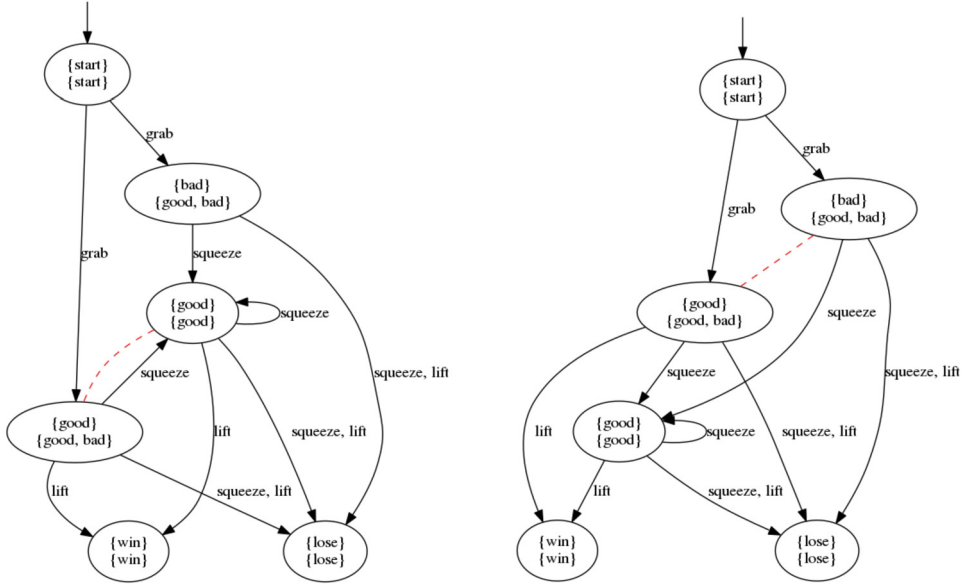
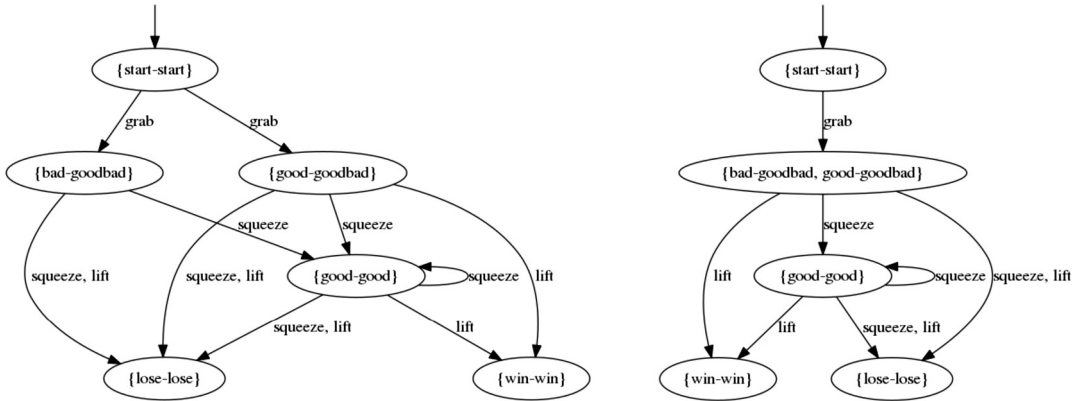**Fig. 6.** The individual projections $\mathbf{G}^K|_0$ and $\mathbf{G}^K|_1$.

Fig. 7. The individual expansions $(\mathbf{G}^K|_0)^K$ and $(\mathbf{G}^K|_1)^K$.

$$(\{(\{good\}, \{good\})\}, \{((\{good\}, \{good\})\})$$

of Fig. 8, while representing second-order knowledge of the two robots, can also be seen as representing the common knowledge of the robots that they both have a good grip. This, in fact, is what justifies that they can simultaneously lift at this point. The notion of common knowledge in the context of MAGIIAN games shall not be explored further in this paper.

5.1. Generalised induced transducers

Now, we discuss the general case of *iterating* the MKBSC construction j times, for any $j \geq 1$, resulting in the expanded game structure $\mathbf{G}^{jK}$, and generalise our results from the preceding section. Let $\mathbf{G}^{0K}$ denote the original game $\mathbf{G}$.

Intuitively, the single-agent game with perfect information $(\mathbf{G}^{jK}|_i)^K$ represents the possible "dynamics" of agent i's $(j+1)$-order knowledge. Similarly to the construction presented in Definition 12, one can combine each $(\mathbf{G}^{jK}|_i)^K$ and individual memoryless strategy $\alpha_i^{(j+1)K}$ into a transducer $A_i(\alpha_i^{(j+1)K})$ for agent i for play in $\mathbf{G}$. To achieve this, however, we first need to formally connect, for every agent $i \in \mathsf{Agt}$, the knowledge states of $(\mathbf{G}^{jK}|_i)^K$ to the observations Obs_i of that agent in $\mathbf{G}$. This can be achieved by observing that each knowledge state $s_i \in S_i$ of $(\mathbf{G}^{jK}|_i)^K$ is a set of locations in $\mathbf{G}^{jK}$, which, when $j > 0$, are tuples that agree on their i-th component, where this i-th component is in turn a knowledge state in $(\mathbf{G}^{(j-1)K}|_i)^K$. We can thus repeat this process until reaching a set of locations in $\mathbf{G}$. Let us denote this set by $\hat{s}_i \subseteq \mathsf{Loc}$. By virtue of the MKBSC construction, $\hat{s}_i$ will be non-empty, and will be a subset of some observation $o_i \in \mathsf{Obs}_i$ in the original game $\mathbf{G}$.
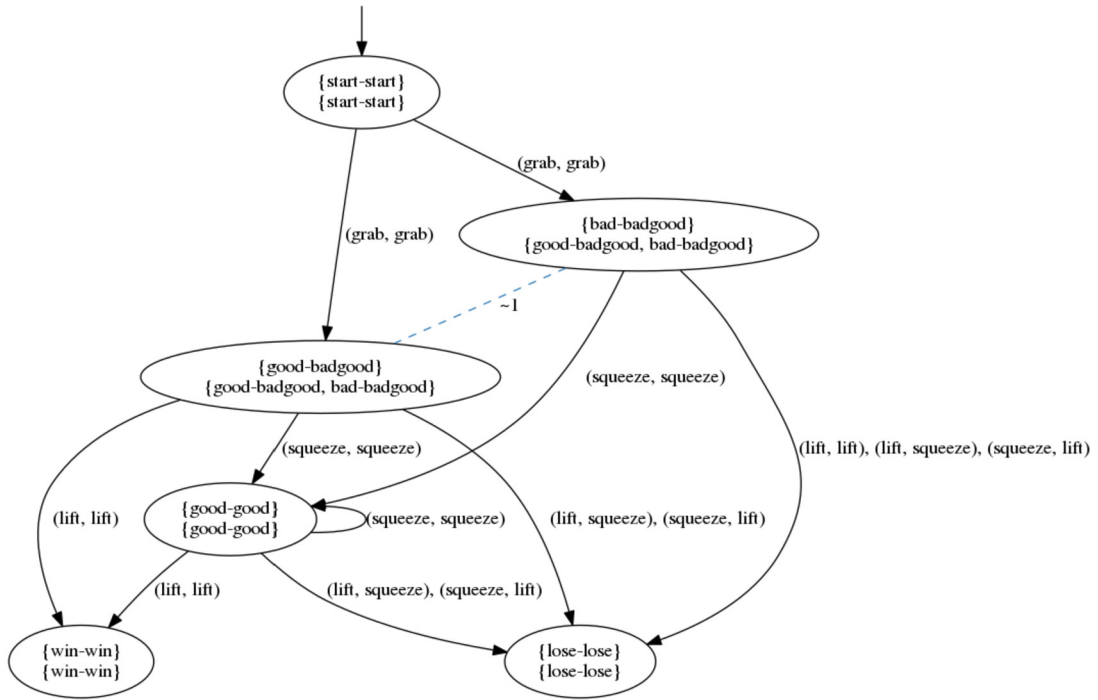
Fig. 8. The pruned product $\mathbf{G}^{2K}$.

For the expanded game structures $\mathbf{G}^{jK}$, the connection between the joint knowledge states s in the latter and the locations in $\mathbf{G}$ is established via iterated intersection of the sets comprising the tuples in s until obtaining a set of locations in $\mathbf{G}$. Let us denote this set by $\Cap s \subseteq$ Loc. Iterated intersection is well-defined by virtue of Lemma 10.

The following result generalises Lemma 8.

Lemma 18. *Let* $\mathbf{G} =$ (Agt, Loc, l_{init}, Act, Δ, Obs) *be a MAGIIAN for a set of agents* Agt, *let* $j \geq 1$, *and let* $\mathbf{G}^{jK} =$ (Agt, S^{jK}, s_I, Act, Δ^{jK}, Obs^{jK}) *be the j-iterated MKBSC expansion of* $\mathbf{G}$. *Further, let* $s \in S^{jK}$, act $\in$ Act *and* $o \in$ Obs^p. *Define the set:*

$$X^{(j)} \stackrel{\text{def}}{=} \left\{ l' \in \cap_{i \in \text{Agt}} o(i) \mid \exists l \in \Cap s. \, (l, \text{act}, l') \in \Delta \right\}$$

Then, $X^{(j)}$ is non-empty if and only if there is $s' \in S^{jK}$ such that $(s, \text{act}, s') \in \Delta^{jK}$ and such that for all $i \in$ Agt, $\hat{s}'(i) \subseteq o(i)$. This $s' \in S^{jK}$ is then unique and we have $X^{(j)} \subseteq \Cap s'$, and if $\mathbf{G}^K$ fulfills the PDK condition, we have $X^{(j)} = \Cap s'$.

Proof. By mathematical induction on j. The base case of $j = 1$ is established by Lemma 8. Assume (as the induction hypothesis) that the result holds for j. We show that the result then follows for $j + 1$.

