

# Reasoning under uncertainty in the game of Cops and Robbers

Dazhu Li<sup>1,2</sup>, Sujata Ghosh<sup>3</sup>, Fenrong Liu<sup>4,5</sup>

<sup>1</sup>Institute of Philosophy, Chinese Academy of Sciences, Beijing, China.

<sup>2</sup>Department of Philosophy, University of Chinese Academy of Sciences, Beijing, China.

<sup>3</sup>Indian Statistical Institute, Chennai, India.

<sup>4</sup>The Tsinghua-UvA JRC for Logic, Department of Philosophy, Tsinghua University, Beijing, China.

<sup>5</sup>Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, The Netherlands.

## Abstract

The game of Cops and Robbers is an important model for studying computational queries in pursuit-evasion environments, among others. As recent logical explorations have shown, its structure exhibits appealing analogies with modal logic. In this paper, we enrich the game with a setting in which players may have imperfect information. We propose a new formal framework, Epistemic Logic of Cops and Robbers (**ELCR**), to make the core notions of the game precise, for instance, players' positions, observational power and inference. Applying **ELCR** to analyze the game, we obtain an automated way to track interactions between players and characterize their information updates during the game. The update mechanism is defined by a novel dynamic operator, and we compare it with some relevant paradigms from the game and logic perspectives. We study various properties of **ELCR** including axiomatization and decidability. To our knowledge, this is the first attempt to explore these games from a formal point of view where (partial) information available to players is taken into account.

**Keywords:** Cops and Robbers, Hide and Seek, Imperfect information, Observation power, Knowledge updates, Dynamic-epistemic logic, Axiomatization, Decidability

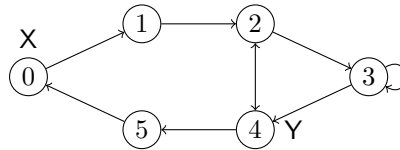
# 1 Introduction

Search missions and pursuit-evasion environments have been investigated in details in the study of robotic systems. Such pursuit-evasion problems, which can also be viewed as adversarial search problems, can be considered as strategizing problems in pursuit-evasion games or their multi-agent counterpart, *Cops and Robbers*. The game of Cops and Robbers refers to a family of pursuit-evasion games played on graphs, where the pursuer (or the cops) must develop optimal responses or search strategies against some worst-case adversary, the evader (or the robbers). The game has a deep root in computer science and serves as an important testbed for studying algorithms and computational complexity in search environments [40, 19].

In recent years, logicians have studied the Hide and Seek game, which shares a similar pursuit-evasion structure, with a focus on modeling the dynamic interaction between players [9, 14]. A two-dimensional modal logic for the game is proposed in [14], and its further properties and extensions are explored in [35, 36, 20, 44]. However, these logical models assume that players have perfect information, which limits the resulting frameworks to tools for reasoning solely about properties of graph structures. In this paper, we shift to an imperfect information setting, examining how agents reason and play the game under uncertainty due to their limited sight or observational power. So, the game of Cops and Robbers studied in the paper can be seen as an imperfect information version of Hide and Seek explored in [35, 36, 20, 44]. Our focus is on how they use partial information available to them during play. We propose a formal framework to support the study of these interactions. From a practical viewpoint, graph games have been used extensively to model reachability problems, social networks, and search problems. Our framework provides a formal language to precisely describe and study a wide range of such issues. This formal study can facilitate the autonomous agent-building efforts with a better know-how in terms of information available to these agents. The corresponding information updates during gameplay help address adversarial search problems.

To illustrate the game’s features, let us first consider a specific example. For simplicity, the game considered in this context is played by two players, a Cop and a Robber, on a *finite directed* graph in which every vertex has successors, an assumption made for ease of presentation.<sup>1</sup>

**Example 1.** *In the graph below, Cop X (female) is at 0, and Robber Y (male) is at 4. They know the graph structure and their own positions. A player can see the other if they are at the same position or at a vertex reachable by an arrow in either direction. Thus, X and Y do not know each other’s exact positions. However, given the graph structure and their observational range, X knows that Y must be at 2, 3, or 4, while Y knows that X must be at 0 or 1.*




---

<sup>1</sup>The game and logic introduced in this paper can be generalized to a setting with more players.

A player whose turn it is has to move along an arrow. Let  $X$  act first. She can only move to 1, and after this, she knows that  $Y$  is not at 2, otherwise, she would see him directly. So,  $X$  knows that  $Y$  must be at 3 or 4. For  $Y$ , he knows after her move that  $X$  is not at 2 (as otherwise he would see her directly) and not at 0 (since none of his previously considered positions for  $X$ —0 or 1—can reach 0 in one step). Therefore,  $Y$  concludes that  $X$  must now be at 1.<sup>2</sup>

Next, let  $Y$  move to 5. Since  $X$  observes that  $Y$  is no longer at 2, she knows that  $Y$  must be at 3, 4, or 5.  $X$  then moves to 2, from where she can see that  $Y$  is not at 3 or 4. Thus,  $X$  concludes that  $Y$  must be at 5 and wins.

Although the example is simple, it indicates several subtleties of the game. For instance, the players' knowledge changes continuously throughout the game, and in the final stage, although the players cannot see each other directly, they can know the positions of the other based on their knowledge about the graph structure and their observational power. Knowledge here is modelled based on observational powers and expresses the uncertainty of the players. In essence, it is quite similar to the way it is modelled in imperfect-information games in general (see e.g., [42]), where information is expressed in terms of equivalence relations, but ultimately models the uncertainties of the players.

In what follows, we will introduce the key notions precisely and present an *Epistemic Logic of Cops and Robbers* (ELCR), which enables us to reason about the action-information interplay and to automatically track knowledge changes during play. We are aware of the well-developed methodology of dynamic-epistemic logic (DEL, [7, 41, 21, 10]) and its various applications in modelling information update. We will compare the two approaches and provide examples to show that our framework may offer a more succinct representation. The interface between logic and games has a long history. Over the past two decades, inspired by board games, graph game logic has emerged as an important research area, [2, 3, 6, 9, 14, 18, 20, 29, 33, 35, 36, 44, 45], with various logical techniques developed to study player interactions on graphs. To our knowledge, this is the first attempt to explore more complex settings where agents have imperfect information. We believe that the ideas and techniques introduced in this article will help open a new line of research.

Before we begin, it is worth noting that this work extends the previous proceedings version [34]. Specifically, compared to the earlier version, this paper offers the following new contributions. The logical language is extended to allow dynamic operators within the scope of knowledge operators, and the models are redesigned to better reflect the game's structure. We establish complete calculi for both ELCR and its static fragment  $\text{ELCR}^-$  (without dynamic operators), further prove that ELCR is decidable. More new examples are included to highlight subtleties of the game and its logic, along with additional properties of the logic to deepen our understanding of the game. Finally, we provide an extensive discussion of related works, including a comparison of ELCR with the DEL-approach to the game proposed in a recent paper [12].

---

<sup>2</sup>At this stage  $Y$  also knows that  $X$  was at 0 before the movement. Although  $Y$  knows the actual action of  $X$ , the knowledge of  $Y$  is obtained by his reasoning with the knowledge about graph and his observational power. There is no inconsistency with the imperfect information nature of the game: for instance, if we let the game begin with the movement of  $Y$  from 4 to 2, then after the movement  $X$  would not know the actual action of  $Y$ , since  $X$  would think it is also possible that, e.g.,  $Y$  moves from 3 to 3.

*Structure of the paper.* In Section 2, we lay out the basics of the game and discuss its alternative designs. Section 3 presents the formal language and models of ELCR, proposes an alternative for a simultaneous-move variant, and compares the two logics. Section 4 applies the new framework to analyze the game of Cops and Robbers. Section 5 studies some basic properties of ELCR. Sections 6 and 7 provide complete Hilbert-style proof systems for the static fragment of the logic without dynamic operators and the whole ELCR respectively and show that ELCR has a decidable satisfiability problem. Section 8 formalizes some recent ideas on the DEL-approach to studying the game, compares it to our logic, and discusses relevant works. Section 9 concludes with directions for future research.

## 2 Game design

Let us first identify basic assumptions regarding players' knowledge. First, given a game, we assume that the graph structure is commonly known by the players. Also, the game is turn-based, and it is common knowledge that whose turn it is to move: here we simply require that in each round, Cop X moves first, and then Robber Y moves.<sup>3</sup> However, they may not know where the opponent moves from and/or where the opponent moves to. Finally, we assume that players can at least remember what they have considered to be possible at the previous stage: when a player moves, the players infer the new possible situations from what they considered to be possible before that movement. For instance, before the movement of X in Example 1, Y considers it possible for X to be at 0 or 1, and once X moves, he gets to know that X is at 1, which is the successor of the previous possibility 0.

As for the winning condition, one option is to stipulate that Cop wins iff she is at same position as Robber [35, 36]. However, our exploration on the role of knowledge allows more alternatives to define the condition, and we will adopt a new criterion:

*Winning condition:* Fix a natural number  $n \in \mathbb{N}$ . Cop X wins iff the position of Robber Y is *known* by X within  $n$  rounds.

*Strategy:* A strategy of a player is a function from the set of positions of the player to the possible moves at that position. A strategy is said to be winning for a player if the player can win the game by playing according to the strategy, whatever be the moves of the other player.<sup>4</sup>

*Ability of players:  $k$ -sight.* To analyze the imperfect information game, it is crucial to identify what players can know, which is dependent on the observational ability of the player. There are two extremes about their ability: players are assumed to know the positions of each other at any stage of a game, and they do not have any ability to ensure they can know something, even their own positions. Inspired by [5, 30, 38], we introduce the notion of  *$k$ -sight of players* to describe their observational power: players always know their own positions, and if their positions are reachable within  $k$

---

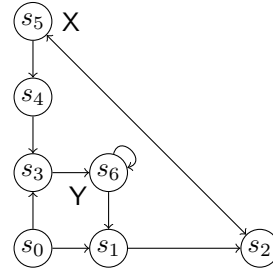
<sup>3</sup>The turn-based assumption is in line with many other works on the game, e.g., [40, 19]. One can also consider any other specific order of their play, and even simultaneous play, but that will not affect our basic idea to analyze the game and the design of the logical tool in Section 3.

<sup>4</sup>For instance, when we assume that the initial positions of X and Y in Example 1 are 3 and 5 respectively, a winning strategy of X is to always stay at 3: one can check that X can know the position of Y after the movement of Y from 0 to 1 in the second round.

steps via the arrows or their converse directions (in which case, we say that *they can see each other*), then they know the positions of each other. For instance, X and Y in Example 1 have sight 1. Below is another example.

**Example 2.** Assume that X and Y have sight 2. Based on our definition of the  $k$ -sight ability, to identify what the players can see, the direction of the arrows in the graph below does not matter.

All of the vertices  $s_0$ - $s_4$  and  $s_6$  are in the sight of Y from  $s_6$ , so Y knows that X is not at any of those vertices (which together with the knowledge about the graph structure would make Y know that X is at  $s_5$ ). From the vertex  $s_5$ , X can see the vertices  $s_1$ - $s_5$ , and based on the knowledge about the graph structure, X knows Y is at either  $s_0$  or  $s_6$ , although she does not know which one is exactly the case. But after X moves from  $s_5$  to  $s_2$ , the position of Y would come in the sight of X, meaning that after the move X would know where Y is.<sup>5</sup>



It is important to emphasize that the  $k$ -sight ability is used to characterize what *at least* players can know, but depending on the concrete situations players may know more. For instance, at the final stage of the game in Example 1, although X cannot observe Y directly based on the 1-sight ability, she can still know where Y is. We emphasize that our goal in this work is to develop logical tools to track changes in the players' knowledge of each other's positions. In terms of knowledge, as a first step we will only consider the knowledge of atomic facts, their Boolean combinations, and the effects of movements on them, but not higher-order knowledge (e.g., X knows that Y knows that X is at the vertex  $a$ ). Although this may look restricted, the proposal developed in this way fits many existing analyses for the imperfect information variants of Cops and Robbers (see e.g., [19]). We leave the work for the more intricate setting involving higher-order knowledge to another occasion.

We end this section by pointing out possible options regarding our assumptions of the game. For instance, as stated earlier, graphs in this context are *serial*, but it is equally reasonable to consider graphs without any restrictions; the players can act simultaneously; different players may have different sights; there can be more players with other kinds of ability, say, they can send each other messages; and there may be other ways to define the winning condition, for instance, when there are  $i > 1$  Cops, Cops may win when the position of Robber is their *distributed knowledge*. Finally, it is also meaningful to extend our current game with explicit probabilities of actions in different situations, a usual manner adopted in imperfect information games, which

<sup>5</sup>If we temporarily let Y move first in this example, then X would know where Y is after the first movement of Y. The reason is as follows. Player Y has two options: moving to  $s_1$  or staying at  $s_6$ . If he moves to  $s_1$ , then X can see him directly. When the latter is the case, X still cannot see Y directly, but the movement makes X know that it is impossible that Y is at another unobservable vertex  $s_0$ , since it has no predecessor from the previous possibilities  $s_0$  and  $s_6$  considered by X, i.e., it would not happen that Y moves to  $s_0$  from  $s_0$  or  $s_6$ . In the latter case, *physically nothing changes, but epistemically the player knows more.*

would affect how players update their knowledge and how they act.<sup>6</sup> Some of these alternatives will be discussed in the later sections. For now, let us move to the details of the logical framework.

### 3 Logical language and models

This section will present a logical framework that characterizes our assumptions about the game and enables us to reason about how players update their knowledge.

Inspired by [4, 8], we will use different *values* to encode different vertices in the graph of a game. Such values and the binary relation in a graph give us a semantic structure of first-order logic (FOL). Also, we will use *variables* to denote players, and then the current position of a player gives us the value of the corresponding variable. So, positions of all players can give us an *assignment* function  $\sigma$  that assigns values to variables, say,  $\sigma(x) = a$  when player  $x$  is at  $a$ .