Let $s \in S^{(j+1)K}$, act $\in$ Act and $o \in$ Obs^p. Further, let $X^{(j+1)}$ be defined as above. Consider the case when $X^{(j+1)}$ is non-empty. (The case when $X^{(j+1)}$ is empty is analogous.) Then, there must be locations $l, l' \in$ Loc in $\mathbf{G}$ such that $l \in \Cap s$, $l' \in \cap_{i \in \text{Agt}} o(i)$ and $(l, \text{act}, l') \in \Delta$. Let s_1 denote the sole element of the (simple) intersection of the sets comprising the tuple s (by Lemma 10, this intersection must be a singleton set). Then $s_1 \in S^{jK}$ and $l \in \Cap s_1$ must be the case. Let $X^{(j)}$ be defined as above (but w.r.t. $s_1 \in S^{jK}$). By the induction hypothesis, $X^{(j)}$ must also be non-empty, and hence, there is exactly one $s_1' \in S^{jK}$ such that $(s_1, \text{act}, s_1') \in \Delta^{jK}$ and for all $i \in$ Agt, $\hat{s_1}'(i) \subseteq o(i)$. Then, $l' \in \Cap s_1'$ must be the case.

Now, let $o_1 \in$ Obs^{jK} be the unique observation profile in $\mathbf{G}^{jK}$ that the agents make in s_1'. By Lemma 8, there is exactly one $s' \in S^{(j+1)K}$ such that $(s, \text{act}, s') \in \Delta^{(j+1)K}$ and for all $i \in$ Agt, $s'(i) \subseteq o_1(i)$. Then, for all $i \in$ Agt, $\hat{s}'(i) \subseteq o(i)$ must be the case. Furthermore, by Lemma 8 and Lemma 10, s_1' must be the sole element of the intersection of the sets comprising the tuple s', and thus, since $l' \in \Cap s_1'$, also $l' \in \Cap s'$ must be the case. This establishes $X^{(j+1)} \subseteq \Cap s'$.

Finally, if $\mathbf{G}^K$ fulfills the PDK condition, because of Lemma 10 all iterated intersections used in the proof must be singletons, and thus $X^{(j+1)} = \Cap s'$. $\quad\square$

As before, this result lifts naturally to observation histories: every sequence π of joint actions and joint observations of $\mathbf{G}$ (where π is not necessarily a path in $\mathbf{G}$) gives rise to at most one path in $\mathbf{G}^{jK}$ such that at each corresponding step of the two sequences, $\hat{s}(i) \subseteq o(i)$ holds for all $i \in$ Agt. Furthermore, every full play π in $\mathbf{G}$ gives rise to exactly one full play in $\mathbf{G}^{jK}$ that is consistent with the actions and the observations of the agents.

Definition 19 *(Generalised induced transducer).* Let $\mathbf{G} = (\mathsf{Agt}, \mathsf{Loc}, l_{\mathsf{init}}, \mathsf{Act}, \Delta, \mathsf{Obs})$ and for any $\mathsf{i} \in \mathsf{Agt}$ and $j \geq 0$, let $(\mathbf{G}^{j\mathsf{K}}|_{\mathsf{i}})^{\mathsf{K}} = (S_{\mathsf{i}}, s_{I,\mathsf{i}}, \mathsf{Act}_{\mathsf{i}}, \Delta_{\mathsf{i}}^{(j+1)\mathsf{K}})$. Let also $\alpha_{\mathsf{i}}^{(j+1)\mathsf{K}} : S_{\mathsf{i}} \to \mathsf{Act}_{\mathsf{i}}$ be a memoryless strategy in $(\mathbf{G}^{j\mathsf{K}}|_{\mathsf{i}})^{\mathsf{K}}$. We define the $\alpha_{\mathsf{i}}^{(j+1)\mathsf{K}}$-**induced transducer** as:

$$A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{(j+1)\mathsf{K}}) \stackrel{\mathrm{def}}{=} (S_{\mathsf{i}}, s_{I,\mathsf{i}}, \mathsf{Obs}_{\mathsf{i}}, \mathsf{Act}_{\mathsf{i}}, \tau_{\mathsf{i}}^{j+1}, \alpha_{\mathsf{i}}^{(j+1)\mathsf{K}})$$

where $\tau_{\mathsf{i}}^{j+1}(s, o_{\mathsf{i}})$ is defined for $s \in S_{\mathsf{i}}$ and $o_{\mathsf{i}} \in \mathsf{Obs}_{\mathsf{i}}$ as the unique $s' \in S_{\mathsf{i}}$ such that $\hat{s}' \subseteq o_{\mathsf{i}}$ and $(s, \alpha_{\mathsf{i}}^{(j+1)\mathsf{K}}(s), s') \in \Delta_{\mathsf{i}}^{(j+1)\mathsf{K}}$, if such an s' exists, and is undefined otherwise.

The transducer $A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{(j+1)\mathsf{K}})$ is, by Definition 4, an observation-based finite-memory strategy for agent i in $\mathbf{G}$. The transducer can be *pruned* by removing, from each memory state s, the outgoing edges for actions other than $\alpha_{\mathsf{i}}^{(j+1)\mathsf{K}}(s)$, then by removing the unreachable memory states, and finally by abstracting away the structure of s (since only the identity of the memory states is relevant).

The following result generalises Theorem 13 to the iterated MKBSC.

Theorem 20 *(Generalised strategy correspondence).* Let $\mathbf{G}$ be a MAGIIAN for a set of agents Agt, let $j \geq 1$, and let $\mathbf{G}^{j\mathsf{K}}$ be the j-iterated MKBSC expansion of $\mathbf{G}$. Let R be an observable reachability objective in $\mathbf{G}$, and $\mathsf{R}^{j\mathsf{K}}$ be its j-iterated translation in $\mathbf{G}^{j\mathsf{K}}$. Finally, let $\left\{\alpha_{\mathsf{i}}^{j\mathsf{K}}\right\}_{\mathsf{i} \in \mathsf{Agt}}$ be a profile of observation-based memoryless strategies in $\mathbf{G}^{j\mathsf{K}}$, and $\left\{A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{j\mathsf{K}})\right\}_{\mathsf{i} \in \mathsf{Agt}}$ be the corresponding profile of generalised induced transducers for $\mathbf{G}$.

(i) If $\left\{A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{j\mathsf{K}})\right\}_{\mathsf{i} \in \mathsf{Agt}}$ is winning for R in $\mathbf{G}$, and $\mathbf{G}^{\mathsf{K}}$ fulfills the PDK condition, then $\left\{\alpha_{\mathsf{i}}^{j\mathsf{K}}\right\}_{\mathsf{i} \in \mathsf{Agt}}$ is winning for $\mathsf{R}^{j\mathsf{K}}$ in $\mathbf{G}^{j\mathsf{K}}$.

(ii) If $\left\{\alpha_{\mathsf{i}}^{j\mathsf{K}}\right\}_{\mathsf{i} \in \mathsf{Agt}}$ is winning for $\mathsf{R}^{j\mathsf{K}}$ in $\mathbf{G}^{j\mathsf{K}}$, then $\left\{A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{j\mathsf{K}})\right\}_{\mathsf{i} \in \mathsf{Agt}}$ is winning for R in $\mathbf{G}$.

Proof. The proof adapts the one of Theorem 13, but refers now to Lemma 18 to relate the plays in $\mathbf{G}$ with those in $\mathbf{G}^{j\mathsf{K}}$.

(i) Let $\left\{A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{j\mathsf{K}})\right\}_{\mathsf{i} \in \mathsf{Agt}}$ be winning for R in $\mathbf{G}$, and let $\mathbf{G}^{\mathsf{K}}$ fulfill the PDK condition. Let $\pi^{j\mathsf{K}} = s_0 \sigma_0 s_1 \sigma_1 \ldots$ be an arbitrary outcome of $\left\{\alpha_{\mathsf{i}}^{j\mathsf{K}}\right\}_{\mathsf{i} \in \mathsf{Agt}}$ in $\mathbf{G}^{j\mathsf{K}}$. Then, by Lemma 18 and Lemma 10, and since G^{K} fulfills the PDK condition, the sequence $\pi = l_0 \sigma_0 l_1 \sigma_1 \ldots$ such that $\{l_k\} = \Cap s_k$ for all $k \geq 0$, must be a full play in $\mathbf{G}$. Now, by Definition 19, π must be an outcome of $\left\{A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{j\mathsf{K}})\right\}_{\mathsf{i} \in \mathsf{Agt}}$ in $\mathbf{G}$. Since $\left\{A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{j\mathsf{K}})\right\}_{\mathsf{i} \in \mathsf{Agt}}$ is winning for R in $\mathbf{G}$, π must be winning for R in $\mathbf{G}$, and hence, by the definition of $\mathsf{R}^{j\mathsf{K}}$, $\pi^{j\mathsf{K}}$ must be winning for $\mathsf{R}^{j\mathsf{K}}$ in $\mathbf{G}^{j\mathsf{K}}$. But $\pi^{j\mathsf{K}}$ is arbitrary, and therefore $\left\{\alpha_{\mathsf{i}}^{j\mathsf{K}}\right\}_{\mathsf{i} \in \mathsf{Agt}}$ must be winning for $\mathsf{R}^{j\mathsf{K}}$ in $\mathbf{G}^{j\mathsf{K}}$.

(ii) Let $\left\{\alpha_{\mathsf{i}}^{j\mathsf{K}}\right\}_{\mathsf{i} \in \mathsf{Agt}}$ be winning for $\mathsf{R}^{j\mathsf{K}}$ in $\mathbf{G}^{j\mathsf{K}}$. Let $\pi = l_0 l_1 l_2 \ldots$ be an arbitrary outcome of $\left\{A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{j\mathsf{K}})\right\}_{\mathsf{i} \in \mathsf{Agt}}$ in $\mathbf{G}$. This outcome induces a corresponding sequence of joint observations, from which, using $\left\{A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{j\mathsf{K}})\right\}_{\mathsf{i} \in \mathsf{Agt}}$, one can recover the corresponding sequence of joint actions. By Lemma 18, these two sequences (of joint actions and joint observations) give rise to a unique play $\pi^{j\mathsf{K}} = s_0 \sigma_0 s_1 \sigma_1 \ldots$ in $\mathbf{G}^{j\mathsf{K}}$ such that $\{l_k\} = \Cap s_k$ for all $k \geq 0$. Now, by Definition 19, $\pi^{j\mathsf{K}}$ must be an outcome of $\left\{\alpha_{\mathsf{i}}^{j\mathsf{K}}\right\}_{\mathsf{i} \in \mathsf{Agt}}$ in $\mathbf{G}^{j\mathsf{K}}$. Since $\left\{\alpha_{\mathsf{i}}^{j\mathsf{K}}\right\}_{\mathsf{i} \in \mathsf{Agt}}$ is winning for $\mathsf{R}^{j\mathsf{K}}$ in $\mathbf{G}^{j\mathsf{K}}$, $\pi^{j\mathsf{K}}$ must be winning for $\mathsf{R}^{j\mathsf{K}}$ in $\mathbf{G}^{j\mathsf{K}}$, and hence, by the definition of $\mathsf{R}^{j\mathsf{K}}$, π must be winning for R in $\mathbf{G}$. But π is arbitrary, and therefore $\left\{A_{\mathsf{i}}(\alpha_{\mathsf{i}}^{j\mathsf{K}})\right\}_{\mathsf{i} \in \mathsf{Agt}}$ must be winning for R in $\mathbf{G}$. $\square$

5.2. Generalised knowledge representation

Let us denote by $A^{(j+1)}$ the domain of knowledge states that can arise in the game structures $(\mathbf{G}^{j\mathsf{K}}|_{\mathsf{i}})^{\mathsf{K}}$, and by $B^{(j)}$ the ones that can arise in $\mathbf{G}^{j\mathsf{K}}$. The elements of $A^{(j+1)}$ are non-empty sets of elements from $B^{(j)}$, and the sets $A^{(j)}$ can thus be defined formally, for $j > 0$, by:

$$A^{(j+1)} \stackrel{\mathrm{def}}{=} 2^{B^{(j)}}$$

The elements of $B^{(j)}$ are tuples of elements of $A^{(j)}$, one for each agent $\mathsf{i} \in \mathsf{Agt}$, and thus, the sets $B^{(j)}$ can be defined inductively as follows:

$$B^{(0)} \stackrel{\mathrm{def}}{=} L$$

$$B^{(j)} \stackrel{\mathrm{def}}{=} (A^{(j)})^{\mathsf{Agt}} \qquad \text{for } j > 0$$
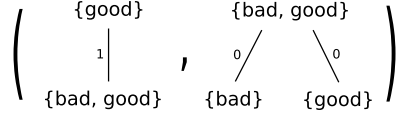
So, as a structure for representing the *j*-**order knowledge** of agents we will use elements of $A^{(j)}$.

As already mentioned, the elements of $A^{(j+1)}$ that can actually arise from the iterated MKBSC have the property that they are sets of tuples from $(A^{(j)})^{\mathsf{Agt}}$ that agree on their i-th component. This observation suggests that there are more compact and meaningful representations of the knowledge structures, as we will now show.

Knowledge trees For teams of two agents, the knowledge states in the MKBSC-iterated games can be represented as pairs of **knowledge trees**, one for each agent, of depth being the iteration index. For example, consider the following knowledge state in $\mathbf{G}^{2K}$ from our running example:

$$(\{(\{\mathsf{good}\}, \{\mathsf{bad}, \mathsf{good}\})\}, \{(\{\mathsf{bad}\}, \{\mathsf{bad}, \mathsf{good}\}), (\{\mathsf{good}\}, \{\mathsf{bad}, \mathsf{good}\})\})$$

By factoring out the common i'th component for each agent i, this structure can be equivalently represented as a pair of trees:

$$\left(\begin{array}{c} \{\mathsf{good}\} \\ {\scriptstyle 1} \mid \\ \{\mathsf{bad}, \mathsf{good}\} \end{array} \quad , \quad \begin{array}{c} \{\mathsf{bad}, \mathsf{good}\} \\ {\scriptstyle 0}\diagup \quad \diagdown {\scriptstyle 0} \\ \{\mathsf{bad}\} \quad \{\mathsf{good}\} \end{array} \right)$$

These trees represent the *second-order knowledge* of the two robots: robot 0 knows that the grip is good, but that robot 1 is uncertain about whether the grip is good or bad, while robot 1 is uncertain about whether the grip is good or bad, but knows that robot 0 knows which of the two is the case.

A visualisation of $\mathbf{G}^{2K}$ for our running example, with states represented by the respective pairs of knowledge trees is shown on Fig. 9. It has been obtained by a modified version of our tool.[13] Knowledge trees can have any finite depth. For instance, a knowledge tree of robot 1 from $\mathbf{G}^{5K}$ is depicted on Fig. 10.

Representing knowledge in the form of trees, rather than as recursive tuples of sets of locations, can play an important role in *explaining to humans* knowledge-based strategies that have been synthesised algorithmically with our method.[14]

Our notion of knowledge trees is akin to the notion of k-trees of [8]. The latter notion is finer than ours in that, in a k-tree, every node (i.e., set) is connected to a particular element of the parent node (set) rather than with the whole set (as it is the case in our representation). Some details on k-trees can be found in Appendix A, where we have slightly modified the original presentation, adapting it to the current set-up. Technically and conceptually similar (yet, not equivalent) are the k-worlds in [7].

Lastly, analogous structures have also been used in decision and game theory, in the context of multi-player decision making, to represent the knowledge of players who use mental models of the other players when deciding how to act, see e.g. the "level-k types" in [25].

5.3. Generalised knowledge update

For $j \geq 0$, taking $A^{(j+1)}$ as the $(j+1)$-order knowledge representation, every individual memoryless strategy $\alpha_i^{(j+1)K}$ in $(\mathbf{G}^{jK}|_i)^K$ can be seen as an individual $(j+1)$-order knowledge-based strategy for agent i in $\mathbf{G}$.

Definition 21 (*Generalised knowledge update*). For $j > 0$, $s \in A^{(j+1)}$, $\mathsf{act}_i \in \mathsf{Act}_i$ and $o_i \in \mathsf{Obs}_i$, the **generalised knowledge update** function of agent $i \in \mathsf{Agt}$ is defined inductively as follows:

$$\delta_i^{j+1}(s, \mathsf{act}_i, o_i) \overset{\mathrm{def}}{=} \left\{ \delta_{i'}^{j}(t(i'), \mathsf{act}(i'), o(i')))_{i' \in \mathsf{Agt}} \,\middle|\, t \in s, \mathsf{act} \in \mathsf{Act}, o \in \mathsf{Obs}^p, \mathsf{act}(i) = \mathsf{act}_i, o(i) = o_i \right\}$$

from which the inconsistent tuples are pruned out (as in the Composition step of Definition 5), and where the base case $\delta_i^1 \overset{\mathrm{def}}{=} \delta_i$ is as given in Definition 14.

As the following result shows, generalised knowledge update δ_i^{j+1} is essentially knowledge update δ_i applied on $\mathbf{G}^{jK}$ (instead of on $\mathbf{G}$). To formalise this, one has to realise that agent observations in the original game $\mathbf{G}$ relate to observations in the expanded games much in the same way as the locations relate to knowledge states (see page 18). For an observation $o_i \in \mathsf{Obs}_i$, let o_i^j denote the set $\left\{ s \in B^{(j)} \,\middle|\, \hat{s}(i) \subseteq o_i \right\}$.

Lemma 22. *Let $j \geq 1$, $s \in A^{(j+1)}$, $\mathsf{act}_i \in \mathsf{Act}_i$ and $o_i \in \mathsf{Obs}_i$. Then:*

$$\delta_i^{j+1}(s, \mathsf{act}_i, o_i) = \left\{ l' \in o_i^j \,\middle|\, \exists \mathsf{act} \in \mathsf{Act}. (\mathsf{act}(i) = \mathsf{act}_i \wedge \exists l \in s. (l, \mathsf{act}, l') \in \Delta^{jK}) \right\}$$

---

[13] Described in [24] and available from github.com/larasik/mkbsc.
[14] At the same time it should be noted that such trees are more efficiently represented by DAGs.