**Remark 1.** *The idea above to define the logic seems similar to FOL, but there is a crucial difference concerning the usages of variables. In FOL, variables are just placeholders without any intrinsic meaning, while variables in our proposal denote positions of players and can take different values in different situations of a game. This use of variables aligns with that in some recent dependence logics (e.g., [8, 47]) and in, e.g., physics (for instance, usually we use “ $v$ ” for velocity).*

We fix a *vocabulary*  $Voc = (Pred, Cons, Var)$ , where  $Pred$  is a set of predicate symbols, containing a specific binary relation symbol  $R$  describing the arrows of a game graph,  $Cons$  is a non-empty, finite set of constants, and  $Var = \{x, y\}$ , meaning the players. As usual, elements of  $Term = Cons \cup Var$  are called *terms*. The language  $\mathcal{L}$  for the *Epistemic Logic of Cops and Robbers* (ELCR) is defined in the following:

**Definition 1.** *Formulas in the language  $\mathcal{L}$  for ELCR are defined as follows:*

$$\begin{aligned}\mathcal{L}_B \ni \alpha &::= Pt \mid t_1 \equiv t_2 \mid \neg\alpha \mid (\alpha \wedge \alpha) \\ \mathcal{L}_{BD} \ni \psi &::= \alpha \mid \neg\psi \mid (\psi \wedge \psi) \mid [z]\psi \\ \mathcal{L} \ni \varphi &::= \psi \mid K_z t \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_z \psi \mid [z]\varphi\end{aligned}$$

where  $t, t_1, t_2 \in Cons \cup Var$  are terms,  $\mathbf{t}$  is a tuple of terms,  $P \in Pred$  is a predicate symbol, and  $z \in Var = \{x, y\}$  is a variable. Other Boolean connectives  $\top, \perp, \vee, \rightarrow, \leftrightarrow$  are defined as usual. We use  $\langle K_z \rangle \varphi$  and  $\langle z \rangle \varphi$  for  $\neg K_z \neg \varphi$  and  $\neg [z] \neg \varphi$ , respectively. Also, for a set  $\mathcal{T} \subseteq Cons \cup Var$ , we use  $K_z \mathcal{T}$  for  $\bigwedge_{t \in \mathcal{T}} K_z t$ .

In the definition, one can merge  $\mathcal{L}_{BD}$  and  $\mathcal{L}_B$  by adding the dynamic operators to  $\mathcal{L}_B$  directly, but to make it easier to reference the static Boolean formulas, we stick to the current form of the definition. We write  $\mathcal{L}^-$  for the fragment of  $\mathcal{L}$  without dynamic operators and  $ELCR^-$  for the corresponding logic. As usual,  $Pt$  and  $t_1 \equiv t_2$  are *atomic formulas* (in particular, formula  $Rt_1 t_2$  means *the value of  $t_2$  is a successor of the value of  $t_1$* ),  $K_z t$  reads *player  $z$  knows the value of  $t$* ,  $K_z \varphi$  means  *$z$  knows that  $\varphi$* , and  $[z]\varphi$  expresses *after any movement of  $z$ ,  $\varphi$  is the case*. The language does not contain

---

<sup>6</sup>Although we do not use probabilities explicitly, how players update their knowledge in our setting in effect is involved with probabilities in an implicit way, which are determined by, e.g., how many positions of a player are considered to be possible by the other, how many successors those possible positions have and what a player can see directly. In line with the implicitness, we will propose a qualitative approach to reason about the current game in Section 3.

formulas for higher-order knowledge (e.g.,  $K_x K_y c$ ). To define both the changes of knowledge for positions and the winning positions in the game, we even do not need the formulas of the form  $K_z \psi$ , but we add them to the language for convenience.

**Definition 2.** A model for ELCR is a tuple  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$ , where

- $\mathbf{D}$  is a non-empty, finite set of values (also called vertices or positions).
- $\mathbf{I}$  is the interpretation function such that
  - For each  $m$ -ary predicate symbol  $P \in \text{Pred}$ ,  $\mathbf{I}(P) \subseteq \mathbf{D}^m$  is an  $m$ -ary relation on  $\mathbf{D}$ . In particular,  $\mathbf{I}(R)$  is a binary relation on  $\mathbf{D}$  such that for any  $s \in \mathbf{D}$ , there is some  $t \in \mathbf{D}$  such that  $(s, t) \in \mathbf{I}(R)$ .
  - For each  $c \in \text{Cons}$ ,  $\mathbf{I}(c) \in \mathbf{D}$ . Moreover, for each  $s \in \mathbf{D}$ , there is  $c \in \text{Cons}$  such that  $\mathbf{I}(c) = s$ .
- $\Sigma \subseteq \mathbf{D}^{\text{Var}}$  is a non-empty set of situations of the players' positions, which are also called assignments.
- For each  $z \in \{x, y\}$ ,  $\sim_z \subseteq \Sigma \times \Sigma$  is an equivalence relation.

When the interpretation function  $\mathbf{I}$  is clear, we often write  $\mathbf{P}$  for  $\mathbf{I}(P)$ . Intuitively,  $(\mathbf{D}, \mathbf{R})$  represents a graph where a game is played, and by definition,  $\mathbf{R}$  is serial;  $\Sigma$  is a collection of possible situations of a game; and the relations  $\sim_{z \in \text{Var}}$  are the indistinguishability relations of players (e.g.,  $\sigma_1 \sim_x \sigma_2$  means that player  $x$  cannot distinguish between situations  $\sigma_1$  and  $\sigma_2$ ).

The class of all models captures the case where players do not have any ability, not even the 0-sight: for instance, it might be the case that  $\sigma_1 \sim_x \sigma_2$  and  $\sigma_1(x) \neq \sigma_2(x)$ , which means that  $x$  does not know where herself is. To capture the  $k$ -sight ability, we first define an auxiliary notion as follows:

**Definition 3.** Let  $(\mathbf{D}, \mathbf{R})$  be a finite graph. For any  $s \in \mathbf{D}$  and  $m \in \mathbb{N}$ , we inductively define the following:

$$\mathbb{D}^0(s) := \{s\}$$

$$\mathbb{D}^{m+1}(s) := \mathbb{D}^m(s) \cup \{t \in \mathbf{D} : \exists u \in \mathbb{D}^m(s) \text{ s.t. } (u, t) \in \mathbf{R} \text{ or } (t, u) \in \mathbf{R}\}.$$

So,  $t \in \mathbb{D}^k(s)$  states that the 'distance' between  $s$  and  $t$  is not more than  $k$ , and more precisely,  $t$  can be reached from  $s$  within  $k$  steps via the symmetric closure of  $\mathbf{R}$ .

**Definition 4.** A  $k$ -sight model is a model  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  such that for each  $z \in \text{Var}$  and  $\sigma, \sigma' \in \Sigma$ , if  $\sigma \sim_z \sigma'$ , then for any  $z' \in \text{Var}$  with  $\sigma(z') \in \mathbb{D}^k(\sigma(z))$ ,  $\sigma(z') = \sigma'(z')$ .

In a  $k$ -sight model, if the distance between the position of a player  $z$  and the position of player  $z'$  is not more than  $k$ , then in all situations that cannot be distinguished by  $z$ , player  $z'$  always has the same position, which means that  $z$  knows the position of  $z'$ . This characterizes the  $k$ -sight ability. In what follows, we work with  $k$ -sight models, and ELCR is defined based on them.

**Remark 2.** The notion of  $k$ -sight models can be adapted to capture cases that players have different sights. For each  $z \in \{x, y\}$ , we use  $k_z$  for the sight of the player. Then, to characterize this more complicated setting, we just need to replace the restriction imposed on  $k$ -sight models with the following: for any  $z \in \text{Var}$  and  $\sigma, \sigma' \in \Sigma$ ,

$$\text{if } \sigma \sim_z \sigma', \text{ then for any } z' \in \text{Var} \text{ with } \sigma(z') \in \mathbb{D}^{k_z}(\sigma(z)), \sigma(z') = \sigma'(z').$$

Given a model  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  and  $\sigma \in \Sigma$ , for any  $t \in \mathbf{Term}$ , we use  $t^{(\mathbf{I}, \sigma)}$  for the value of  $t$ : if  $t \in \mathbf{Cons}$ , then  $t^{(\mathbf{I}, \sigma)} := \mathbf{I}(t)$ ; and if  $t \in \mathbf{Var}$ , then  $t^{(\mathbf{I}, \sigma)} := \sigma(t)$ .

Now we move to presenting the semantics for the logic. Truth conditions for the fragment  $\mathcal{L}^-$  are straightforward, and key clauses are as follows:

$$\begin{aligned} M, \sigma \models P(t_1, \dots, t_n) & \text{ iff } (t_1^{(\mathbf{I}, \sigma)}, \dots, t_n^{(\mathbf{I}, \sigma)}) \in \mathbf{I}(P) \\ M, \sigma \models t_1 \equiv t_2 & \text{ iff } t_1^{(\mathbf{I}, \sigma)} = t_2^{(\mathbf{I}, \sigma)} \\ M, \sigma \models K_z t & \text{ iff for all } \sigma' \in \Sigma, \text{ if } \sigma' \sim_z \sigma, \text{ then } t^{(\mathbf{I}, \sigma)} = t^{(\mathbf{I}, \sigma')} \\ M, \sigma \models K_z \varphi & \text{ iff for all } \sigma' \in \Sigma, \text{ if } \sigma' \sim_z \sigma, \text{ then } M, \sigma' \models \varphi \end{aligned}$$

Recall that in our models  $M$ , for each vertex, there is a constant true at the vertex. So, given a situation  $\sigma$ , for any term  $t \in \mathbf{Term}$ , there is a constant  $c$  such that  $t \equiv c$  at  $\sigma$  (if  $t$  is a constant, then  $t \equiv c$  is globally true in  $\Sigma$ ). Based on this,  $K_z t$  amounts to  $\bigwedge_{c \in \mathbf{Cons}} ((K_z) t \equiv c \rightarrow K_z t \equiv c)$ , but we stick to using  $K_z t$  for syntactic succinctness.

It remains to show the truth condition for  $[z]\varphi$ . In terms of the game, a desired clause would give us an automatic mechanism to capture the effects of movements, which are involved with both the changes of positions and the updates of the knowledge. Let us first define binary relations  $\mathbf{R}^{z \in \mathbf{Var}}$  on *all* assignments  $\mathbf{D}^{\mathbf{Var}}$ :

$$\text{For any } \sigma, \sigma' \in \mathbf{D}^{\mathbf{Var}}, \text{ we write } \mathbf{R}^z \sigma \sigma' \text{ if } (\sigma(z), \sigma'(z)) \in \mathbf{R} \text{ and for } z' \in \mathbf{Var} \setminus \{z\}, \\ \sigma(z') = \sigma'(z').$$

Therefore, when  $\mathbf{R}^z \sigma \sigma'$ , the only difference of  $\sigma$  and  $\sigma'$  concerns the values of  $z$ . It intuitively describes the fact that after  $z$  moves from  $\sigma(z)$  to  $\sigma'(z)$ , the situation  $\sigma$  becomes  $\sigma'$ . Given a class  $\Sigma \subseteq \mathbf{D}^{\mathbf{Var}}$  of situations and a variable  $z$ , we define

$$\mathbf{R}^z(\Sigma) := \{\sigma' \mid \text{there is } \sigma \in \Sigma \text{ s.t. } \mathbf{R}^z \sigma \sigma'\}.$$

When  $\Sigma$  is a singleton  $\{\sigma\}$ , we write  $\mathbf{R}^z(\sigma)$  for  $\mathbf{R}^z(\{\sigma\})$ . Also, we use  $\Sigma|\sigma$  to mean the subset  $\Sigma' = \{\sigma' \in \Sigma \mid \sigma \sim_z \sigma' \text{ for some } z \in \mathbf{Var}\}$ . Now we can show the truth condition for the movement operators  $[z]\varphi$ :<sup>7</sup>

**Definition 5.** *For the case that both the players have the same sight  $k$ , the truth condition for  $[z]\varphi$  is as follows:*

$$(\mathbf{D}, \mathbf{I}, \Sigma, \sim), \sigma_1 \models [z]\varphi \text{ iff for all } \sigma_2 \in \mathbf{R}^z(\sigma_1), (\mathbf{D}, \mathbf{I}, \Sigma', \sim'), \sigma_2 \models \varphi,$$

where the new  $\Sigma'$  is given by the following:

- (a) if  $\sigma_2(x) \in \mathbb{D}^k(\sigma_2(y))$ , then  $\Sigma' = \{\sigma_2\}$ ,
- (b) if  $\sigma_2(x) \notin \mathbb{D}^k(\sigma_2(y))$ , then  $\Sigma' = \{\sigma' \in \mathbf{R}^z(\Sigma|\sigma_1) \mid \sigma'(x) \notin \mathbb{D}^k(\sigma'(y))\}$ ,

and the new relations  $\sim'_{z \in \{x, y\}}$  on  $\Sigma'$  are obtained by the following:

- (c)  $\sigma'_1 \sim'_z \sigma'_2$  iff  $\sigma'_1(z) = \sigma'_2(z)$ .

The clause (a) aims to deal with the case that after the movement, the two players are in the sight of each other. In this case, both of them know the actual situation. Moreover, the clause (b) tackles the case that after the movement, the two players are not in the sight of each other. Different from that of (a), players only consider the situations in which they are not in the sight of each other to be possible.

With the definition above, one can see that different  $\sigma_2 \in \mathbf{R}^z(\sigma_1)$  can give us different  $\Sigma'$ . It always holds that  $\Sigma' \subseteq \mathbf{R}^z(\Sigma)$  and  $\sigma_2 \in \Sigma'$ . When  $M$  is a  $k$ -sight

<sup>7</sup>The authors would like to thank Alexandru Baltag for a useful discussion on the notion of the update.

model, the resulting model obtained by the updating  $M$  is again a  $k$ -sight model. We will provide some examples for the updates further in Section 4. In what follows, we will use some ordinary notions directly, including *satisfiability*, *validity* and *logical consequence*, which can be defined in the usual manner.

In the remainder of the section, we provide some comments on the update mechanism and offer the method of updates for *the simultaneous movements*. For convenience, we will often use tuples of values to specify the positions of  $x, y$  (in this order), to denote situations. For instance, we write  $\sigma$  as  $(s_1, s_2)$  if  $\sigma(x) = s_1$  and  $\sigma(y) = s_2$ . Also, we will highlight the actual situation with an underline, e.g.,  $(\underline{s_1}, s_2)$ .