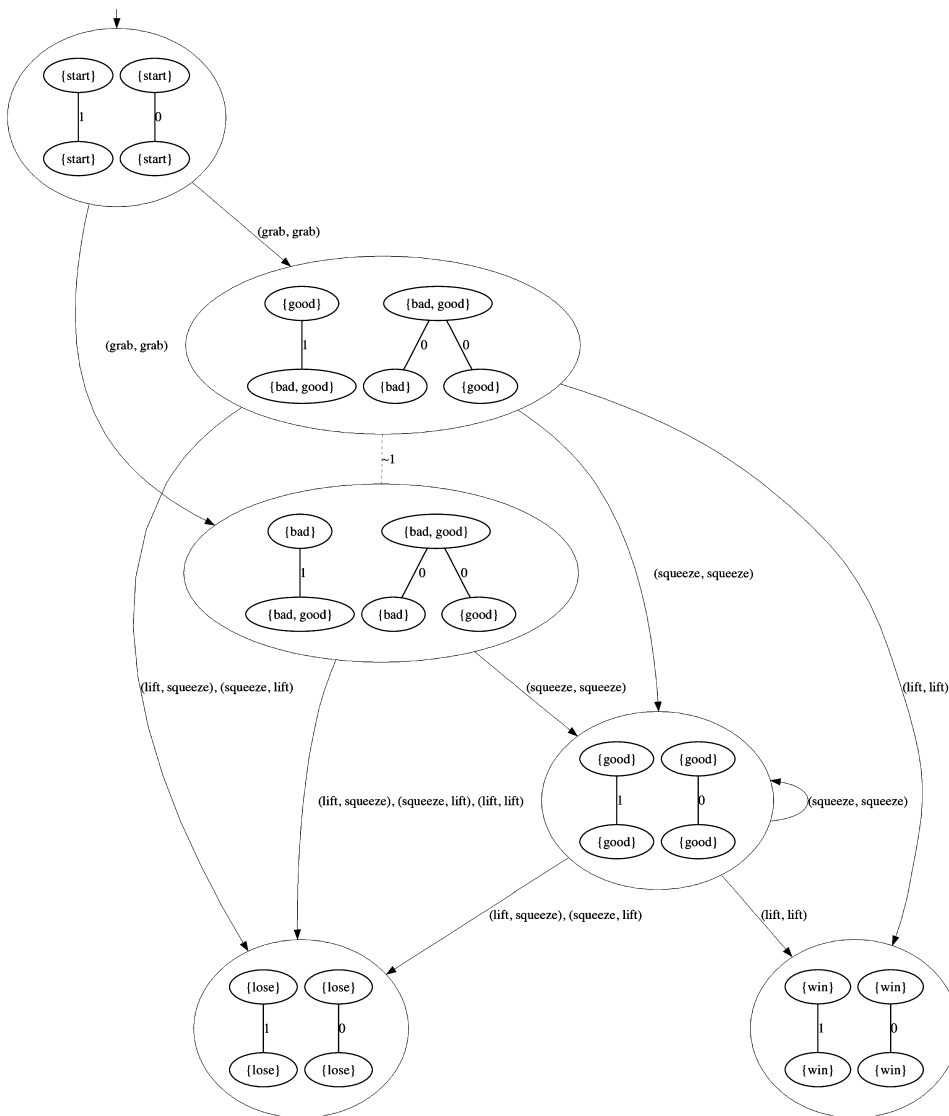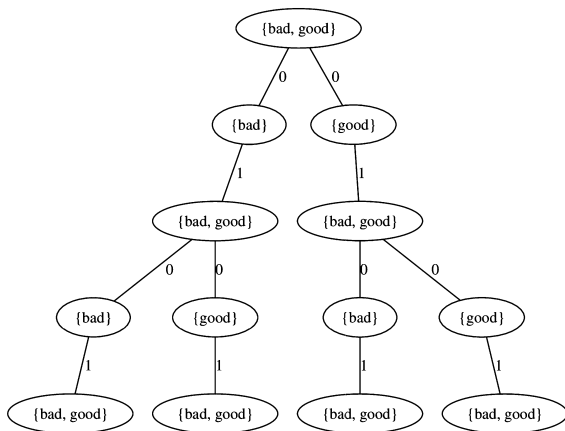**Fig. 9.** The game $\mathbf{G}^{2K}$ with explicit knowledge trees.

Fig. 10. A knowledge tree from $\mathbf{G}^{5K}$.

Proof. By mathematical induction on j. The base case of $j = 1$ is established in the proof of Theorem 15. Assume (as the induction hypothesis) that the result holds for j. We show that the result then follows for $j + 1$.

Let $s \in A^{(j+1)}$, $\mathsf{act}_i \in \mathsf{Act}_i$, $o_i \in \mathsf{Obs}_i$, and let o_i^j be as defined above. Then:

$$
\begin{aligned}
& \delta_i^{j+1}(s, \mathsf{act}_i, o_i) \\
&= \left\{ (\delta_{i'}^j(t(i'), \mathsf{act}(i'), o(i')))_{i' \in \mathsf{Agt}} \,\middle|\, t \in s, \mathsf{act} \in \mathsf{Act}, o \in \mathsf{Obs}^p, \mathsf{act}(i) = \mathsf{act}_i, o(i) = o_i \right\} && \{\text{Definition } 21\} \\
&= \left\{ (\{l' \in o_{i'}^{j-1} \mid \exists \mathsf{act} \in \mathsf{Act}. (\mathsf{act}(i') = \mathsf{act}_{i'} \wedge \exists l \in t(i'). (l, \mathsf{act}, l') \in \Delta^{(j-1)\mathsf{K}})\})_{i' \in \mathsf{Agt}} \right. \\
&\qquad \left. \middle|\, t \in s, \mathsf{act} \in \mathsf{Act}, o \in \mathsf{Obs}^p, \mathsf{act}(i) = \mathsf{act}_i, o(i) = o_i \right\} && \{\text{Ind.hyp.}\} \\
&= \left\{ l' \in o_i^j \,\middle|\, \exists \mathsf{act} \in \mathsf{Act}. (\mathsf{act}(i) = \mathsf{act}_i \wedge \exists l \in s. (l, \mathsf{act}, l') \in \Delta^{j\mathsf{K}}) \right\} && \{\text{Definition } 5\} \quad \square
\end{aligned}
$$

The following result generalises Theorem 15 to the iterated MKBSC: the strategy profiles resulting from the translation to transducers and to knowledge-based strategies are equivalent, i.e., give rise to the same sets of outcomes.

Theorem 23 (Generalised strategy equivalence). *Let $\mathbf{G}$ be a MAGIIAN for a set of agents Agt, let $j \geq 1$, and let $\mathbf{G}^{j\mathsf{K}}$ be the j-iterated MKBSC expansion of $\mathbf{G}$. Let $\left\{ \alpha_i^{j\mathsf{K}} \right\}_{i \in \mathsf{Agt}}$ be a profile of observation-based memoryless strategies in $\mathbf{G}^{j\mathsf{K}}$, and $\left\{ A_i(\alpha_i^{j\mathsf{K}}) \right\}_{i \in \mathsf{Agt}}$ be the corresponding profile of induced transducers for $\mathbf{G}$. Then, the strategy profile $\left\{ A_i(\alpha_i^{j\mathsf{K}}) \right\}_{i \in \mathsf{Agt}}$, and the profile of j-order knowledge-based strategies based on $\left\{ \alpha_i^{j\mathsf{K}} \right\}_{i \in \mathsf{Agt}}$ and $\left\{ \delta_i^j \right\}_{i \in \mathsf{Agt}}$, give rise to the same set of outcomes in $\mathbf{G}$.*

Proof. We show that $\tau_i^j(s, o_i) = \delta_i^j(s, \alpha_i^{j\mathsf{K}}(s), o_i)$ for all $j \geq 1$, $s \in A^{(j+1)}$ and $o_i \in \mathsf{Obs}_i$. The result then follows from Definition 19, Definition 4, and the definition of j-order knowledge-based strategies.

The case when $j = 1$ is established by Theorem 15. Let $j \geq 1$, $s \in A^{(j+1)}$, $\mathsf{act}_i \in \mathsf{Act}_i$, $o_i \in \mathsf{Obs}_i$, and let o_i^j be as defined above. Then:

$$
\begin{aligned}
& \tau_i^{j+1}(s, o_i) \\
&= \text{the unique } s' \in S_i \text{ such that } \hat{s}' \subseteq o_i \text{ and } (s, \alpha_i^{(j+1)\mathsf{K}}(s), s') \in \Delta_i^{(j+1)\mathsf{K}} && \{\text{Definition } 19\} \\
&= \left\{ l' \in o_i^j \,\middle|\, \exists l \in s. (l, \alpha_i^{(j+1)\mathsf{K}}(s), l') \in \Delta_i^{j\mathsf{K}} \right\} && \{\text{Definition } 5.2\} \\
&= \left\{ l' \in o_i^j \,\middle|\, \exists \mathsf{act} \in \mathsf{Act}. (\mathsf{act}(i) = \alpha_i^{(j+1)\mathsf{K}}(s) \wedge \exists l \in s. (l, \mathsf{act}, l') \in \Delta^{j\mathsf{K}}) \right\} && \{\text{Definition } 5.1\} \\
&= \delta_i^{j+1}(s, \alpha_i^{(j+1)\mathsf{K}}(s), o_i) && \{\text{Lemma } 22\}
\end{aligned}
$$

if such an s' exists; otherwise, by Lemma 18, $\delta_i^{j+1}(s, \alpha_i^{(j+1)\mathsf{K}}(s), o_i)$ is undefined. $\square$

As we showed in Lemma 10, for all $j > 1$, the expansions $\mathbf{G}^{j\mathsf{K}}$ satisfy the PDK condition. By composing our results we obtain that, for all $j > 1$, our strategy synthesis method is *complete* for the class of j-order knowledge-based strategies with respect to observable reachability objectives whenever $\mathbf{G}^{\mathsf{K}}$ satisfies the PDK condition: if a winning profile of j-order knowledge-based strategies exists, it will be found with our method.

Further, since the observation-based memoryless strategies are preserved by the MKBSC (see second paragraph following the proof of Theorem 9), the sets of strategies produced by the iterative approaches above grow monotonically: the class of j-order knowledge-based strategies subsumes any lower-order class. In other words, *increasing the order of knowledge increases the strategic ability of the team*.

6. Follow-up discussion

In this section we discuss some aspects of the MKBSC that we consider of importance, and which will be explored in follow-up work.

6.1. Utilising the iterated MKBSC

There are two dual approaches to using the constructions presented here for synthesising knowledge-based strategies:

- **Global:** Keep applying iteratively the MKBSC on the game graph, and then search for an observation-based memoryless strategy profile in the resulting expanded game until – if ever – such strategy profile is found. Then convert it to a knowledge-based strategy profile.
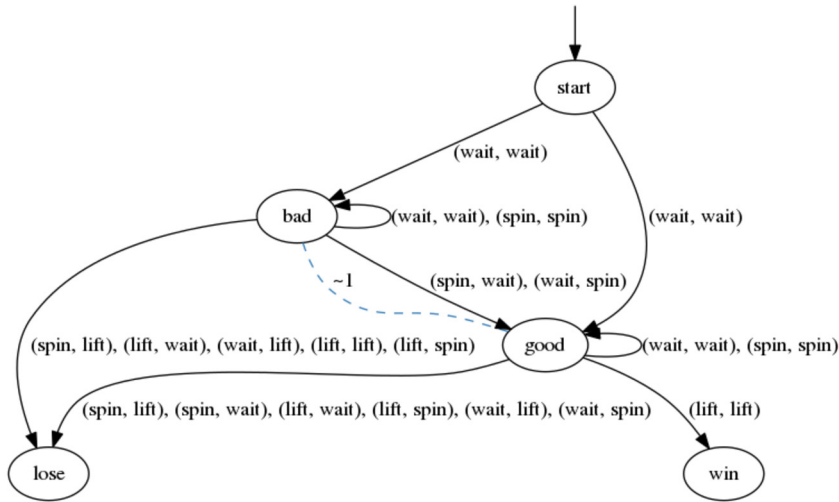
Fig. 11. A stable game.

- **Local:** Keep incrementing the epistemic depth, and then explore the reachable part of the game graph produced (at the current depth) by means of the knowledge update functions. Then, search for a memoryless strategy in that graph (which, if found, will by construction be a knowledge-based one).

As indicated earlier, these two approaches are equivalent.

6.2. Stabilisation of the iterated MKBSC

For some MAGIIAN games the iterated construction eventually stabilises, in the sense that it starts producing isomorphic games, even though the internal structure of the locations (the tuples of knowledge trees) grows unboundedly. For instance, for the game graph of our running example, $\mathbf{G}^{3K}$ is isomorphic to $\mathbf{G}^{2K}$. We then say that game graph $\mathbf{G}^{2K}$ is *stable*.

Thus, the global approach outlined above can be augmented with a check for stabilisation. Furthermore, as explained in the end of Section 5, if we know that the construction will stabilise, we may directly iterate until that point and only search for memoryless strategies in the stable game (since we are guaranteed to not miss any such strategy that would have been found in some of the intermediate games).

This leads to a number of interesting questions about stabilisation and the properties of stable game graphs:

- To begin with, what does stabilisation mean in terms of the knowledge encoded in the states of stabilised games, and in terms of existence of knowledge-based strategies in them?
  We only note here that stabilisation corresponds to the existence of a finite knowledge representation that contains the higher-order knowledge of the agents of *any* order. The knowledge encoded in the locations of stable game graphs could thus be "folded" into a cyclic representation.
- What structural conditions on a game are necessary, or sufficient, for its iterated MKBSC to eventually stabilise?
- For which classes of objectives does it suffice to search for memoryless strategies in their stable expansions?
- What are the implications of non-stabilisation?

These questions will be explored in follow-up work.

6.3. Limitations of the MKBSC

As stated in the Introduction, our knowledge representation is just one possible choice, albeit one with a good intuitive justification. As the following example illustrates, however, it is not the case that whenever there is a profile of observation-based finite-memory strategies that is winning for a given reachability objective, then there is also a winning profile of j-order knowledge-based strategies (of the type studied here) for some j.

Example 24. Consider the game shown in Fig. 11. It models a similar scenario as our cup-lifting game, but now the cup needs to be oriented suitably before it can be lifted. The game is stable, meaning that whatever can be achieved by a profile of j-order knowledge-based strategies for any $j \geq 1$, it can also be achieved by a profile of observation-based memoryless strategies. However, for this game, there is no winning profile of observation-based memoryless strategies, while there

clearly is a winning profile of observation-based finite-memory ones: after the initial wait, robot 1 waits once more, while robot 0 spins the cup if needed; then the two robots lift the cup.

7. Related work

As pointed out in the introduction, the present work relates in more, or less, essential ways with several major research areas, including decentralised cooperative decision making, multi-agent planning, knowledge-based programs, games with imperfect information and strategy synthesis in them, etc., where variations of the general problem in focus of this paper have been explored, some of them quite extensively. Here we provide a reasonably detailed, yet inevitably incomplete list of the conceptually and technically closest to our work research areas and topics, with some relevant key references on them.

7.1. Games with imperfect information and knowledge-based strategies

Most of the basic notions and components (including terminology and notation) of our framework are adopted from general studies of games with imperfect information, such as [11,2,1,4].

In particular, the *knowledge-based subset construction* (KBSC) for single-agent games against Nature has been introduced and studied in [2,3]. The generalisation MKBSC for MAGIIAN explored here was first proposed in [14] by the third co-author. Our presentation of the MKBSC is equivalent to that one, but makes explicit the different stages of the construction (the original proposal defines directly the expanded game). This "deconstruction" of the original definition has allowed us to define a translation of strategies to transducers (Definition 12), and to propose a heuristic for strategy synthesis based on single-agent games with perfect information (Section 4.4). Furthermore, [14] did not provide any formal characterisation of the construction, but discovered the phenomenon of stabilisation and made the observation that iterating the MKBSC in effect computes higher-order knowledge, illustrating this on the cup-lifting game. Finally, [14] made the observation of the limitation of the construction discussed here in Section 6.3.

The notion of knowledge-based strategies proposed and explored here is closely related to the notion of *knowledge-based programs*, introduced and studied, e.g., in [7,5,6]. Our presentation is more abstract and less algorithmic, but our results can easily be adapted to the latter notion. These may be useful for the synthesis of protocols and for programming intelligent agents.

Constructions representing and using agents' higher-order knowledge have been proposed in [7–10]. All these are essentially related to our MKBSC construction. The notion of *k*-trees proposed in [8] was already discussed in Section 5.2. Among the constructions, special mention deserves the *epistemic unfolding* introduced and studied in [10]. The construction essentially translates a MAGIIAN, in a strategy-preserving fashion, to a single-agent game with perfect information. The resulting (expanded) game, however, is generally infinite (even when collapsed to so-called "homomorphic cores"), and the construction is thus not guaranteed to terminate. In contrast, our MKBSC construction always results in a finite game, but does not necessarily remove the imperfect information. Another difference is that the unfolding is based on an "epistemic model" common to all agents, and thus addresses the (YN) case discussed in Section 2.2.

7.2. Logics for multi-agent knowledge and strategic reasoning

While we have not involved formal logical languages and systems in this study, we note that it is essentially related to various logics and models of multi-agent knowledge and strategic reasoning.

First, there is a clear link of our work with *multi-agent epistemic logic*. Every agent in our framework is implicitly assumed to be a perfect (ideal) reasoner and its knowledge satisfies the standard S5 principles. Thus, the knowledge, both factual and higher-order, of all agents is modelled in multi-agent epistemic frames by epistemic indistinguishability relations [27], which are equivalence relations that partition the state space into families of information/knowledge states (see e.g. [6]). These are mutually inter-definable with *Aumann structures*, and higher-order knowledge of the agents can be computed in either of them in a standard way (see again [6]). These epistemic structures naturally arise in the MAGIIAN models, but we do not define and deal with them explicitly, since we do not need to involve explicitly epistemic logic, or any other formal logical system in the present work.

Furthermore, there are 2 hierarchies of epistemic structures naturally arising in MAGIIAN models and the problems we study. The first one is the hierarchy of *static higher-order knowledge*, associated with the hierarchy of iterations of the MKBSC construction applied to the original game. Every iteration increases the order of explicitly represented agents' knowledge, starting with 0-order (the factual knowledge about the current state of the game), then 1-order (the knowledge about the other agents' 0-order knowledge), 2nd order, etc. This is static knowledge because it is not associated with any particular play. The second one is the hierarchy of *(0-order) dynamic knowledge*, associated with each particular play in the given model, where the agents re-compute it after every transition and new observation they make, in the way described in this paper. Here we also mention the close conceptual links with *dynamic epistemic logic (DEL)* [27], especially the mechanism of epistemic updates of the knowledge representing models. Some recent studies relating DEL and its use in solving concurrent games include [28,29], technical ideas in which can be used for further extension of the present work to a non-cooperative setting.

These two hierarchies of knowledge can be interleaved in a natural way, by producing a combined hierarchy of *higher-order dynamic knowledge*, the capturing and utilisation of which for synthesising winning strategy profiles for the team of agents we have explored here. We leave the study of that combined hierarchy from the perspective of multi-agent epistemic logic to future work, but we note its relationship with the hierarchy of *k*-trees, defined and explored originally in [8], and further works, including [30,31].

Another natural link can also be made with general game description languages (GDL) as a different framework for knowledge representation in games, and [32] provides a comparative study of the GDL approach with the DEL-based framework that can also be useful for further work on the topic.

An important related line of research on *models and logics for games with (perfect and) imperfect information* comprises a large number of works, going back to [33–35], focusing mainly on the logical properties of semantics, expressiveness, deciding satisfiability, etc. More recent and also more computationally focused works in that line of research include:

- [36], developing a model checking algorithm for a variant of the alternating time temporal logic ATL with knowledge operators, assuming imperfect information and perfect recall, but also communication between the agents, so the strategies considered there employ the distributed knowledge of the agents.
- [37], which develops (bounded) model checking methods for reachability and other problems, expressible in the logic ATL, under the assumption of imperfect information and perfect recall, but with bounded horizon.
- [38], which considers a variant of the logic NatATL ("ATL with Natural Strategies") introduced in [39] with imperfect information and studies the model checking problem for it. In this logic, bounds are imposed on the complexity of the admissible strategies, assuming that they are represented by lists of guarded actions.

More related references can be found in [40].

The closest link of our study with that research line are the concurrent game models with incomplete and imperfect information that are essentially used as models in our work, too. The major distinctions in our study as compared to these are, first, that all agents work as a team; second, that there is a non-deterministic environment deciding the outcome of the action profiles of the team; third, that we do not employ (yet) a formal logical language here, and fourth, that the strategy profile is designed externally and then communicated, either fully, or only locally, to the individual agents. These make the methods and results of our work substantially different from those presented in the literature mentioned above.