**Remark 3.** *In the clause (b) of Definition 5, we require  $\Sigma'$  is a subset of  $\mathbf{R}^z(\Sigma|\sigma)$ , although it looks natural to just require it to be a subset of  $\mathbf{R}^z(\Sigma)$ . However, restricting our attention to  $\mathbf{R}^z(\Sigma|\sigma)$  is vital to ensure that the update would not cause disorders. To see this, let us consider an example where we just require  $\Sigma'$  to be a subset of  $\mathbf{R}^z(\Sigma)$ .*



We consider  $\Sigma = \{(s_1, s_4), (s_2, s_5)\}$ , where  $(s_1, s_4)$  is the actual situation. Also, we assume that both  $x$  and  $y$  have sight 1, but due to some reason they happen to be able to distinguish between the two situations (so both  $K_x y$  and  $K_y x$  hold at  $(s_1, s_4)$ ).

Now, let  $x$  move, after which the only possible situation is  $(s_2, s_4)$ . However,

$$\Sigma' = \{\sigma' \in \mathbf{R}^x(\Sigma) \mid \sigma'(x) \notin \mathbb{D}^k(\sigma'(y))\} = \{(\underline{s_2}, s_4), (s_2, s_5), (s_3, s_5)\}$$

At the new situation  $(s_2, s_4)$ , by our clause for the epistemic relations, we still have  $K_y x$ , but  $x$  now cannot distinguish between  $(s_2, s_4)$  and  $(s_2, s_5)$ , i.e.,  $x$  forgot the position of  $y$  immediately after the movement of  $x$  herself! However, this should not be the case in our setting: see Fact 6.<sup>8</sup>

Finally, it is worth noting that the difference between the two requirements just make sense for the initial updates of a given model, in that for the resulting  $\Sigma'$  after any update associated to the new situation  $\sigma'$ , it is always the case that  $\Sigma'|\sigma' = \Sigma'$ .

**Digression: Simultaneous movements.** Let us end this part by a brief discussion on the logic for the setting that players  $x, y$  move simultaneously. To capture the effects of the new pattern, it is crucial to define suitable ‘group movement operators’  $[Var]\varphi$ , and as what we are going to show, such a framework can be easily obtained by generalizing the idea behind the updates induced by  $[z]$ . Let us define  $\mathbf{R}^{Var}$  such that

$$\text{for any } \sigma, \sigma' \in \mathbf{D}^{Var}, \mathbf{R}^{Var} \sigma \sigma' \text{ iff for any } z \in Var, (\sigma(z), \sigma'(z)) \in \mathbf{R},$$

which captures the changes of situations caused by the simultaneous movements of both  $x$  and  $y$ . For a set  $\Sigma \subseteq \mathbf{D}^{Var}$ , we can define

$$\mathbf{R}^{Var}(\Sigma) := \{\sigma' \mid \text{there is } \sigma \in \Sigma \text{ s.t. } \mathbf{R}^{Var} \sigma \sigma'\},$$

and again, when  $\Sigma$  is a singleton  $\{\sigma\}$ , we write  $\mathbf{R}^{Var}(\sigma)$  for  $\mathbf{R}^{Var}(\{\sigma\})$ . Now, the truth condition for  $[Var]\varphi$  is given by the following:

$$(\mathbf{D}, \mathbf{I}, \Sigma, \sim), \sigma_1 \models [Var]\varphi \quad \text{iff} \quad \text{for all } \sigma_2 \in \mathbf{R}^{Var}(\sigma_1), (\mathbf{D}, \mathbf{I}, \Sigma', \sim'), \sigma_2 \models \varphi,$$

---

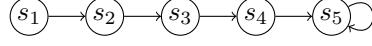
<sup>8</sup>It is important to study the case that players have only *restricted memory* [39, 25], and we leave this to future inquiry.

where  $\Sigma'$  and  $\sim'$  are given by the same clauses as that in Definition 5, except we now need to use  $R^{Var}$  instead of  $R^z$ . The logic for the simultaneous movements is interesting in its own right, as suggested by the following:

**Fact 1.** *Let  $\mathcal{L}^+$  be the language extending  $\mathcal{L}$  with  $[Var]\varphi$ . The following is not valid:*

$$[x][y]\varphi \leftrightarrow [Var]\varphi. \text{ }^9$$

*Proof.* It suffices to find an instance. Consider the following model:

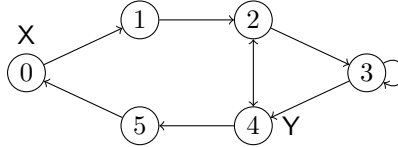


Assume that both  $x, y$  have the sight 1, and the actual situation is  $(s_1, s_3)$ . Also, the original  $\Sigma$  and the indistinguishability relations are given by the 1-sight ability and the actual situation of their positions. Now,  $[x][y]K_x y$  holds, but  $[Var]K_x y$  fails.  $\square$

As shown by the example in the proof above, in  $[x][y]\varphi$ , a first movement  $[x]$  may make  $x$  know more about the position of  $y$ , but  $[Var]$  leaves no room for  $x$  to know that. In what follows, we continue to study ELCR and leave the exploration on the simultaneous variant to another occasion.

## 4 Applications of ELCR to Cops and Robbers

Now we will use the formal framework to track players' knowledge, to describe the winning conditions for players, among others. In this part, we assume that the round-restriction  $n \in \mathbb{N}$  imposed on the winning condition is big enough.



Let us revisit the game in Example 1. We will write  $x$  and  $y$  for Cop and Robber respectively. Recall that they have the sight 1 and that in each round,  $x$  acts first. As before, we write tuples for situations and usually highlight actual situations with underlines. For the example, we denote by  $\sigma_0 = (0, 4)$  the initial situation.

Based on the 1-sight ability,  $x$  knows her own position. Also,  $x$  knows that  $y$  is not at 0, 1 or 5, but at 2, 3 or 4. So, for player  $x$ , all of the following situations are possible:

$$x : \quad \underline{(0, 4)}, (0, 2), (0, 3)$$

For  $y$ , he knows where himself is, but he cannot distinguish between the following:

$$y : \quad \underline{(0, 4)}, (1, 4)$$

<sup>9</sup>The order of  $[x]$  and  $[y]$  in the equivalence does not matter, i.e.,  $[y][x]\varphi \leftrightarrow [Var]\varphi$  is not valid as well.

All these aforementioned situations together form the set  $\Sigma_0$  of possible situations of any models matching the game. The indistinguishability relations  $\sim_x$  and  $\sim_y$  among these situations are as described above.

**Remark 4.** *Here one can see that it is crucial to restrict ourselves to the setting without higher-order knowledge: if higher-order knowledge is permitted, then  $\langle K_x \rangle K_y x$  holds at the actual situation, but w.r.t. the game, the formula is not a correct description the knowledge of  $x$ . Generalized proposals for settings with higher-order knowledge exist, but we leave these for future work.*

Now, the game begins and player  $x$  has only one option:

$x$  moves to 1. The actual situation becomes  $(1, 4)$ , written  $\sigma_1$ . Since they are not in the sight of each other, by the definition of the update, we can obtain the following:

$$\Sigma_1 = \{\sigma \in \mathbb{R}^x(\Sigma_0|\sigma_0) \mid \sigma(x) \notin \mathbb{D}^1(\sigma(y))\} = \{(1, 4), (1, 3)\}.$$

Here  $\Sigma_0|\sigma_0$  is exactly  $\Sigma_0$ . Again, by the definition of updates, the indistinguishability relations among the situations are as follows:

$$x : \quad \boxed{(1, 4), (1, 3)} \qquad y : \quad \boxed{(1, 4)}$$

Recall the situations that cannot be distinguished by  $x$  in the round 0. She now does not consider the case that  $y$  is at 2 to be possible: in the actual situation  $(1, 4)$  she cannot see directly  $y$ , but  $(1, 2)$  is such a case. Besides, for  $y$ , since 0 is not a successor of any previous possibilities 0, 1 of the position of  $x$  considered by him,  $y$  does not consider the case  $(0, 4)$  to be possible now, which is formally captured by the requirement that  $\Sigma_1$  should be a subset of  $\mathbb{R}^x(\Sigma_0|\sigma_0)$ .

Let us now move to the next stage:

$y$  moves to vertex 5. We write  $\sigma_2$  for the new situation  $(1, 5)$ . Now the players are not in the sight of each other as well, and the new class  $\Sigma_2$  of situations and the indistinguishability relations given by our update policy are as follows:

$$\Sigma_2 = \{\sigma \in \mathbb{R}^y(\Sigma_1|\sigma_1) \mid \sigma(x) \notin \mathbb{D}^1(\sigma(y))\} = \{(1, 5), (1, 3), (1, 4)\}.$$

$$x : \quad \boxed{(1, 5), (1, 3), (1, 4)} \qquad y : \quad \boxed{(1, 5)}$$

Player  $y$  still knows the actual situation, and player  $x$  considers the unobservable successors of previous possibilities 3, 4 to be possible at this new stage.

Finally, the game ends by the following step:

$x$  moves to 2. Now the actual situation, i.e.,  $(2, 5)$ , is the only possibility considered by each of the players, which can be seen from the following:

$$\{\sigma \in \mathbb{R}^x(\Sigma_2|\sigma_2) \mid \sigma(x) \notin \mathbb{D}^1(\sigma(y))\} = \{(2, 5)\}.$$

Although the final  $\{(2, 5)\}$  is a singleton set, it is obtained not by the clause (a) in Definition 5, but by the clause (b): this indicates that the realization of the actual situation depends on their indirect reasoning (with the knowledge about the graph structure and the  $k$ -sight ability), but not on the direct observation.

Generally speaking, the language can also be used to verify whether a player has a winning strategy. For instance, when the round-restriction  $n$  is 2, Cop has a winning strategy iff the following is true at the initial situation:

$$K_x y \vee \langle x \rangle K_x y \vee \langle x \rangle [y] K_x y \vee \langle x \rangle [y] \langle x \rangle K_x y \vee \langle x \rangle [y] \langle x \rangle [y] K_x y. \quad {}^{10}$$

This suggests another potential application of the logic to the game: the model-checking problem for the formula of ELCR corresponds to the verification of existence of winning strategies of the players, and a complexity study would provide us an upper bound of the complexity of deciding this imperfect information game. As proved in [34], the static fragment  $\text{ELCR}^-$  has a P-complete model-checking problem, and it remains to determine the exact complexity of the model-checking problem for ELCR.

## 5 Some properties of ELCR

Having seen the basics of the logic, we now turn to exploring some of its properties, involving validities and the effects of dynamic operators. Let us start by pointing out that many valid schemata of FOL fail in this new setting. For instance, formula  $t_1 \equiv t_2 \rightarrow (\varphi \leftrightarrow \varphi[t_1/t_2])$  is *not* a validity of ELCR, where  $\varphi[t_1/t_2]$  is obtained by replacing  $t_2$  in  $\varphi$  with  $t_1$ .<sup>11</sup> But when we restrict the formula  $\varphi$  to those not involving knowledge, the schema still holds.

For any natural number  $n \in \mathbb{N}$  and terms  $t_1, t_2$ , we define the following, which expresses the distance between  $t_1$  and  $t_2$ :

$$\begin{aligned} D^0 t_1 t_2 &:= t_1 \equiv t_2, \\ D^{n+1} t_1 t_2 &:= D^n t_1 t_2 \vee \bigvee_{t \in \text{Var} \cup \text{Cons}} (D^n t_1 t \wedge (R t t_2 \vee R t_2 t)). \end{aligned}$$

Since  $\text{Var} \cup \text{Cons}$  is finite, the formulas above are well-defined. Now  $D^k z t$  means that the position of  $t$  is in the sight of  $z$ .

**Fact 2.** *The following formulas are valid w.r.t.  $k$ -sight models:*

- (1)  $D^k z t \rightarrow K_z t$ , given that  $z \in \text{Var}$ .
- (2)  $(K_z \mathcal{T} \wedge P(t_1, t_2, \dots, t_m)) \rightarrow K_z P(t_1, t_2, \dots, t_m)$ , given that  $\{t_1, \dots, t_m\} \subseteq \mathcal{T}$ .

Formula (1) means if  $t$  is in the sight of  $z$ , then  $z$  knows where  $t$  is, and (2) indicates that knowing the values of  $\mathcal{T}$  means knowing the atomic facts involving  $\mathcal{T}$ : when  $P$  is  $R$ , it characterizes the assumption that *players know the structures of game graphs*.

In terms of the validities of knowledge operators  $K_z \alpha \in \mathcal{L}$ , although  $K_z \alpha \rightarrow \alpha$  is always the case, we do not have the ordinary  $K_z \alpha \rightarrow K_z K_z \alpha$  or  $\neg K_z \alpha \rightarrow K_z \neg K_z \alpha$  for *the positive reflection* and *the negative reflection*, respectively, since syntactically we restrict ourselves to the setting without high-order knowledge. However, we can *mimic* them at the semantic level, in the sense of the following:

**Fact 3.** *Let  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$ ,  $\sigma \in \Sigma$  and  $K_z \alpha \in \mathcal{L}$ .*

- (1)  $M, \sigma \models K_z \alpha$  iff for any  $\sigma' \in \Sigma$  such that  $\sigma \sim_z \sigma'$ ,  $M, \sigma' \models K_z \alpha$ .
- (2)  $M, \sigma \models \neg K_z \alpha$  iff for any  $\sigma' \in \Sigma$  such that  $\sigma \sim_z \sigma'$ ,  $M, \sigma' \models \neg K_z \alpha$ .
- (3) *As a consequence, for any  $\varphi$  of the form  $K_z \alpha_1 \wedge \dots \wedge K_z \alpha_n \wedge \langle K_z \rangle \beta_1 \wedge \dots \wedge \langle K_z \rangle \beta_m$  with  $1 \leq m + n$ ,  $M, \sigma \models \varphi$  iff for any  $\sigma' \in \Sigma$  s.t.  $\sigma \sim_z \sigma'$ ,  $M, \sigma' \models \varphi$ .*

<sup>10</sup>For the setting without any round restrictions, the syntactic length would become infinite, which makes the formula not well-defined. To analyze this, one may need some enhanced tools such as *the modal logic for substitution* [46] that is motivated by the perfect information version of the game [36].