7.3. Decentralised partially observable Markov decision processes

Technically, our framework of MAGIIAN models is a special case of *decentralised partially observable Markov decision processes (Dec-POMDP)* [41,21,42,20], modelling multi-agent planning and decision-making under uncertainty, where the policy planning is centralised whereas the execution is assumed decentralised because of the lack of (adequate) communication between the agents in the execution phase. The essential differences of the general framework of Dec-POMDPs from our MAGIIAN models are as follows:

- The transitions in Dec-POMDPs are determined by transition probability functions with explicitly specified distributions associated with each action profile, whereas in MAGIIAN models they are nondeterministically settled for each action profile by Nature. (The possible outcomes can be assumed uniformly distributed, but without using that assumption for optimising the team's joint policy.)
- The *reward function* in Dec-POMDPs is typically quantitative and the aim is to maximize it, whereas in MAGIIAN models it is a qualitative objective, typically reachability or safety.
- The agents' observations in Dec-POMDPs are stochastic, subject to given probability distributions, whereas in MAGIIAN models they are deterministic, i.e. uniquely determined by the states.
- A finitely bounded *horizon* (number of transition steps for optimising the reward) is usually explicitly assumed. On the other hand, the horizon on the problem that we study here is implicitly assumed unbounded.

The typical decision problem studied for Dec-POMDPs is as follows: given a Dec-POMDP M, a positive integer horizon T and a reward threshold K, the question is whether there is a joint policy for all agents that yields a total reward in T steps which is at least K. This problem is clearly decidable and the main research problem in the studies cited above is to analyse and determine its complexity under various additional assumptions. Typically, it is NEXPTIME-complete, even in the 2-agent case. This is where the main difference with our study occurs. The reward function in our framework is very simple: it assigns a reward 1 if the objective is reached, otherwise assigns 0, and we do not assume a pre-defined finite horizon, which makes the respective reachability problem in the focus of our study generally semi-decidable – just like the infinite-horizon Dec-POMDP problems under various optimality criteria, cf. [41]. Respectively, our main aim is to develop semi-decision procedures for constructing successful strategy (policy) profiles, and a major problem of our study is to ensure termination of these procedures. And, very importantly, we essentially employ the higher-order knowledge of the agents in the design of these strategy profiles, which is (at least explicitly) not taken into account in the alternative approaches and methods mentioned further.

From the large body of literature on Dec-POMDPs, we have identified the following directions and works as the closest to our study:

- Approximation algorithms for solving the infinite-horizon problems in Dec-POMDPs with quantitative reward functions have been proposed, e.g. in [43], using a joint controller with a correlation device that sends signals to all agents, plus a bounded policy iteration algorithm for improving the agents' finite-state controllers; followed by [44] where an optimal policy iteration algorithm for solving DEC-POMDPs is developed, using stochastic finite-state controllers to represent policies. Other works in this direction include [45], presenting a best-first search algorithm for computing an optimal policy vector, and [46], based on a memory-bounded optimisation approach using nonlinear optimisation techniques. In [47], an incremental policy generation method is applied to Dec-POMDPs with finite horizon using one-step reachability analysis, but the same approach can also be used in infinite-horizon policy iteration. In [48] a new MacDec-POMDP planning algorithm is presented that searches over policies represented as finite-state controllers, which can be much more concise and easier to interpret than representations based on policy trees, and can operate over an infinite horizon. In [49], the infinite-horizon assumption is replaced by indefinite-horizon.
  The methods and results mentioned above, as well as other related works in the area, are well surveyed in [21,42,22]. It is important to emphasize that all these methods are specifically applicable to optimising *quantitative* reward functions, but essentially not – at least, not naturally and efficiently – in the case of qualitative reachability objectives studied here.
- [16] proposes transforming a Dec-POMDP into a deterministic MDP, based on "complete information-states" which represent the joint history of the individual "decision rules" applied by the agents, starting with a given initial belief state. That enables reduction of finding an optimal separable joint policy in the original Dec-POMDP to the same problem in the resulting complete information-state MDP, by applying various methods developed for the latter problem. Again, the methods explored there apply to optimising *quantitative* reward functions.
- [18] studies the decision problem with finite horizon, but where the objectives are *qualitative*, reachability goals, rather than maximizing quantitative long-term reward functions, as in general Dec-POMDPs. It is shown there that, under certain assumptions (including an explicit finite horizon and a shared initial belief state) the complexity of the problem is as hard as the one for standard Dec-POMDP, and a method for computing a solution for the deterministic case based on compilation to classical planning is presented.
- We also note the recent works [17,50] which show, inter alia, that all (deterministic) joint policies for QDec-POMDPs can be (succinctly) represented as multi-agent knowledge-based programs, but without discussing the question of how such policies (and their representing programs) for a given objective can be constructed. This result is quite close in spirit to our observation that the intensional and extensional views on knowledge-based strategies are equivalent.

In [51] the different, yet quite similar to Dec-POMDPs framework of *Multiagent Team Decision Problem (MTDP)* is presented, where possible explicit communication between the agents is also considered, and the assumption of agents' perfect recall for the agents is made. In particular, "state-estimator functions" are added and used to model and update the current belief states based on the agents' communication and recall. These are conceptually similar to our abstract mechanism for knowledge updates. Again, only quantitative reward functions are considered in that work, and the complexities of solving the respective decision problems are studied and shown in [21] to be of the same complexity as for Dec-POMDPs.

In summary, none of the works on solving Dec-POMDPs or related problems surveyed above addresses the case of reachability objectives with unbounded horizon studied here, nor do they propose a solution that is adequate and efficient for solving that problem. Thus, whereas technically our work falls in the broader framework of Dec-POMDPs, both the decision problem studied here and our approach to its solution are substantially different from those mentioned above.

7.4. Planning under uncertainty

From a more general perspective, our work also relates, albeit in less essential ways, to other formal frameworks and studies of cooperative multi-agent planning under uncertainty [52–55] and, in particular, multi-agent epistemic planning [56–59]. Some of these works, e.g. [52], also assume incompleteness of the information about the domain, viz. that some of that information is unavailable at plan time, but can be acquired at runtime by the agents executing the plan, by also taking into account the higher-order knowledge or beliefs of the agents.

Major differences from most of these works are that in our framework agents are assumed to cooperate but not communicate with each other, their collective goal is not epistemic but ontic, and they act collectively against Nature. Still, some ideas and techniques from these works are quite relevant to our study and worth exploring further.

We also note the general link with (multi-agent) temporal planning [60–62]. Although the models and problems studied here, and the method we developed for their solutions, are different from those explored in that area, some ideas and approaches from the latter, such as easy-to-use modelling languages, can be beneficial for further progress on the topic explored here.

7.5. Other related topics and works

Of the many other related topics and works, we only mention the implicit link of our study to *theories of mind* (see, e.g., [63]), though in our case the "minds" of the agents are only represented by their knowledge known to the other agents, and possibly by the strategies which they follow, but not by beliefs, intentions, and other attitudes. Still, we believe that many interesting phenomena of "mind", especially in the context of Artificial Intelligence, can be studied and explored in our simplified framework of knowledge.

8. Conclusion

We have studied agents' (first- and higher-order) knowledge representation in multi-agent games with imperfect information against Nature and its use for synthesising knowledge-based strategy profiles for a team of agents by a supervisor, which then provides each agent with their own strategy and lets them play without them being able to communicate with each other. In particular, we have introduced and studied the generalised knowledge-based construction MKBSC and have established connections between (transducer-based) finite-memory strategies and knowledge-based strategies in the original game, and observation-based memoryless strategies in its iterated MKBSC expansions.

Conclusions From our results one can draw the following conclusions. First, higher-order (nested) knowledge can be based on the notion of "most precise estimate of the current state-of-affairs". The higher the order (i.e., the nesting depth) of knowledge, the higher the strategic abilities of the team. Also, the higher the uncertainty of the agents, i.e., the less they observe and know, the higher the benefit from nested knowledge. Next, for the class of knowledge-based strategies considered here (i.e., for the proposed notion of knowledge) and the classes of reachability and safety objectives, for a given bound on the nesting depth of knowledge and under the PDK condition, we have an *algorithm* for strategy synthesis; without such a bound it is only a semi-algorithm. Then, there is a *duality* between the extensional and the intensional views. The former is more suitable for strategy synthesis, while the latter can be more convenient in the play, and can also be used to explain the synthesised strategies. And finally, on some games the iterated MKBSC *stabilises*. If there is no winning memoryless observation-based strategy in the stable expansion, then, under the PDK condition, there is no winning knowledge-based strategy of *any* order in the original game. However, there might still be a winning finite-memory observation-based strategy in the original game.

Future work In future work we plan to characterise the class of objectives that can be achieved with knowledge-based strategies of the type defined here. Next, we plan to capture formally the notion of "degree of imperfectness" of information and the intuition (expressed in Section 4.2) that the MKBSC decreases this degree. We also plan to study the strategy synthesis problem after relaxing some of the assumptions made here, such as the case (YN) when agents are permitted to know each others strategies, or when agents do have some (limited) communication. Further, we plan to study in depth the stabilisation phenomenon of the MKBSC and, in particular, characterise the structural conditions for stabilisation. Furthermore, we will explore other knowledge representations, comparing the respective classes of objectives that they are sufficient for, and define the corresponding expansions following the general scheme. We will also design strategy synthesis algorithms and heuristics, and investigate their complexity. Further, we will explore temporal (epistemic) logic as a means for defining objectives, and epistemic logic as a means for representing the individual knowledge-based strategies. Related to this, we plan to investigate the notion of common knowledge in the context of MAGIIAN games. Then, we will explore the connection of our work to multi-agent epistemic planing. And finally, we plan to evaluate the practical utility of the strategy synthesis method proposed here, and investigate in depth potential application areas.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. *k*-trees

Following [8], we define the set of *k*-**trees** $\mathcal{T}_k$ and the set of i-**objective** *k*-**trees** $\mathcal{T}_{k,i}$, for all agents i ∈ Agt. The set of i-objective *k*-forests is defined as $\mathcal{F}_{k,i} \stackrel{\text{def}}{=} 2^{\mathcal{T}_{k,i}}$.

Definition 25 *(k-tree)*. Let **G** be a MAGIIAN over the set of locations Loc, with agents Agt $= \{1, 2, \ldots, n\}$. We recursively define the sets:

$$\mathcal{T}_0 \stackrel{\text{def}}{=} \mathsf{Loc} \quad (\text{or rather, } \{\langle l, \varnothing, \dots, \varnothing \rangle \mid l \in \mathsf{Loc}\})$$

$$\mathcal{T}_{0,\mathsf{i}} \stackrel{\text{def}}{=} \mathsf{Loc}$$

$$\mathcal{T}_{k+1} \stackrel{\text{def}}{=} \left\{ \langle l, F_1, \dots, F_n \rangle \mid l \in \mathsf{Loc} \text{ and } F_i \in \mathcal{F}_{k,\mathsf{i}} \text{ for all } \mathsf{i} \in \mathsf{Agt.} \right\}$$

$$\mathcal{T}_{k+1,\mathsf{i}} \stackrel{\text{def}}{=} \left\{ \langle l, F_1, \dots, F_n \rangle \in \mathcal{T}_{k+1} \mid F_i = \varnothing \right\}$$

Example 26. In a 2-agent game against Nature, consider the 2-tree:

$$t_1^{(2)} = \left\langle l_2, F_1^{(1)}, F_2^{(1)} \right\rangle$$

where:

$$F_1^{(1)} = \{\langle l_1, \varnothing, \{l_1\} \rangle, \langle l_2, \varnothing, \{l_2, l_3\} \rangle\}$$
$$F_2^{(1)} = \{\langle l_2, \{l_1, l_2\}, \varnothing \rangle, \langle l_3, \{l_3\}, \varnothing \rangle\}$$

This 2-tree models a state of affairs in which the game is in location l_2, and agent 1 knows (i.e., considers possible, as modelled by $F_1^{(1)}$) that the game is either in location l_1 or in l_2, and that in the former case agent 2 knows that the game is in l_1, while in the latter case, in l_2 or l_3. The knowledge of agent 2 is analogous, as modelled by $F_2^{(1)}$.

Next we define, by mutual recursion, a **global update** function G_k, and a family of local update functions $H_{k,\mathsf{i}}$:

$$G_k : \mathcal{T}_k \times \mathsf{Act} \times \mathsf{Loc} \to \mathcal{T}_k$$

$$H_{k,\mathsf{i}} : \mathcal{F}_{k,\mathsf{i}} \times \mathsf{Act}_\mathsf{i} \times \mathsf{Obs}_\mathsf{i} \to \mathcal{F}_{k,\mathsf{i}}$$

Definition 27 (*Knowledge update*). Let $\mathbf{G} = (\mathsf{Loc}, l_{\mathsf{init}}, \mathsf{Act}, \Delta, \mathsf{Obs})$ be a MAGIIAN. We recursively define the functions:

$$G_0(l, \mathsf{act}, l') \stackrel{\text{def}}{=} l'$$

$$G_{k+1}(\langle l, F_1, \dots, F_n \rangle, (\mathsf{act}_1, \dots, \mathsf{act}_n), l') \stackrel{\text{def}}{=}$$
$$\left\langle l', H_{k,1}(F_1, \mathsf{act}_1, \mathsf{obs}_1(l')), \dots, H_{k,n}(F_n, \mathsf{act}_n, \mathsf{obs}_n(l')) \right\rangle$$

$$H_{k,\mathsf{i}}(F_i, \mathsf{act}_\mathsf{i}, o_\mathsf{i}) \stackrel{\text{def}}{=}$$
$$\left\{ G_k(t^{(k)}, \mathsf{act}, l) \mid t^{(k)} \in F_i, \mathsf{act(i)} = \mathsf{act}_\mathsf{i}, (root(t^{(k)}), \mathsf{act}, l) \in \Delta, l \in o_\mathsf{i} \right\}$$

The above definition is slightly more general than the original one from [8], since the game model presented there does not involve named actions.

References

[1] A. Pnueli, R. Rosner, Distributed reactive systems are hard to synthesize, in: Proceedings of FOCS'90, IEEE Computer Society, 1990, pp. 746–757.
[2] J.H. Reif, The complexity of two-player games of incomplete information, Comput. Syst. Sci. 29 (2) (1984) 274–301.
[3] K. Chatterjee, L. Doyen, T.A. Henzinger, J. Raskin, Algorithms for omega-regular games with imperfect information, Log. Methods Comput. Sci. 3 (3) (2007) 1–23, https://doi.org/10.2168/LMCS-3(3:4)2007.
[4] L. Doyen, J. Raskin, Games with imperfect information: theory and algorithms, in: K.R. Apt, E. Grädel (Eds.), Lectures in Game Theory for Computer Scientists, Cambridge University Press, 2011, pp. 185–212, http://www.cambridge.org/gb/knowledge/isbn/item5760379.
[5] R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi, Knowledge-based programs, Distrib. Comput. 10 (4) (1997) 199–225, https://doi.org/10.1007/s004460050038.
[6] R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi, Reasoning About Knowledge, MIT Press, Cambridge, MA, USA, 2003.
[7] R. Fagin, J.Y. Halpern, M.Y. Vardi, A model-theoretic analysis of knowledge, J. ACM 38 (2) (1991) 382–428, https://doi.org/10.1145/103516.128680.
[8] R. van der Meyden, Common knowledge and update in finite environments, Inf. Comput. 140 (2) (1998) 115–157.
[9] D. Berwanger, Ł. Kaiser, Information tracking in games on graphs, J. Log. Lang. Inf. 19 (4) (2010) 395–412, https://doi.org/10.1007/s10849-009-9115-8.
[10] D. Berwanger, L. Kaiser, B. Puchala, A perfect-information construction for coordination in games, in: Proceedings of FSTTCS 2011, in: LIPIcs, vol. 13, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2011, pp. 387–398.
[11] G.L. Peterson, J.H. Reif, Multiple-person alternation, in: Proceedings of FOCS 1979, IEEE Computer Society, 1979, pp. 348–363.
[12] C. Dima, F.L. Tiplea, Model-checking ATL under imperfect information and perfect recall semantics is undecidable, arXiv:1102.4225.
[13] S. Vester, Alternating-time temporal logic with finite-memory strategies, in: G. Puppis, T. Villa (Eds.), Proceedings Fourth International Symposium on Games, Automata, Logics and Formal Verification, GandALF 2013, Borca di Cadore, Dolomites, Italy, 29-31th August 2013, in: EPTCS, vol. 119, 2013, pp. 194–207.
[14] E. Lundberg, Collaboration in Multi-Agent Games, Tech. Rep., KTH Royal Institute of Technology, School of Computer Science and Communication, 2017, https://kth.diva-portal.org/smash/get/diva2:1115157/FULLTEXT01.pdf.
[15] P. Kazmierczak, T. Ågotnes, W. Jamroga, Multi-agency is coordination and (limited) communication, in: H.K. Dam, J.V. Pitt, Y. Xu, G. Governatori, T. Ito (Eds.), Proceedings of PRIMA 2014, in: Lecture Notes in Computer Science, vol. 8861, Springer, 2014, pp. 91–106.

[16] J.S. Dibangoye, C. Amato, O. Buffet, F. Charpillet, Optimally solving Dec-POMDPs as continuous-state MDPs, J. Artif. Intell. Res. 55 (2016) 443–497, https://doi.org/10.1613/jair.4623.

[17] A. Saffidine, F. Schwarzentruber, B. Zanuttini, Knowledge-based policies for qualitative decentralized POMDPs, in: S.A. McIlraith, K.Q. Weinberger (Eds.), Proceedings of AAAI-18 and EAAI-18, AAAI Press, 2018, pp. 6270–6277, https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17029.

[18] R.I. Brafman, G. Shani, S. Zilberstein, Qualitative planning under partial observability in multi-agent domains, in: M. desJardins, M.L. Littman (Eds.), Proceedings of AAAI 2013, AAAI Press, 2013, http://www.aaai.org/ocs/index.php/AAAI/AAAI13/paper/view/6388.

[19] H. Nylén, A. Jacobsson, Investigation of a Knowledge-based Subset Construction for Multi-player Games of Imperfect Information, Tech. Rep., KTH Royal Institute of Technology, School of Computer Science and Communication, 2018, https://kth.diva-portal.org/smash/get/diva2:1221520/FULLTEXT01.pdf.

[20] F.A. Oliehoek, C. Amato, A Concise Introduction to Decentralized POMDPs, Springer Briefs in Intelligent Systems, Springer, 2016.

[21] S. Seuken, S. Zilberstein, Formal models and algorithms for decentralized decision making under uncertainty, Auton. Agents Multi-Agent Syst. 17 (2) (2008) 190–250, https://doi.org/10.1007/s10458-007-9026-5.

[22] C. Amato, G. Chowdhary, A. Geramifard, N.K. Ure, M.J. Kochenderfer, Decentralized control of partially observable Markov decision processes, in: Proceedings of CDC 2013, IEEE, 2013, pp. 2398–2405.

[23] J. Pilecki, M.A. Bednarczyk, W. Jamroga, SMC: synthesis of uniform strategies and verification of strategic ability for multi-agent systems, J. Log. Comput. 27 (7) (2017) 1871–1895, https://doi.org/10.1093/logcom/exw032.

[24] E. Handberg, L. Rostami, Epistemic Structures for Strategic Reasoning in Multi-Player Games, Tech. Rep., KTH Royal Institute of Technology, School of Computer Science and Communication, 2019, https://kth.diva-portal.org/smash/get/diva2:1338657/FULLTEXT01.pdf.

[25] D.O. Stahl, P.W. Wilson, On players′ models of other players: theory and experimental evidence, Games Econ. Behav. 10 (1) (1995) 218–254, https://doi.org/10.1006/game.1995.1031.

[26] J.Y. Halpern, Y. Moses, Knowledge and common knowledge in a distributed environment, J. ACM 37 (3) (1990) 549–587, https://doi.org/10.1145/79147.79161.