<sup>11</sup>It is not hard to find a counterexample to, e.g.,  $c \equiv y \rightarrow (K_x c \leftrightarrow K_x y)$ :  $y$  is at  $c$ ,  $x$  knows the value of  $c$ , but  $x$  may not know where  $y$  is.

*Proof.* The key reason for these is that  $\sim_z$  is an equivalence. We skip the details.  $\square$

Also, by induction on  $\mathcal{L}_B$ -formulas, we can show the following:

**Fact 4.** *Let  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  and  $M' = (\mathbf{D}, \mathbf{I}, \Sigma', \sim')$  be models, and  $\alpha \in \mathcal{L}_B$ . Let  $\sigma \in \Sigma, \sigma' \in \Sigma'$  s.t. for any variable  $z$  occurring in  $\alpha$ ,  $\sigma(z) = \sigma'(z)$ . Then, it holds that:*

$$M, \sigma \models \alpha \text{ iff } M', \sigma' \models \alpha.$$

In what follows, for any  $\mathcal{T} \subseteq \text{Cons}$  and  $u \in \text{Term}$ , we define the following:

$$Ru = \mathcal{T} := \bigwedge_{t \in \mathcal{T}} Rut \wedge \bigwedge_{t \in \text{Cons} \setminus \mathcal{T}} \neg Rut$$

expressing that  $\mathcal{T}$  is exactly the set of (the names of)  $\mathbf{R}$ -successors of  $z$ . For instance,  $Rc_1 = \{c_2\}$  means that via the relation  $\mathbf{R}$ ,  $c_1$  can only reach  $c_2$  but not all other constants. Now we have the following about the effects of updates on  $\mathcal{L}_B$ -formulas:

**Fact 5.** *Let  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  be a model,  $\sigma_1 \in \Sigma$ . For any  $\alpha \in \mathcal{L}_B$  and  $z \in \{x, y\}$ ,*

*$M, \sigma_1 \models \bigwedge_{\mathcal{T} \subseteq \text{Cons}} (Rz = \mathcal{T} \rightarrow \bigwedge_{c \in \mathcal{T}} \alpha[c/z])$  iff for all  $\sigma_2 \in R^z(\sigma_1)$ ,  $M', \sigma_2 \models \alpha$  where  $M' = (\mathbf{D}, \mathbf{I}, \Sigma', \sim')$  is the update of  $M$  (associated to  $\sigma_2$ ) induced by  $[z]$ .*

*Proof.* Notice that the  $\mathcal{T}$  making  $Rz = \mathcal{T}$  true is exactly  $\{c \in \text{Cons} \mid M, \sigma_1 \models Rz c\}$ . Given this  $\mathcal{T}$ , we can fix  $R^z(\sigma_1)$  as  $\{\sigma_1[z := \mathbf{I}(c)] \mid c \in \mathcal{T}\}$ , where  $\sigma_1[z := c]$  might be different from  $\sigma_1$  on the value of  $z$ : the former assigns the value of  $c$  to  $z$ . Now, the fact can be proved by induction on formulas. The cases for Boolean connectives are trivial. The proofs for  $Pt$  and  $t_1 \equiv t_2$  are similar, and we just consider the former.

$$\begin{aligned} M, \sigma_1 \models \bigwedge_{c \in \mathcal{T}} Pt[c/z] & \quad \text{iff} & \quad \text{for all } c \in \mathcal{T}, M, \sigma_1[z := \mathbf{I}(c)] \models Pt \\ & \quad \text{iff} & \quad \text{for all } \sigma_2 \in R^z(\sigma_1), M', \sigma_2 \models Pt \end{aligned}$$

The second holds by the connection between  $\mathcal{T}$  and  $R^z(\sigma_1)$  and the fact that all of  $\mathbf{D}$ ,  $\mathbf{I}$  and the value of the other variable are the same in  $M, \sigma_1$  and  $M', \sigma_2$ .  $\square$

While the above preliminaries concern the static part  $\mathcal{L}^-$ , the result below is involved with the dynamic operators, which means that after a movement of a player, the player still remembers what was known in the previous situation (Remark 3):

**Fact 6.** *Let  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  be a model,  $\sigma \in \Sigma$ , and  $\{z, z'\} = \{x, y\}$ . If  $M, \sigma \models K_z z'$ , then  $M, \sigma \models [z]K_z z'$ .*

*Proof.* W.l.o.g., let  $z := x$  and  $z' := y$ . Suppose  $M, \sigma \models K_x y$ . Now, let  $\sigma' \in R^x(\sigma)$  and we consider the updated model  $M'$  associated to  $\sigma'$ . There are different cases.

First, if  $\sigma'(x) \in \mathbb{D}^k(\sigma'(y))$ , then the  $\Sigma'$  of  $M'$  is  $\{\sigma'\}$ , which gives us  $M', \sigma' \models K_x y$ .

Next, assume that  $\sigma'(x) \notin \mathbb{D}^k(\sigma'(y))$ . Now, let  $\sigma'_1 \in \Sigma'$  s.t.  $\sigma' \sim'_x \sigma'_1$ . Given  $\sigma'_1 \in \Sigma'$ , there is some  $\sigma_1 \in \Sigma \mid \sigma$  s.t.  $\sigma'_1 \in R^x(\sigma_1)$ . Notice that  $\sigma_1(y) = \sigma'_1(y)$  and  $\sigma(y) = \sigma'(y)$ . So, it suffices to show that  $\sigma_1(y) = \sigma(y)$ , which then can give us  $\sigma'_1(y) = \sigma'(y)$ . Since  $\sigma_1 \in \Sigma \mid \sigma$ , we have  $\sigma_1 \sim_x \sigma$  or  $\sigma_1 \sim_y \sigma$ : if  $\sigma_1 \sim_x \sigma$ , then by  $M, \sigma \models K_x y$ , it holds that  $\sigma_1(y) = \sigma(y)$ , otherwise by the  $0 \leq k$ -sight ability,  $\sigma_1(y) = \sigma(y)$  holds as well.  $\square$

## 6 Axiomatization of the static fragment: $\text{ELCR}^-$

Having seen the basic properties of the logic, we will proceed to provide a Hilbert-style calculus for  $\text{ELCR}$  and show that it has a decidable satisfiability problem. In this

section we will first consider the static part without action modalities  $[x]$  or  $[y]$ . In what follows, we use  $z, z'$  for variables in  $\{x, y\}$ . Also, for any  $\mathcal{T} \subseteq \text{Cons}$ , we define

$$K_z z' = \mathcal{T} := \bigwedge_{c \in \mathcal{T}} \langle K_z \rangle z' \equiv c \wedge \bigwedge_{c \in \text{Cons} \setminus \mathcal{T}} K_z \neg z' \equiv c$$

expressing that  $\mathcal{T}$  consists of the possible positions of  $z'$  considered by  $z$ .

**Table 1** Proof system **ELCR**<sup>-</sup> for the static part **ELCR**<sup>-</sup>.

<b>I:</b> General axioms and rules for Boolean connectives and $\equiv$	
(Tau)	Propositional tautologies
(A1)	$t_1 \equiv t_1$
(A2)	$t_1 \equiv t_2 \rightarrow t_2 \equiv t_1$
(A3)	$t_1 \equiv t_2 \wedge t_2 \equiv t_3 \rightarrow t_1 \equiv t_3$
(A4)	$t_1 \equiv t_2 \rightarrow (\alpha \leftrightarrow \alpha[t_1/t_2])$ given that $\alpha \in \mathcal{L}_B$ .
(MP)	From $\varphi \rightarrow \psi$ and $\varphi$ , infer $\psi$ .
<b>II:</b> Axioms for basics of the games	
(Seriality)	$\bigvee_{t \in \text{Cons}} Rct$
(At-Some-Where)	$\bigvee_{c \in \text{Cons}} z \equiv c$ , given $z \in \{x, y\}$
(k-sight)	$D^k zt \rightarrow K_z t$ , where $z \in \{x, y\}$ and $t \in \text{Term}$ .
<b>III:</b> Axioms and rules for $K_z \alpha$	
(K)	$K_z(\alpha \rightarrow \beta) \rightarrow (K_z \alpha \rightarrow K_z \beta)$ , where $\alpha, \beta \in \mathcal{L}_B$ .
(T)	$K_z \alpha \rightarrow \alpha$ , where $\alpha \in \mathcal{L}_B$ .
(Knowledge-Ground)	$K_z z' = \mathcal{T} \rightarrow (K_z \alpha \leftrightarrow \bigwedge_{c \in \mathcal{T}} \alpha[c/z'])$ , given $\mathcal{T} \subseteq \text{Cons}$ , $\{z, z'\} = \{x, y\}$ and $\alpha \in \mathcal{L}_B$ .
(K-Additivity)	From $\varphi \rightarrow \alpha$ , infer $\varphi \rightarrow K_z \alpha$ , where $\varphi$ is of the form $K_z \alpha_1 \wedge \dots \wedge K_z \alpha_n \wedge \langle K_z \rangle \beta_1 \wedge \dots \wedge \langle K_z \rangle \beta_m$ s.t. $1 \leq m + n$ and $\alpha, \alpha_1 \leq i \leq n, \beta_1 \leq i \leq m \in \mathcal{L}_B$ .
(K-Elimination)	From $\varphi \rightarrow (K_z \alpha \rightarrow \beta)$ , infer $\varphi \rightarrow (\alpha \rightarrow \beta)$ , where $\varphi$ has the form $K_{z'} \alpha_1 \wedge \dots \wedge K_{z'} \alpha_n \wedge \langle K_{z'} \rangle \beta_1 \wedge \dots \wedge \langle K_{z'} \rangle \beta_m$ with $1 \leq m + n$ , $\alpha, \beta, \alpha_1 \leq i \leq n, \beta_1 \leq i \leq m \in \mathcal{L}_B$ and $\{z, z'\} = \{x, y\}$ .
<b>IV:</b> Interaction axioms for $K_z \alpha$ and $K_z t$	
(De-Re-Knowledge)	$t \equiv c \rightarrow (K_z t \leftrightarrow K_z c)$ , where $z \in \{x, y\}$ and $t \in \text{Term}$ .
(Structure-Knowledge)	$(K_z \mathcal{T} \wedge \alpha(t_1, t_2, \dots, t_m)) \rightarrow K_z \alpha(t_1, t_2, \dots, t_m)$ , given that $\{t_1, \dots, t_m\} \subseteq \mathcal{T}$ and $\alpha \in \mathcal{L}_B$ .

The details of the proof system are given in Table 1. Notice that some axioms are redundant, but we keep them for convenience. Let us briefly comment on the axioms.

- In the part **II**, (**Seriality**) means a state in a graph always has **R**-successors, and (**At-Some-Where**) means that a player is always at some vertex in game graphs.
- The part **III** is about knowledge operators. Formula (**Knowledge-Ground**) lays out the foundation of knowledge: when  $\mathcal{T}$  is the ‘uncertainty scope’ of  $z$  regarding the position of the other player  $z'$ ,  $z$  knows  $\alpha$  iff every possible position of  $z'$  considered by  $z$  together with other relevant parameters can satisfy  $\alpha$ .
- The last part **IV** is about the interactions between  $K_z t$  and  $K_z \alpha$ . The axiom (**De-Re-Knowledge**), motivated by [4], states that when  $t$  and  $c$  have the same value,  $z$  knows the value of  $t$  iff  $z$  knows the fact that they have the same value. Finally, the axiom (**Structure-Knowledge**) means that when  $z$  knows the values of terms in a true statement  $\alpha$ ,  $z$  knows the fact that  $\alpha$ .

We write  $\vdash \varphi$  if  $\varphi$  is provable in the calculus  $\mathbf{ELCR}^-$ . Also, we write  $\Phi \vdash \varphi$  if there exist finitely many  $\varphi_1, \dots, \varphi_n \in \Phi$  such that  $\vdash \varphi_1 \wedge \dots \wedge \varphi_n \rightarrow \varphi$ , and when  $\Phi$  is a singleton, e.g.,  $\{\psi\}$ , we employ  $\psi \vdash \varphi$  for simplicity.

**Fact 7.**  $\mathbf{ELCR}^-$  is sound for  $\mathbf{ELCR}^-$ .

*Proof.* See Appendix A. □

**Fact 8.** The following formulas are provable and rules are derivable:

- (1)  $\vdash K_z z$  and  $\vdash K_z z \equiv z$ .
- (2) From  $\alpha$ , infer  $K_z \alpha$ , where  $\alpha \in \mathcal{L}_B$ . (Nec)
- (3)  $\vdash K_z c$  for any  $c \in \text{Cons}$ .
- (4)  $\vdash K_z \text{Cons} \cup \{z\}$
- (5)  $\vdash \alpha \rightarrow K_z \alpha$ , given that  $\alpha \in \mathcal{L}_B$  does not contain the other variable.
- (6)  $\vdash K_z z' \wedge \alpha \rightarrow K_z \alpha$ , given that  $\{z, z'\} = \{x, y\}$  and  $\alpha \in \mathcal{L}_B$ .
- (7) From  $K_z \alpha \rightarrow \beta$ , infer  $\alpha \rightarrow \beta$ , where  $\alpha, \beta \in \mathcal{L}_B$ .

*Proof.* We prove the first three items, and the others are omitted to save space.

(1) For the first item, it goes as follows:

- |     |   |                                       |
|-----|---|---------------------------------------|
| (1) | $z \equiv z \rightarrow (K_z z \leftrightarrow K_z z \equiv z)$ | (De-Re-Knowledge)                     |
| (2) | $z \equiv z$  | (A1)                                  |
| (3) | $z \equiv z \rightarrow D^k z z$                                | (by the definition of $D^k t_1 t_2$ ) |
| (4) | $D^k z z$   | (MP, 2, 3)                            |
| (5) | $D^k z z \rightarrow K_z z$                                     | ( $k$ -sight)                         |
| (6) | $K_z z$   | (MP, 4, 5)                            |
| (7) | $K_z z \equiv z$  | (propositional logic, 1, 2, 6)        |

(2). Assume that  $\alpha$  is the case. Then,  $K_z z \equiv z \rightarrow \alpha$ . So, by (K-Additivity), we have  $K_z z \equiv z \rightarrow K_z \alpha$ . By the first item,  $K_z z \equiv z$ . Using (MP), we can obtain  $K_z \alpha$ .