[27] H. van Ditmarsch, W. van der Hoek, B. Kooi, Dynamic Epistemic Logic, Springer, Dordecht, 2008.

[28] B. Maubert, S. Pinchinat, F. Schwarzentruber, S. Stranieri, Concurrent games in dynamic epistemic logic, in: C. Bessiere (Ed.), Proceedings of IJCAI 2020, ijcai.org, 2020, pp. 1877–1883.

[29] B. Maubert, S. Pinchinat, F. Schwarzentruber, Reachability games in dynamic epistemic logic, in: S. Kraus (Ed.), Proceedings of IJCAI 2019, Macao, China, August 10-16, 2019, ijcai.org, 2019, pp. 499–505.

[30] X. Huang, R. van der Meyden, Synthesizing strategies for epistemic goals by epistemic model checking: an application to pursuit evasion games, in: Proceedings of AAAI 2012, 2012.

[31] R. van der Meyden, M.Y. Vardi, Synthesis from knowledge-based specifications, CoRR, arXiv:1307.6333 [abs], 2013.

[32] T. Engesser, R. Mattmüller, B. Nebel, M. Thielscher, Game description language and dynamic epistemic logic compared, Artif. Intell. 292 (2021) 103433, https://doi.org/10.1016/j.artint.2020.103433.

[33] W. van der Hoek, M. Wooldridge, Cooperation, knowledge, and time: alternating-time temporal epistemic logic and its applications, Stud. Log. 75 (1) (2004) 125–157.

[34] W. Jamroga, W. van der Hoek, Agents that know how to play, Fundam. Inform. 63 (2–3) (2004) 185–219.

[35] W. Jamroga, T. Ågotnes, Constructive knowledge: what agents can achieve under incomplete information, J. Appl. Non-Class. Log. 17 (4) (2007) 423–475.

[36] D.P. Guelev, C. Dima, C. Enea, An alternating-time temporal logic with knowledge, perfect recall and past: axiomatisation and model-checking, J. Appl. Non-Class. Log. 21 (1) (2011) 93–131, https://doi.org/10.3166/jancl.21.93-131.

[37] X. Huang, Bounded model checking of strategy ability with perfect recall, Artif. Intell. 222 (2015) 182–200, https://doi.org/10.1016/j.artint.2015.01.005.

[38] W. Jamroga, V. Malvone, A. Murano, Natural strategic ability under imperfect information, in: E. Elkind, M. Veloso, N. Agmon, M.E. Taylor (Eds.), Proceedings of AAMAS 2019, International Foundation for Autonomous Agents and Multiagent Systems, 2019, pp. 962–970, http://dl.acm.org/citation.cfm?id=3331791.

[39] W. Jamroga, V. Malvone, A. Murano, Natural strategic ability, Artif. Intell. 277 (2019), https://doi.org/10.1016/j.artint.2019.103170.

[40] T. Ågotnes, V. Goranko, W. Jamroga, M. Wooldridge, Knowledge and ability, in: H. van Ditmarsch, J. Halpern, W. van der Hoek, B. Kooi (Eds.), Handbook of Epistemic Logic, College Publications, 2015, pp. 543–589.

[41] D.S. Bernstein, R. Givan, N. Immerman, S. Zilberstein, The complexity of decentralized control of Markov decision processes, Math. Oper. Res. 27 (4) (2002) 819–840, https://doi.org/10.1287/moor.27.4.819.297.

[42] F.A. Oliehoek, Decentralized POMDPs, in: M.A. Wiering, M. van Otterlo (Eds.), Reinforcement Learning, in: Adaptation, Learning, and Optimization, vol. 12, Springer, 2012, pp. 471–503.

[43] D.S. Bernstein, E.A. Hansen, S. Zilberstein, Bounded policy iteration for decentralized POMDPs, in: L.P. Kaelbling, A. Saffiotti (Eds.), Proceedings of IJCAI-05, Professional Book Center, 2005, pp. 1287–1292, http://ijcai.org/Proceedings/05/Papers/0379.pdf.

[44] D.S. Bernstein, C. Amato, E.A. Hansen, S. Zilberstein, Policy iteration for decentralized control of Markov decision processes, J. Artif. Intell. Res. 34 (2009) 89–132, https://doi.org/10.1613/jair.2667.

[45] D. Szer, F. Charpillet, An optimal best-first search algorithm for solving infinite horizon DEC-POMDPs, in: J. Gama, R. Camacho, P. Brazdil, A. Jorge, L. Torgo (Eds.), Proceedings of ECML 2005, in: Lecture Notes in Computer Science, vol. 3720, Springer, 2005, pp. 389–399.

[46] C. Amato, D.S. Bernstein, S. Zilberstein, Optimizing memory-bounded controllers for decentralized POMDPs, in: R. Parr, L.C. van der Gaag (Eds.), Proceedings of UAI 2007, AUAI Press, 2007, pp. 1–8, https://dl.acm.org/doi/abs/10.5555/3020488.3020489.

[47] C. Amato, J.S. Dibangoye, S. Zilberstein, Incremental policy generation for finite-horizon DEC-POMDPs, in: A. Gerevini, A.E. Howe, A. Cesta, I. Refanidis (Eds.), Proceedings of ICAPS 2009, AAAI, 2009, http://aaai.org/ocs/index.php/ICAPS/ICAPS09/paper/view/711.

[48] C. Amato, G.D. Konidaris, A. Anders, G. Cruz, J.P. How, L.P. Kaelbling, Policy search for multi-robot coordination under uncertainty, Int. J. Robot. Res. 35 (14) (2016) 1760–1778, https://doi.org/10.1177/0278364916679611.

[49] C. Amato, S. Zilberstein, Achieving goals in decentralized POMDPs, in: C. Sierra, C. Castelfranchi, K.S. Decker, J.S. Sichman (Eds.), Proceedings of AAMAS 2009, vol. 1, IFAAMAS, 2009, pp. 593–600, https://dl.acm.org/citation.cfm?id=1558095.

[50] B. Zanuttini, J. Lang, A. Saffidine, F. Schwarzentruber, Knowledge-based programs as succinct policies for partially observable domains, Artif. Intell. 288 (2020) 103365, https://doi.org/10.1016/j.artint.2020.103365.

[51] D.V. Pynadath, M. Tambe, The communicative multiagent team decision problem: analyzing teamwork theories and models, J. Artif. Intell. Res. 16 (2002) 389–423, https://doi.org/10.1613/jair.1024.

[52] S. Sardiña, G.D. Giacomo, Y. Lespérance, H.J. Levesque, On the limits of planning over belief states under strict uncertainty, in: P. Doherty, J. Mylopoulos, C.A. Welty (Eds.), Proceedings of KR 2006, AAAI Press, 2006, pp. 463–471, http://www.aaai.org/Library/KR/2006/kr06-048.php.

[53] A. Torreño, E. Onaindia, O. Sapena, FMAP: distributed cooperative multi-agent planning, Appl. Intell. 41 (2) (2014) 606–626, https://doi.org/10.1007/s10489-014-0540-2.

[54] C.J. Muise, V. Belle, P. Felli, S.A. McIlraith, T. Miller, A.R. Pearce, L. Sonenberg, Planning over multi-agent epistemic states: a classical planning approach, in: B. Bonet, S. Koenig (Eds.), Proceedings of AAAI 2015, AAAI Press, 2015, pp. 3327–3334, http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9974.

[55] A. Torreño, E. Onaindia, A. Komenda, M. Stolba, Cooperative multi-agent planning: a survey, ACM Comput. Surv. 50 (6) (2018) 84:1–84:32, https://doi.org/10.1145/3128584.

[56] T. Bolander, M.B. Andersen, Epistemic planning for single and multi-agent systems, J. Appl. Non-Class. Log. 21 (1) (2011) 9–34, https://doi.org/10.3166/jancl.21.9-34.

[57] M.C. Cooper, A. Herzig, F. Maffre, F. Maris, P. Régnier, A simple account of multi-agent epistemic planning, in: G.A. Kaminka, M. Fox, P. Bouquet, E. Hüllermeier, V. Dignum, F. Dignum, F. van Harmelen (Eds.), Proceedings of ECAI 2016, in: Frontiers in Artificial Intelligence and Applications, vol. 285, IOS Press, 2016, pp. 193–201.

[58] T. Engesser, T. Bolander, R. Mattmüller, B. Nebel, Cooperative epistemic multi-agent planning for implicit coordination, in: S. Ghosh, R. Ramanujam (Eds.), Proceedings of M4M@ICLA 2017, in: EPTCS, vol. 243, 2017, pp. 75–90.

[59] Y. Li, Y. Wang, Multi-agent knowing how via multi-step plans: a dynamic epistemic planning based approach, in: P. Blackburn, E. Lorini, M. Guo (Eds.), Proceedings of LORI 2019, in: Lecture Notes in Computer Science, vol. 11813, Springer, 2019, pp. 126–139.

[60] W. Cushing, S. Kambhampati, Mausam, D.S. Weld, When is temporal planning really temporal?, in: M.M. Veloso (Ed.), Proceedings of IJCAI 2007, 2007, pp. 1852–1859, http://ijcai.org/Proceedings/07/Papers/299.pdf.

[61] M.C. Cooper, F. Maris, P. Régnier, Tractable monotone temporal planning, in: L. McCluskey, B.C. Williams, J.R. Silva, B. Bonet (Eds.), Proceedings of ICAPS 2012, AAAI, 2012, http://www.aaai.org/ocs/index.php/ICAPS/ICAPS12/paper/view/4689.

[62] M.C. Cooper, F. Maris, P. Régnier, Monotone temporal planning: tractability, extensions and applications, J. Artif. Intell. Res. 50 (2014) 447–485, https://doi.org/10.1613/jair.4358.

[63] I. van de Pol, I. van Rooij, J. Szymanik, Parameterized complexity results for a model of theory of mind based on dynamic epistemic logic, Electron. Proc. Theor. Comput. Sci. 215 (2016) 246–263, https://doi.org/10.4204/eptcs.215.18.