(3) For the third one, the details are as follows:

- |     |   |                   |
|-----|---|-------------------|
| (1) | $c \equiv c \rightarrow (K_z c \leftrightarrow K_z c \equiv c)$ | (De-Re-Knowledge) |
| (2) | $c \equiv c$  | (A1)              |
| (3) | $K_z c \leftrightarrow K_z c \equiv c$                          | (MP, 1, 2)        |
| (4) | $K_z c \equiv c$  | (Nec, 2)          |
| (5) | $K_z c$   | (MP, 3, 4)        |

This completes the proof. □

We say that a set  $\Delta$  of formulas is *inconsistent* if there exists a finite set  $\Gamma \subseteq \Delta$  such that  $\vdash \bigwedge \Gamma \rightarrow \perp$ , and that  $\Delta$  is *consistent* if it is not inconsistent. We also say that  $\Delta$  is a *maximally consistent set* (MCS for short) if  $\Delta$  is consistent and it holds that  $\varphi \in \Delta$  or  $\neg \varphi \in \Delta$  for every formula  $\varphi$  without the dynamic operators.

Let  $\Delta_1, \Delta_2 \in \text{MCS}$ . We write  $\Delta_1 =^{\text{Cons}} \Delta_2$  if they contain the same  $\mathcal{L}_B$ -formulas  $\alpha$  whose terms are constants. Also, let  $\mathcal{T} \subseteq \text{Term}$ , and we write  $\Delta_1 =_{\mathcal{T}} \Delta_2$  if

for any  $t \in \mathcal{T}$  and  $c \in \text{Cons}$ ,  $t \equiv c \in \Delta_1$  iff  $t \equiv c \in \Delta_2$ .

It is easy to see that when  $\mathcal{T} \subseteq \text{Cons}$ ,  $\Delta_1 =^{\text{Cons}} \Delta_2$  implies  $\Delta_1 =_{\mathcal{T}} \Delta_2$ .

**Fact 9.** Let  $\mathcal{T} \subseteq \text{Term}$  and  $\Delta_1, \Delta_2 \in \text{MCS}$  s.t.  $\Delta_1 =^{\text{Cons}} \Delta_2$  and  $\Delta_1 =_{\mathcal{T}} \Delta_2$ . Also, let  $\alpha \in \mathcal{L}_{\mathcal{B}}$  s.t. the terms occurring in it are among  $\mathcal{T}$ . Then,  $\alpha \in \Delta_1$  iff  $\alpha \in \Delta_2$ .

It can be proved by induction on  $\alpha \in \mathcal{L}_{\mathcal{B}}$ . Now we move to introducing the definition of the canonical models:

**Definition 6.** Let  $\Delta_0$  be a MCS containing the following:

$$K_x y = \mathcal{T}_y, K_y x = \mathcal{T}_x, x \equiv c_x, y \equiv c_y,$$

where  $\mathcal{T}_y \cup \mathcal{T}_x \cup \{c_x, c_y\} \subseteq \text{Cons}$ . We defined the canonical model  $M^{\Delta_0} = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  induced by  $\Delta_0$  in the following manner:

- $\mathbf{D} = \{[c] \mid c \in \text{Cons}\}$ , where  $[c] := \{c' \in \text{Cons} \mid c' \equiv c \in \Delta_0\}$
- $\mathbf{I}$  is defined as follows:  
 $\mathbf{I}(P) := \{([c_1], \dots, [c_n]) \mid P(c_1, \dots, c_n) \in \Delta_0\}$  and  $\mathbf{I}(c) := [c]$ .
- $\Sigma$  is given by the following:

$$\begin{aligned} & \{\Delta_0\} \cup \\ & \{\Delta_0 =^{\text{Cons}} \Delta \in \text{MCS} \mid K_x y = \mathcal{T}_y \in \Delta, K_y x \equiv c_x \in \Delta, y \equiv c, \neg c \equiv c_y \in \Delta \ \& \ c \in \mathcal{T}_y\} \cup \\ & \{\Delta_0 =^{\text{Cons}} \Delta \in \text{MCS} \mid K_y x = \mathcal{T}_x \in \Delta, K_x y \equiv c_y \in \Delta, x \equiv c, \neg c \equiv c_x \in \Delta \ \& \ c \in \mathcal{T}_x\} \end{aligned}$$

such that for any  $z \in \{x, y\}$ ,  $\Delta(z) = [c]$  if  $z \equiv c \in \Delta$ .

- For any  $\Delta_1, \Delta_2 \in \Sigma$ ,  $\Delta_1 \sim_z \Delta_2$  iff  $K_z \Delta_1 = K_z \Delta_2$ , where for any  $\Delta \in \Sigma$ ,  $K_z \Delta := \{\alpha \mid K_z \alpha \in \Delta\}$ .

The set  $\Sigma$  defined above is desired, in the sense of the following:

**Fact 10.** Let  $\Delta_0 \in \text{MCS}$  s.t.  $K_x y = \mathcal{T}_y$ ,  $K_y x = \mathcal{T}_x$ ,  $x \equiv c_x$ ,  $y \equiv c_y \in \Delta_0$ , where  $\mathcal{T}_y \cup \mathcal{T}_x \cup \{c_x, c_y\} \subseteq \text{Cons}$ . Let  $M^{\Delta_0} = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  be the induced canonical model.

- (1) Let  $\Gamma := \{\alpha \in \mathcal{L}_{\mathcal{B}} \mid \alpha \in \Delta_0 \text{ s.t. terms in } \alpha \text{ are constants}\}$ . For any  $c \in \mathcal{T}_y$  with  $\neg c \equiv c_y \in \Delta_0$ ,  $\{K_x y = \mathcal{T}_y, y \equiv c, \neg c \equiv c_y \ \& \ K_y x \equiv c_x\}$  is consistent with  $\Gamma$ . For any  $c \in \mathcal{T}_x$  with  $\neg c \equiv c_x \in \Delta_0$ ,  $\{K_y x = \mathcal{T}_x, x \equiv c, \neg c \equiv c_x \ \& \ K_x y \equiv c_y\}$  is consistent with  $\Gamma$ . So, when those constants  $c$  exist, there are MCSs of  $\Sigma$  containing the sets.
- (2) For any  $\Delta \in \Sigma$ , each of  $x, y$  has exactly one value.
- (3) For any  $\Delta_1, \Delta_2 \in \Sigma$ , if  $\Delta_1(x) = \Delta_2(x)$  and  $\Delta_1(y) = \Delta_2(y)$ , then  $\Delta_1 = \Delta_2$ .

*Proof.* The second item is easy to see. We merely prove the first and the third.

(1) For the first item, it suffices to consider the first part, and the second part is similar. Let  $c \in \mathcal{T}_y$ . Suppose for reductio that the set is not consistent with  $\Gamma$ . Then,

$$K_x y = \mathcal{T}_y \wedge K_y x \equiv c_x \vdash \neg c \equiv c_y \rightarrow (\bigwedge \Phi \rightarrow \neg y \equiv c), \quad \text{for some finite } \Phi \subseteq \Gamma$$

Applying the rule (K-ELIMINATION) can give us the following:

$$\vdash K_x y = \mathcal{T}_y \rightarrow (x \equiv c_x \rightarrow (\neg c \equiv c_y \rightarrow (\bigwedge \Phi \rightarrow \neg y \equiv c))).$$

By (K-ADDITIVITY),  $\vdash K_x y = \mathcal{T}_y \rightarrow K_x(x \equiv c_x \rightarrow (\neg c \equiv c_y \rightarrow (\bigwedge \Phi \rightarrow \neg y \equiv c)))$ . Then,  $\vdash K_x y = \mathcal{T}_y \rightarrow (K_x x \equiv c_x \rightarrow (K_x \neg c \equiv c_y \rightarrow (K_x \bigwedge \Phi \rightarrow K_x \neg y \equiv c)))$ . However, since  $K_x y = \mathcal{T}_y \wedge K_x x \equiv c_x \wedge K_x \neg c \equiv c_y \wedge K_x \bigwedge \Phi \in \Delta_0$ , we have  $K_x \neg y \equiv c$ , which contradicts  $c \in \mathcal{T}_y$ .

(2) We now move to the third item. By the construction, it is easy to see that one of  $\Delta_1$  and  $\Delta_2$  is  $\Delta_0$  iff  $\Delta_1 = \Delta_2 = \Delta_0$ . In what follows, we consider the case that  $\Delta_1 \neq \Delta_0$  and  $\Delta_2 \neq \Delta_0$ . Then, by construction, exactly one of  $\Delta_1(x) \neq [c_x]$  and  $\Delta_1(y) \neq [c_y]$  is the case, but they cannot hold at the same time. It suffices to consider the case that  $\Delta_1(x) = [c_x]$  and  $\Delta_1(y) \neq [c_y]$ , and the other case is analogous.

By assumption,  $\Delta_1 = \vee_{ar} \Delta_2$ , and by construction,  $K_x y = \mathcal{T}_y \in \Delta_1 \cap \Delta_2$ . By the former and Fact 9, the  $\mathcal{L}_B$ -parts of  $\Delta_1$  and  $\Delta_2$  are the same. Also, the arguments for the formulas  $K_x t$  and  $K_y t$  are simple. We now move to considering  $K_x \alpha$  and  $K_y \alpha$ .

Now suppose that  $K_x \alpha \in \Delta_1$ . Then,  $\bigwedge_{t \in \mathcal{T}_y} \alpha[c_x/x][t/y] \in \Delta_1$ . Again, by Fact 9,  $\bigwedge_{t \in \mathcal{T}_y} \alpha[c_x/x][t/y] \in \Delta_2$ , which together with  $K_x y = \mathcal{T}_y, x \equiv c_x \in \Delta_2$  can give us  $K_x \alpha \in \Delta_2$ . The converse is similar.

Next, suppose that  $K_y \alpha \in \Delta_1$ . Let  $c'_y \in \mathcal{T}_y$  with  $c'_y \equiv y \in \Delta_1$ . So, by assumption,  $c'_y \equiv y \in \Delta_2$ . It is easy to see that  $\alpha[c_x/x][c'_y/y] \in \Delta_1$ . By Fact 9,  $\alpha[c_x/x][c'_y/y] \in \Delta_2$ . With this, we know that  $K_y \alpha \in \Delta_2$ . Again, the converse is similar.  $\square$

**Theorem 1.** *For any  $\Delta_0 \in MCS$ , the induced  $M^{\Delta_0}$  is a  $k$ -sight model.*

*Proof.* (1) We first show that the components are well-defined. It is simple to check that the definition itself does not depend on the choice of representatives of equivalence classes  $[c]$ . Also, as indicated by Fact 10,  $\Sigma$  is a well-defined set of assignments.

(2) By the axiom (**Ser**iality), the relation  $\mathbf{I}(R)$  is serial.

(3) It is straightforward to see that  $\sim_x$  and  $\sim_y$  are equivalence relations.

(4) It remains to show that players have the  $k$ -sight ability. Assume that  $\Delta_1 \sim_x \Delta_2$ , and we show that for any  $z \in \{x, y\}$ , if  $\Delta_1(z) \in \mathbb{D}^k(\Delta_1(x))$ , then  $\Delta_1(z) = \Delta_2(z)$ .

By the definition of  $\mathbb{D}^k$ , we have  $\mathbb{D}^k x z$ . Then, by the axiom ( $k$ -sight), it holds that  $K_x z$ . Also, by the axiom (**At-Some-Where**), using the fact that  $\Delta_1$  is a  $MCS$ , there is some  $c \in Cons$  s.t.  $z \equiv c \in \Delta_1$ . Now, using the axiom (**De-Re-Knowledge**), we can obtain  $K_x z \equiv c \in \Delta_1$ . Using the axiom (**T**), we have  $z \equiv c \in \Delta_1$ . Also, since  $\Delta_1 \sim_x \Delta_2$ , it follows from  $K_x z \equiv c \in \Delta_1$  that  $z \equiv c \in \Delta_2$ . So,  $\Delta_1(z) = \Delta_2(z) = [c]$ .  $\square$

Now, we show the following *Existence Lemma*:

**Lemma 1.** *Let  $\Delta_0$  be a  $MCS$  such that  $K_x y = \mathcal{T}_y, K_y x = \mathcal{T}_x, x \equiv c_x, y \equiv c_y \in \Delta_0$ , where  $\mathcal{T}_y \cup \mathcal{T}_x \cup \{c_x, c_y\} \subseteq Cons$ . Also, let  $M^{\Delta_0}$  be the induced canonical model of  $\Delta_0$ . For any  $z \in \{x, y\}$  and any  $\Delta \in MCS$  from  $M^{\Delta_0}$ , when  $K_z \alpha \notin \Delta$ , there is some  $\Delta'$  from  $M^{\Delta_0}$  s.t.  $\Delta \sim_z \Delta'$  and  $\alpha \notin \Delta$ .*

*Proof.* W.l.o.g., let  $z := x$ . When  $\Delta(x) \neq [c_x]$ , by construction,  $K_x y \equiv c_y \in \Delta$ . Now,  $K_x \alpha \notin \Delta$  gives us  $\neg \alpha \in \Delta$ , so a desired  $\Delta'$  is  $\Delta$  itself. Let us assume that  $\Delta(x) = [c_x]$ . Now, it follows from  $K_x \alpha \notin \Delta$  that there is  $c_y^0 \in \mathcal{T}_y$  such that  $\neg \alpha[c_x/x][c_y^0/y] \in \Delta$ .

If  $c_y^0 \equiv c_y \in \Delta$ , then  $\Delta = \Delta_0$  (cf. the proof for the item (3) of Fact 10). Now,  $\Delta_0$  itself is such a  $\Delta'$ . Let us now assume that  $\neg c_y^0 \equiv c_y \in \Delta$ . We consider  $\Delta_1 \in MCS$  such that  $\Delta' = {}^{Cons} \Delta_0$  and  $\{K_x y = \mathcal{T}_y, K_y x \equiv c_x, y \equiv c_y^0, \neg c_y^0 \equiv c_y\} \subseteq \Delta'$ . We have seen from the item (1) of Fact 10 that  $\Sigma$  does contain such a  $\Delta_1$ . It is simple to see that  $\neg \alpha \in \Delta_1$ . Now it remains to show that  $\Delta \sim_x \Delta_1$ , which is guaranteed by the fact  $\Delta_1 = {}^{Cons} \Delta_0, K_x y = \mathcal{T}_y \in \Delta \cap \Delta_1$  and the axiom (**Knowledge-Ground**).  $\square$

As a consequence, we have the following:

**Lemma 2.** *Let  $\Delta_0 \in MCS$  and  $M^{\Delta_0}$  be its induced canonical model. For  $\{z, z'\} = \{x, y\}$  and any  $\Delta \in MCS$  from  $M^{\Delta_0}$ , when  $\neg K_z z' \in \Delta$ , there is a  $\Delta'$  from  $M^{\Delta_0}$  such that  $\Delta' \sim_z \Delta$  and for some  $c \in Cons$ ,  $c \equiv z' \in \Delta$  and  $c \not\equiv z' \in \Delta'$ .*

*Proof.* Since  $K_z z' \leftrightarrow \bigwedge_{c \in Cons} ((K_z)z' \equiv c \rightarrow K_z z' \equiv c)$ , it holds by Lemma 1.  $\square$

Next we proceed to show the crucial *Truth Lemma*:

**Lemma 3.** *Let  $\Delta_0 \in MCS$  and  $M^{\Delta_0}$  be its induced canonical model. For any  $\varphi \in \mathcal{L}^-$  and  $\Delta$  from  $M^{\Delta_0}$ ,  $M^{\Delta_0}, \Delta \models \varphi$  iff  $\varphi \in \Delta$ .*

*Proof.* It goes by induction on formulas. The cases for Boolean connectives  $\neg, \wedge$  are straightforward by induction hypothesis, and we consider others.

(1) Formula  $\varphi$  is  $P(t_1, \dots, t_n)$ . Then, we have the following equivalences:

$$\begin{aligned} M^{\Delta_0}, \Delta \models \varphi & \text{ iff } (t_1^{(\mathbf{I}, \Delta)}, \dots, t_n^{(\mathbf{I}, \Delta)}) \in \mathbf{I}(P) \\ & \text{ iff } ([c_1], \dots, [c_n]) \in \mathbf{I}(P) \\ & \text{ iff } P(c_1, \dots, c_n) \in \Delta_0 \\ & \text{ iff } P(c_1, \dots, c_n) \in \Delta \\ & \text{ iff } P(t_1, \dots, t_n) \in \Delta \end{aligned}$$

In the second equivalence, those  $c_i$  are constants, and when  $t_i$  is a constant,  $c_i$  can be  $t_i$  itself, and when  $t_i$  is a variable, we put  $c_i := c$  for some  $c \in Cons$  s.t.  $c \equiv t_i \in \Delta$  (due to **(At-Some-Where)**, such a constant  $c$  always exists). By Fact 9 and  $\Delta =^{Cons} \Delta_0$  (the latter follows from  $\Delta \in \Sigma$ ), the fourth equivalence holds. The last one holds directly (if all those  $t_i$  are constants) or holds by **(A4)** (if some of those  $t_i$  are variables).

(2) Formula  $\varphi$  is  $t_1 \equiv t_2$ . The reasoning is similar to the case above.

(3) Formula  $\varphi$  is  $K_z t$ . When  $t$  is a constant or  $z$ , we have  $K_z t \in \Delta$  (Fact 8), and meanwhile, for the semantic aspect, it follows from the construction of canonical models that  $M^{\Delta_0}, \Delta \models K_z t$ . We now move to the case that  $t$  is the other variable  $z'$ .

By Lemma 2, when  $M^{\Delta_0}, \Delta \models K_z t$ ,  $K_z t \in \Delta$ . For the other direction, assume that  $K_z t \in \Delta$ ,  $\Delta \sim_z \Delta'$ ,  $\Delta(z') = [c_1]$  and  $\Delta'(z') = [c_2]$ . We will show  $[c_1] = [c_2]$ , for which it suffices to prove that  $c_1 \equiv c_2 \in \Delta'$ . Now,  $c_1 \equiv z' \in \Delta$  and  $c_2 \equiv z' \in \Delta'$ . Given  $c_1 \equiv z' \in \Delta$  and  $K_z t \in \Delta$ , using **(De-Re-Knowledge)** we obtain  $K_z z' \equiv c_1 \in \Delta$ . Since  $\Delta \sim_z \Delta'$ , it holds that  $z' \equiv c_1 \in \Delta'$ . So,  $c_1 \equiv c_2 \in \Delta'$ .

(4) Finally, we consider the case that  $\varphi$  is  $K_z \alpha$ . We proceed as follows:

$$\begin{aligned} M^{\Delta_0}, \Delta \models K_z \alpha & \text{ iff for all } \Delta' \in \Sigma, \text{ if } \Delta \sim_z \Delta', \text{ then } M^{\Delta_0}, \Delta' \models \alpha \\ & \text{ iff for all } \Delta' \in \Sigma, \text{ if } \Delta \sim_z \Delta', \text{ then } \alpha \in \Delta' \\ & \text{ iff } K_z \alpha \in \Delta \end{aligned}$$

The second equivalence holds by induction hypothesis. One direction of the last equivalence holds by Lemma 1, and the converse holds by  $K_z \alpha \in \Delta$  and  $\Delta \sim_z \Delta'$ .  $\square$

Finally we can show the following strong completeness result for **ELCR<sup>-</sup>**:

**Theorem 2.** *For any set  $\Gamma \subseteq \mathcal{L}^-$ , if  $\Gamma$  is consistent, then it is satisfiable. As a consequence, the static part is compact.*

*Proof.* Given a consistent set  $\Gamma \subseteq \mathcal{L}^-$ , by a Lindenbaum-style argument [16], we can extend it to some  $\Delta_0 \in MCS$ . Then, there is a canonical model  $M^{\Delta_0}$  induced by  $\Delta_0$ . By Lemma 3, for any  $\Delta$  from  $M^{\Delta_0}$  and any  $\varphi \in \mathcal{L}^-$ ,  $M^{\Delta_0}, \Delta \models \varphi$  iff  $\varphi \in \Delta$ . Notice that  $\Delta_0$  is also an assignment in  $M^{\Delta_0}$ . Since  $\Gamma \subseteq \Delta_0$ ,  $M^{\Delta_0}, \Delta_0 \models \Gamma$ , as desired.  $\square$

Moreover, we can show the following:

**Corollary 1.**  $ELCR^-$  is decidable.

*Proof.* Notice that in  $M^{\Delta_0} = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$ , both  $\mathbf{D}$  and  $\Sigma$  are finite, so a static formula that is satisfiable can be satisfied by a finite model. Moreover, given that the fragment is finitely axiomatizable, we can obtain its decidability immediately.  $\square$

## 7 Axiomatization of ELCR

So far, we have shown a complete calculus for the static base  $ELCR^-$ , and this section is devoted to providing a complete calculus for ELCR. To do so, it is enough to find an effective way to eliminate the occurrences of dynamic operators, like the techniques of *recursion axioms* developed for DEL.

First, we consider the following formulas that can reduce  $\mathcal{L}_{BD}$  to  $\mathcal{L}_B$ :

$$[z]\alpha \leftrightarrow \bigwedge_{\mathcal{T} \subseteq Cons} (Rz = \mathcal{T} \rightarrow \bigwedge_{c \in \mathcal{T}} \alpha[c/z]), \quad \text{given } \alpha \in \mathcal{L}_B \quad (\mathbf{R1})$$

$$[z]\neg\varphi \leftrightarrow \neg[z]\varphi \quad (\mathbf{R2})$$

$$[z](\varphi \wedge \psi) \leftrightarrow [z]\varphi \wedge [z]\psi \quad (\mathbf{R3})$$

Notice that there is always a set  $\mathcal{T} \subseteq Cons$  making  $Rz = \mathcal{T}$  true, and when  $\alpha$  does not contain any occurrence of  $z$ , (R1) amounts to  $[z]\alpha \leftrightarrow \alpha$ .<sup>12</sup> Moreover, different from the ordinary format of recursion axioms for DEL, the  $\alpha$  in (R1) need not be an atom.

Although the syntax of  $\mathcal{L}_{BD}$  allows nested occurrences of dynamic operators, we can always start with the elimination of some innermost occurrence of a dynamic operator, as the case for DEL.

Next, we move to dealing with  $[z]K_{z'}t$ :

$$[z]K_{z'}t \leftrightarrow \top, \quad \text{where } t \in Cons \cup \{z'\} \quad (\mathbf{R4})$$

$$[z]K_{z'}z' \leftrightarrow K_{z'}z' \vee$$

$$\bigwedge_{\mathcal{T}, \mathcal{T}_1 \subseteq Cons} (K_{z'}z' = \mathcal{T} \wedge Rz = \mathcal{T}_1 \rightarrow \quad (\mathbf{R5})$$

$$\bigwedge_{a \in \mathcal{T}_1} (D^k a z' \vee \bigwedge_{b \in \mathcal{T}} (b \neq z' \rightarrow D^k ab))), \quad \text{where } \{z, z'\} = \{x, y\}$$

$$[z]K_{z'}z \leftrightarrow \bigwedge_{\mathcal{T}, \mathcal{T}_1 \subseteq Cons} (Rz = \mathcal{T} \wedge K_{z'}z = \mathcal{T}_1 \rightarrow ((\bigwedge_{t \in \mathcal{T}} D^k z't) \vee$$

---

<sup>12</sup>Recall that we have (Seriality) as an axiom of the static part, so it cannot be the case that  $[z]\perp$ .

$$\bigwedge_{t_1, t_2 \in \mathcal{T}_1, t'_1, t'_2 \in \text{Cons}} (Rt_1 t'_1 \wedge Rt_2 t'_2 \wedge \neg D^k z' t'_1 \wedge \neg D^k z' t'_2 \rightarrow t'_1 \equiv t'_2)), \quad \text{given that } \{z, z'\} = \{x, y\} \quad (\text{R6})$$

The formula (R4) suggests that for any  $z, z' \in \{x, y\}$ ,  $[z]K_{z'}\text{Cons} \cup \{z'\}$  is always the case. By (R5), for different  $z, z'$ ,  $[z]K_z z'$  holds if, and only if, one of the following holds:

- (i)  $z$  has already known the position of  $z'$  before the movement (i.e.,  $K_z z'$ ).
- (ii) Given the possible positions  $\mathcal{T}$  of  $z'$  considered by  $z$  and the set  $\mathcal{T}_1$  of  $\mathbf{R}$ -successors of  $z$ , for any possible new position  $a$  of  $z$ , *either*  $z$  can see directly where  $z'$  is or all other vertices different from the position of  $z'$  can be observed directly by  $z$ .

The last (R6) means that for different  $z$  and  $z'$ , when we assume that  $Rz = \mathcal{T}$  and  $K_{z'} z = \mathcal{T}_1$ ,  $[z]K_{z'} z$  is the case if, and only if, one of the following is the case:

- (i) All those of  $\mathcal{T}$  are in the sight of  $z'$ .
- (ii) When  $t_1, t_2 \in \mathcal{T}_1$  have successors that are not in the sight of  $z'$ , all those successors not in the sight of  $z$  are the same.

Now we proceed to tackle  $[z]K_{z'}\alpha$ . In what follows, for any  $\alpha \in \mathcal{L}_{\mathbf{B}}$  and  $z \in \{x, y\}$ , we use  $\alpha(z)$  to highlight that  $z$  does occur in  $\alpha$ . The details are as follows:

$$[z]K_z \alpha \leftrightarrow [z]\alpha, \quad \text{given that } \alpha \in \mathcal{L}_{\mathbf{B}} \text{ does not contain the other variable} \quad (\text{R7})$$

$$[z]K_{z'} \alpha \leftrightarrow \alpha, \quad \text{given } \{z, z'\} = \{x, y\} \text{ and } \alpha \in \mathcal{L}_{\mathbf{B}} \text{ does not contain } z \quad (\text{R8})$$

$$\begin{aligned} [z]K_z \alpha(z') \leftrightarrow & \bigwedge_{\mathcal{T}, \mathcal{T}_1 \subseteq \text{Cons}} (Rz = \mathcal{T} \wedge K_z z' = \mathcal{T}_1 \rightarrow ((K_z z' \wedge \bigwedge_{c \in \mathcal{T}} \alpha[c/z]) \vee \\ & (\neg K_z z' \wedge \bigwedge_{t \in \mathcal{T}} ((D^k t z' \wedge \alpha[t/z]) \vee \\ & (\neg D^k t z' \wedge \bigwedge_{c \in \mathcal{T}_1} (\neg D^k t c \rightarrow \alpha[t/z][c/z'])))))), \\ & \text{given } \{z, z'\} = \{x, y\} \text{ and } \alpha \in \mathcal{L}_{\mathbf{B}} \end{aligned} \quad (\text{R9})$$

$$\begin{aligned} [z]K_{z'} \alpha(z) \leftrightarrow & \bigwedge_{\mathcal{T}, \mathcal{T}_1 \subseteq \text{Cons}} (Rz = \mathcal{T} \wedge K_{z'} z = \mathcal{T}_1 \rightarrow (\bigwedge_{t \in \mathcal{T}} \alpha[t/z]) \wedge ((\bigwedge_{t \in \mathcal{T}} D^k z' t) \vee \\ & \bigwedge_{t_1 \in \mathcal{T}_1, t_2 \in \text{Cons}} (Rt_1 t_2 \wedge \neg D^k z' t_2 \rightarrow \alpha[t_2/z]))), \\ & \text{given } \{z, z'\} = \{x, y\} \text{ and } \alpha \in \mathcal{L}_{\mathbf{B}} \end{aligned} \quad (\text{R10})$$

Among these, the formulas (R7) and (R8) are simple, and principles (R9) and (R10) are their complements, respectively, for the more complicated settings: one can check that the latter two amount to the former ones when  $\alpha$  does not contain the corresponding variables. For the formula (R9), we assume that  $Rz = \mathcal{T}$  and  $K_z z' = \mathcal{T}_1$ , and the formula expresses that  $[z]K_z \alpha(z')$  iff one of the following is the case:

- (i)  $z$  already knew the position of  $z'$  before the movement, and after any movement of  $z$ ,  $\alpha$  is the case.

(ii)  $z$  did not know the position of  $z'$  before the movement, but for any  $t$  where  $z$  can move to, either  $z'$  can be observed from  $t$  and  $\alpha$  is the case, or  $z'$  cannot be observed from  $t$  but any unobservable possibility in  $\mathcal{T}$  from  $t$  makes  $\alpha$  true.

Finally, for the principle (R10), we assume that  $Rz = \mathcal{T}$  and  $K_{z'}z = \mathcal{T}_1$ , and it states that  $[z]K_{z'}\alpha(z)$  is true iff  $\bigwedge_{t \in \mathcal{T}} \alpha[t/z]$  and one of the following holds:

- (i)  $z$  can only move into the observable range of  $z'$ .
- (ii) For any possibility  $t_1 \in \mathcal{T}_1$  (considered by  $z'$  before the movement), any of its unobservable  $\mathbf{R}$ -successor from  $z'$  makes  $\alpha$  true.

We write **ELCR** for the resulting calculus obtained by adding (R1)-(R10) to the proof system **ELCR**<sup>-</sup>, and generalize the usages of  $\vdash$  to the setting of **ELCR**. We show that **ELCR** is both sound and complete.

**Theorem 3.** *The calculus **ELCR** is sound for ELCR.*

*Proof.* See Appendix B. □

**Theorem 4.** *For any set  $\Gamma \subseteq \mathcal{L}$ , if  $\Gamma$  is consistent, then it is satisfiable. As a consequence, ELCR is also compact.*

*Proof.* For any  $\varphi \in \Gamma$ , by (R1)-(R10), there is always a static  $\psi \in \mathcal{L}^-$  such that  $\vdash \varphi \leftrightarrow \psi$ . Based on this, there is a set  $\Gamma' \subseteq \mathcal{L}^-$  that is provably equivalent to  $\Gamma$ .  $\Gamma$  is consistent, so is  $\Gamma'$ . Then, by Theorem 2,  $\Gamma'$  is satisfiable. By the soundness (Theorem 3),  $\Gamma$  is satisfiable, as desired. □

Finally, the arguments in the proof above suggest the following:

**Corollary 2.** *ELCR is decidable.*

*Proof.* By the proof for the completeness of **ELCR**, any formula of  $\mathcal{L}$  is equivalent to a static formula of  $\mathcal{L}^-$ . Now it follows from Corollary 1 that ELCR is also decidable. □

So, different from the undecidable proposal in [36] for the perfect information game of Hide and Seek, we now have a decidable framework for the imperfect information version of the game.<sup>13</sup>

## 8 Related approaches

There are many other logical milestones in the realm of knowledge dynamics, especially the paradigm of DEL. A DEL approach to Cops and Robbers is discussed in [12], and we will formalize its key points and apply the ideas to the same Example 1. This would form a visual comparison between the two frameworks, indicating the succinctness of our proposal. In addition, this section also offers an overview on the recent logical frameworks developed for games played on graphs and examines logics addressing various dependencies that are technically relevant to our work.

---

<sup>13</sup>As stated in [36], the culprit of the undecidability of that logic is the equality. Our language also contains the symbol, and one reason for the decidability of ELCR is that we confine ourselves to the finite setting. But it is important to notice that generalizing ELCR into infinite case does not necessarily lead us to an undecidable framework. For more discussion on when the equality is dangerous, see [44].

## 8.1 DEL-approach to the game

While exploring ELCR, one may wonder whether some appropriate modifications to the techniques developed for DEL, e.g., *product update* [7], may apply to the game under consideration. As stated above, some such ideas are sketched in [12], and we now formalize them with some necessary modifications. In this part, to avoid digressing too far, we will omit many basics of DEL and discuss the key points in a concise manner.

Our main focus is on the product updates of *epistemic models* and *event models*. The former are exactly our *k-sight models*, and for the latter, the game graphs themselves serve as an important parameter, but with a different reading now: edges are now moves from a vertex to another.

**Definition 7.** Let  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  be a *k-sight model* and  $\sigma \in \Sigma$ . There are event model  $E_\sigma^x$  for  $x$  associated to  $\sigma$  and event model  $E_\sigma^y$  for  $y$  associated to  $\sigma$ , which are analogous. We just define the former, and skip that of the latter. Details are as follows:

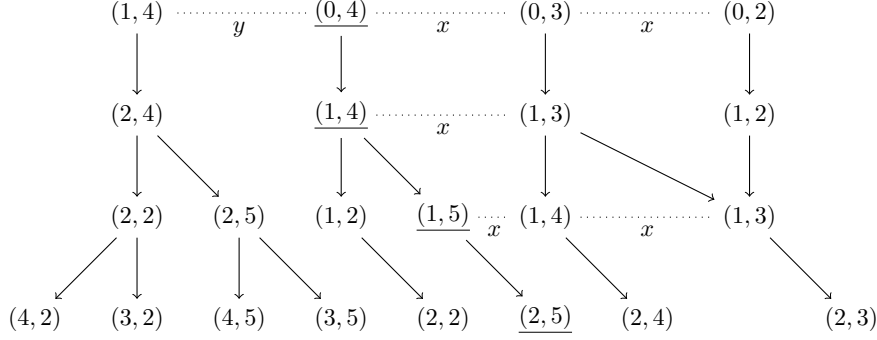
$E_\sigma^x = (\mathbf{I}(R), \approx, \{pre_{(s,t)} \mid (s,t) \in \mathbf{I}(R)\}, \{post_{(s,t)} \mid (s,t) \in \mathbf{I}(R)\})$ , where

- For each  $(s,t) \in \mathbf{I}(R)$ , its pre-condition  $pre_{(s,t)}$  is that  $x$  is at  $s$ , and its post-condition  $post_{(s,t)}$  is that  $x$  is at  $t$ , which together mean the change of the position of  $x$  from  $s$  to  $t$  caused by the action.
- The indistinguishability relation  $\approx_x$  for  $x$  is the identity relation on the pairs  $\mathbf{I}(R)$ , since  $x$  always knows which movement she is taking.
- The case for the indistinguishability relation  $\approx_y$  for  $y$  is more complicated, since we need to take the *k-sight ability* into account, which determines whether or not  $y$  knows the action of  $x$ :
  - For those  $(s,t) \in \mathbf{I}(R)$  with  $s, t \in \mathbb{D}^k(\sigma(y))$ ,  $\approx_y$  is the identity relation.
  - For those  $(s,t) \in \mathbf{I}(R)$  with  $s \in \mathbb{D}^k(\sigma(y))$  and  $t \notin \mathbb{D}^k(\sigma(y))$ ,  
 $(s,t) \approx_y (s',t')$  iff  $s = s'$  and  $t' \notin \mathbb{D}^k(\sigma(y))$ .
  - For those  $(s,t)$  with  $s \notin \mathbb{D}^k(\sigma(y))$  and  $t \in \mathbb{D}^k(\sigma(y))$ ,  
 $(s,t) \approx_y (s',t')$  iff  $t = t'$  and  $s' \notin \mathbb{D}^k(\sigma(y))$ .
  - For those  $(s,t)$  with  $s \notin \mathbb{D}^k(\sigma(y))$  and  $t \notin \mathbb{D}^k(\sigma(y))$ ,  
 $(s,t) \approx_y (s',t')$  iff  $s, t, s', t' \notin \mathbb{D}^k(\sigma(y))$ .

**Definition 8.** Let  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  be a *k-sight model*,  $\sigma \in \Sigma$  and  $E_\sigma^x$  be the event model associated to  $\sigma$ . The product update  $(M, \sigma) \times E_\sigma^x$  is a new *k-sight model*  $(\mathbf{D}, \mathbf{I}, \Sigma', \sim')$  defined as follows:

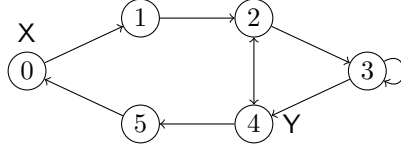
- $\Sigma' = \{(\sigma_1, x, (s,t)) \mid \sigma_1(x) = s, \sigma_1 \in \Sigma\}$ , where  $(\sigma_1, x, (s,t))$  is a new situation such that  $(\sigma_1, x, (s,t))(x) = t$  and  $(\sigma_1, x, (s,t))(y) = \sigma_1(y)$ .
- For any variable  $z \in \text{Var}$  and situations  $(\sigma_1, x, (s_1, t_1)), (\sigma_2, x, (s_2, t_2)) \in \Sigma'$ ,  $(\sigma_1, x, (s_1, t_1)) \sim'_z (\sigma_2, x, (s_2, t_2))$  if, and only if,
  - (i)  $\sigma_1 \sim_z \sigma_2, (s_1, t_1) \approx_z (s_2, t_2)$ .
  - (ii) If  $\sigma_1(x) \in \mathbb{D}^k(\sigma_1(y))$ , then  $\sigma_1 = \sigma_2$ , otherwise  $\sigma_2(x) \notin \mathbb{D}^k(\sigma_2(y))$ .

In the definition above, the clause (i) is the usual way to obtain the indistinguishability relations in product updates, and the clause (ii) is an additional requirement



**Fig. 1** Example analyses: A DEL-approach. We use the dotted links labeled with  $x$  and  $y$  to represent the indistinguishability relations of the two players (we omit the self-loops and transitive links). Also, there are four layers connected with solid arrows. Each of the layers is an epistemic model, and for simplicity, we just draw the corresponding classes of situations and the indistinguishability relations, and highlight the actual situations with underlines: the first layer is the original epistemic model  $M$ , the second one is  $M_1 = (M, (0, 4)) \times E_{(0,1)}^x$ , the third is  $M_2 = (M_1, (1, 4)) \times E_{(4,5)}^y$ , and the last layer is  $(M_2, (1, 5)) \times E_{(1,2)}^x$ . Finally, the solid arrows indicate how the classes of situations in the latter stages come from the previous stages.

that ensures that the resulting models are still  $k$ -sight models.<sup>14</sup> Now we can formally re-analyze Example 1 with the DEL-approach specified above. Details are given in Figure 1. We add the game structure below to remind the readers.



As mentioned in Figure 1, each of the four layers is an epistemic model. Notice that there is a striking analogy between the ELCR-based analyses and the discussion here: in each layer with the class of situations  $\Sigma$  and the actual situation  $(s_1, s_2)$ , the *part*  $\Sigma|(s_1, s_2)$  corresponds to *complete* situations in the ELCR setting. Although the resulting sets of situations based on the ELCR-updates may increase (or decrease) in size, every such situation makes sense for the analyses of the game. We do not need to pay attention to the irrelevant situations. Following this viewpoint we claim that ELCR is a more succinct proposal.

## 8.2 Other perspectives

As stated earlier, this work is an extension of [34]. The game and its various variants have been well-studied in the field of computer science, with algorithmic and combinatorial perspectives (see e.g., [19, 40]). Complementary to these approaches, subsequent

<sup>14</sup>As the case in [12], one can also use two types of event models: one for movements, which can be obtained by removing the clause (ii); and the other for “inspection”, which can make the players know whether or not they are in the sight of each other. Then the desired outcome can be obtained by the product of the original epistemic model, the event model for movements and the event model for inspection (in this order). But here we mix the two types of event models for convenience.

studies from the logical aspects have been conducted in recent years for the perfect information version of the game, also termed as Hide and Seek in the literature. The first logical proposal in this direction is defined in [14], which constitutes a broad program that promotes the study of graph game design in tandem with matching new modal logics. Afterwards, this is studied extensively in [35, 36, 20], involving its expressive power at the levels of models and frames, computational behavior, axiomatization and its connections with other relevant paradigms like *product logics with the diagonal constant* (e.g., [31, 37]). More recently, the original logic for the game has been extended with formulas from *hybrid logic* (e.g., [17]) in [44], which are important to establish a complete Hilbert-style proof system for the resulting logic. In addition, based on the approach of substitutions [11], [46] develops a logical proposal that is crucial to capture the winning positions of players in a natural infinite setting.

This work and the series of logical studies on Hide and Seek belong to the broader exploration on the interaction between graph game and logic, a trend starting from *sabotage games* [9, 3, 2] and its matching *sabotage modal logic*. Different from Cops and Robbers in which the game graphs are fixed, in each round of a sabotage game, a blocker tries to stop the other player from moving to a given goal region by removing a link from the graph. Many variants of sabotage games have been studied from a logical perspective [43, 28, 6].<sup>15</sup> Also, [32] studies the graph games with a particular policy of link deletion that is performed under certain conditions that can be expressed explicitly in a given language. Distinct from link modifications, [18, 29] develop logics to capture the so-called *poison games* [24], in which a player can poison a node, to make it unavailable to the opponent. The logics of poison games are further explored in [23], involving their axiomatization and computational behavior. In addition, [45] studies a dynamic logic of local fact changes that captures a class of graph games in which properties of vertices might be affected by other vertices. All these studies on graph games are about the settings in which players have perfect information. Based on a DEL-approach, [12] analyzes some imperfect information versions, involving the sabotage games and the game of Hide and Seek. For more on this topic, we refer to [14] for a broad research program, [33] for extensive references to modal logics for graph games, and [15] for the latest developments of this area.

Finally, the design of the static fragment of our language is inspired by the works on different sorts of dependencies: *the logic of epistemic dependency* [4] and *the logic of functional dependency* [8]. The latter is about the dependencies between variables, and contains formulas of the form  $D_x y$ , expressing that fixing the value of variable  $x$  would determine the value of variable  $y$ . More relevantly, the logical language of the former contains formulas  $K_a x$  expressing that the agent  $a$  knows the value of  $x$ , where  $a$  is an index for agents and  $x$  is a variable of the object language. In contrast, our work introduces formulas of the form  $K_x y$ , meaning  $x$  knows the location of  $y$ , with both  $x$  and  $y$  as variables. Moreover, [4] explores the dynamic scenarios induced by public announcement operators that are central to DEL, our focus is on agent movements, essential to the game of Cops and Robbers. Combining these two approaches

---

<sup>15</sup>There are many further works on the sabotage modal logic. In [13], the logic is axiomatized in a broader setting with hybrid formulas, and [27] offers an upper bound of the complexity to determine the notion of bisimulation for the logic. Also, in addition to the sabotage modal logic that contains an operator to delete links, there is a class of *relation-changing logics* that contain operators to swap and add links [26, 1, 22].

to dynamics – public announcements and movement-based updates – would certainly be an interesting direction for further study.

## 9 Conclusion and future work

*Summary.* To study the game of Cops and Robbers with uncertainty among players, this paper provided a formal framework ELCR to capture players’ reasoning about knowledge and actions. As illustrated, many validities of the logic characterized natural assumptions on the game. Axiomatization and decidability of the static version  $\text{ELCR}^-$  and the full dynamic logic ELCR were explored. From a broader perspective, we showed that the ideas underlying the design of ELCR can be easily adjusted to fit other important variants of the game and provided a formal connection between ELCR and the corresponding setting with simultaneous moves of players. In addition, we formalized the ideas in the literature that advocate using the DEL-approach for studying the game. In process, we set the stage for proving a succinctness result with respect to our update mechanism.

*Future work.* Several further directions have been identified in the article. In addition, there are other promising directions to explore. For example, there is extensive literature on the complexity of different versions of the Cops and Robbers game [19], and we intend to do the same for the one introduced in our work and its different variants, and in process, design efficient algorithms to construct winning strategies of players. On the logic side, an immediate extension would be to provide a complete Hilbert-style proof system for the logic designed for the simultaneous movements. This could be obtained by adapting the recursion axioms provided in Section 7. Also, although Section 8.1 illustrates the succinctness of ELCR, it remains to be examined how much more succinct it is than the DEL-approach. Another important direction is to study other logical properties of ELCR, including its expressiveness and frame correspondence. Moreover, it is crucial to study the setting involving higher-order knowledge, especially the cases that players have limited abilities to reason about each other’s knowledge [39]. Finally, from the games perspective, graph games with imperfect information warrant a more detailed study, which could be facilitated by usage of logic tools developed in this work.

**Acknowledgements.** This research was inspired by a question from Alexandru Baltag when the logic of the hide and seek game was presented at the January 2022 workshop ‘Exploring Baltag’s Universe’. We thank Alexandru Baltag, Johan van Benthem, Davide Grossi, and Katsuhiko Sano for their valuable feedback, as well as the editors of the LORI special issue and the anonymous referees for their helpful comments. Dazhu Li is supported by the National Social Science Foundation of China [22CZX063]. Sujata Ghosh acknowledges financial support from the Department of Science and Technology, Government of India (Ref. No. DST/CSRI/2018/202, CSRI). Fenrong Liu is supported by the Tsinghua University Initiative Scientific Research Program.

## References

- [1] C. Areces, R. Fervari, G. Hoffmann, and M. Martel. Satisfiability for relation-changing logics. *Journal of Logic and Computation*, 28:1443–1470, 2018.
- [2] G. Aucher, J. van Benthem, and D. Grossi. Sabotage modal logic: Some model and proof theoretic aspects. In W. van der Hoek, W. Holliday, and W. Wang, editors, *Proceedings of LORI 2015*, volume 9394 of *LNCS*, pages 1–13, Heidelberg, 2015. Springer.
- [3] G. Aucher, J. van Benthem, and D. Grossi. Modal logics of sabotage revisited. *Journal of Logic and Computation*, 28:269–303, 2018.
- [4] A. Baltag. To know is to know the value of a variable. In L. Beklemishev, S. Demri, and A. Máté, editors, *Advances in Modal Logic 11*, page 135–155, London, 2016. College Publications.
- [5] A. Baltag, Z. Christoff, R. K. Rendsvig, and S. Smets. Dynamic epistemic logics of diffusion and prediction in social networks. *Studia Logica*, 107:489–531, 2019.
- [6] A. Baltag, D. Li, and M. Y. Pedersen. A modal logic for supervised learning. *Journal of Logic, Language and Information*, 31:213–234, 2022.
- [7] A. Baltag, L. S. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. In I. Gilboa, editor, *Proceedings of TARK 1998*, pages 43–56, 1998.
- [8] A. Baltag and J. van Benthem. A simple logic of functional dependence. *Journal of Philosophical Logic*, 50:939–1005, 2021.
- [9] J. van Benthem. An essay on sabotage and obstruction. In D. Hutter and W. Stephan, editors, *Mechanizing Mathematical Reasoning*, volume 2605 of *LNCS*, pages 268–276. Springer, Heidelberg, 2005.
- [10] J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, Cambridge, 2011.
- [11] J. van Benthem. An abstract look at the fixed-point theorem for provability logic. In N. Bezhanishvili, R. Iemhoff, and F. Yang, editors, *Dick de Jongh on Intuitionistic and Provability Logics*, volume 28 of *Outstanding Contributions to Logic*, pages 75–88. Springer, Cham, 2024.
- [12] J. van Benthem. Dynamic-epistemic logic for games. In J. van Benthem and F. Liu, editors, *Graph Games and Logic Design: Recent Developments and Future Directions* (To Appear). Springer, Cham, 2025.
- [13] J. van Benthem, L. Li, C. Shi, and H. Yin. Hybrid sabotage modal logic. *Journal of Logic and Computation*, 33:1216–1242, 2023.
- [14] J. van Benthem and F. Liu. Graph games and logic design. In F. Liu, H. Ono, and J. Yu, editors, *Knowledge, Proof and Dynamics*, Logic in Asia: Studia Logica Library, pages 125–146, Singapore, 2020. Springer.
- [15] J. van Benthem and F. Liu, editors. *Graph Games and Logic Design: Recent Developments and Future Directions* (To Appear). Springer, Cham, 2025.
- [16] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, UK, 2001.
- [17] P. Blackburn and J. Seligman. Hybrid languages. *Journal of Logic, Language and Information*, 4:251–272, 1995.
- [18] F. Z. Blando, K. Mierzewski, and C. Areces. The modal logics of the poison

- game. In F. Liu, H. Ono, and J. Yu, editors, *Knowledge, Proof and Dynamics*, Logic in Asia: Studia Logica Library, pages 3–23, Singapore, 2020. Springer.
- [19] A. Bonato and R. J. Nowakowski. *The Game of Cops and Robbers on Graphs*, volume 61 of *The Student Mathematical Library*. AMS, Providence, 2011.
- [20] Q. Chen and D. Li. Logic of the hide and seek game: Characterization, axiomatization, decidability. In N. Gierasimczuk and F. R. Velázquez-Quesada, editors, *Proceedings of DaLí 2023*, volume 14401 of *LNCS*, pages 20–34, Cham, 2024. Springer.
- [21] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer, Dordrecht, 2008.
- [22] P. Du and Q. Chen. Axiomatization of hybrid logic of link variations. In N. Gierasimczuk and F. R. Velázquez-Quesada, editors, *Proceedings of DaLí 2023*, volume 14401 of *LNCS*, pages 35–51, 2024.
- [23] P. Du, F. Liu, and D. Li. Modal logics for the poison game: Axiomatization and undecidability. Manuscript, 2025.
- [24] P. Duchet and H. Meyniel. Kernels in directed graphs: A poison game. *Discrete Mathematics*, 115:273–276, 1993.
- [25] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. The MIT Press, Cambridge, 1995.
- [26] R. Fervari. *Relation-Changing Modal Logics*. PhD thesis, Universidad Nacional de Córdoba, Argentina, 2014.
- [27] S. Ghosh, S. Gupta, and L. Li. Bisimulation in model-changing modal logics: An algorithmic study. *Journal of Logic and Computation*, 34:399–427, 2024.
- [28] N. Gierasimczuk, L. Kurzen, and F. Velázquez-Quesada. Learning and teaching as a game: A sabotage approach. In X. He, J. Horty, and E. Pacuit, editors, *Proceedings of LORI 2009*, volume 5834 of *LNCS*, pages 119–132, Heidelberg, 2009. Springer.
- [29] D. Grossi and S. Rey. Credulous acceptability, poison games and modal logic. In N. Agmon, M. E. Taylor, E. Elkind, and M. Veloso, editors, *Proceedings of AAMAS 2019*, page 1994–1996, 2019.
- [30] D. Grossi and P. Turrini. Short sight in extensive games. In *AAMAS '12: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, page 805–812, 2012.
- [31] S. P. Kikot. Axiomatization of modal logic squares with distinguished diagonal. *Mathematical Notes*, 88:238–250, 2010.
- [32] D. Li. Losing connection: The modal logic of definable link deletion. *Journal of Logic and Computation*, 30:715–743, 2020.
- [33] D. Li. *Formal Threads in the Social Fabric: Studies in the Logical Dynamics of Multi-Agent Interaction*. PhD thesis, Department of Philosophy, Tsinghua University and ILLC, University of Amsterdam, 2021.
- [34] D. Li, S. Ghosh, and F. Liu. Knowing is winning: An epistemic approach to the hide and seek game. In J. van Benthem and F. Liu, editors, *Graph Games and Logic Design: Recent Developments and Future Directions* (To Appear). Springer, Cham, 2025.
- [35] D. Li, S. Ghosh, F. Liu, and Y. Tu. On the subtle nature of a simple logic of the

- hide and seek game. In A. Silva, R. Wassermann, and R. de Queiroz, editors, *Proceedings of WoLLIC 2021*, volume 13038 of *LNCS*, pages 201–218, Cham, 2021. Springer.
- [36] D. Li, S. Ghosh, F. Liu, and Y. Tu. A simple logic of the hide and seek game. *Studia Logica*, 111:821–853, 2023.
- [37] D. Li and K. Sano. Finite axiomatization for the hybrid product logics with diagonal constant. Manuscript, 2025.
- [38] C. Liu, F. Liu, and K. Su. A dynamic-logical characterization of solutions in sight-limited extensive games. In Q. Chen, P. Torroni, S. Villata, J. Hsu, and A. Omicini, editors, *PRIMA 2015: Principles and Practice of Multi-Agent Systems*, volume 9387 of *LNCS*, pages 467–480, 2015.
- [39] F. Liu. Diversity of agents and their interaction. *Journal of Logic, Language and Information*, 18:23–53, 2009.
- [40] R. Nowakowski and P. Winkler. Vertex-to-vertex pursuit in a graph. *Discrete Math*, 13:235–239, 1983.
- [41] J. A. Plaza. Logics of public communications. *Synthese*, 158:165–179, 2007.
- [42] E. Rasmusen. *Games and Information: An Introduction to Game Theory*. Wiley-Blackwell, 4 edition, 2006.
- [43] P. Rohde. *On Games and Logics over Dynamically Changing Structure*. PhD thesis, RWTH Aachen University, 2005.
- [44] K. Sano, F. Liu, and D. Li. Hybrid logic of the hide and seek game. To appear in *Studia Logica*, 2024.
- [45] D. Thompson. Local fact change logic. In F. Liu, H. Ono, and J. Yu, editors, *Knowledge, Proof and Dynamics*, Logic in Asia: Studia Logica Library, pages 73–96, Singapore, 2020. Springer.
- [46] Y. Tu, S. Ghosh, F. Liu, and D. Li. A modal approach towards substitutions. Manuscript, 2025.
- [47] J. Väänänen. *Dependence Logic: A New Approach to Independence Friendly Logic*. Cambridge University Press, Cambridge, 2007.

## Appendix A Proof for Fact 7

*Proof.* We will show the validity of axiom (**Knowledge-Ground**), and prove that both (**K-Additivity**) and (**K-Elimination**) preserve validity. The others are left as exercises to the reader. Let  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  be a  $k$ -sight model and  $\sigma \in \Sigma$ .

(1) First, we show (**Knowledge-Ground**) is valid. W.l.o.g., let  $z := x$  and  $z' := y$ . Assume that  $M, \sigma \models K_x y = \mathcal{T}$ , where  $\mathcal{T} \subseteq \text{Cons}$ . The proof for the case that  $\alpha$  does not contain any occurrence of  $y$  is easy. Assume that the formula does contain  $y$ .

(1.1) Assume that  $M, \sigma \not\models \bigwedge_{c \in \mathcal{T}} \alpha[c/y]$ . Then,  $M, \sigma \not\models \alpha[c/y]$  for some  $c \in \mathcal{T}$ . Since  $c \in \mathcal{T}$ , it follows from  $M, \sigma \models K_x y = \mathcal{T}$  that there is some  $\sigma' \in \Sigma$  s.t.  $\sigma \sim_x \sigma'$  and  $\sigma'(y) = \mathbf{I}(c)$ . By  $M, \sigma \not\models \alpha[c/y]$ ,  $M, \sigma' \not\models \alpha$ , which then gives us  $M, \sigma \not\models K_x \alpha$ .

(1.2) Suppose that  $M, \sigma \not\models K_x \alpha$ . Then, there is a situation  $\sigma' \in \Sigma$  such that  $\sigma \sim_x \sigma'$  and  $M, \sigma' \not\models \alpha$ . By the axiom (**At-Some-Where**), there is some  $c' \in \text{Cons}$  such that  $M, \sigma' \models c_1 \equiv y$ . Clearly,  $c_1 \in \mathcal{T}$ . Since  $M, \sigma' \not\models \alpha$ , it holds that  $M, \sigma' \not\models \alpha[c'/y]$ .

Recall that  $\sigma \sim_x \sigma'$ ,  $\sigma(x) = \sigma'(x)$  (due to the  $0 \leq k$ -sight ability), so  $M, \sigma \not\models \alpha[c'/y]$ , which then shows that  $M, \sigma \not\models \bigwedge_{c \in \mathcal{T}} \alpha[c/y]$ , as needed.

(2) We move to (K-Additivity). Assume that  $\varphi \rightarrow \alpha$  is valid on any  $k$ -sight models based on  $(\mathbf{D}, \mathbf{I})$ . Let  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  be a  $k$ -sight model and  $\sigma, \sigma' \in \Sigma$  with  $\sigma \sim_z \sigma'$  and  $M, \sigma \models \varphi$ . By Fact 3,  $M, \sigma' \models \varphi$ . So,  $M, \sigma' \models \alpha$ , which indicates  $M, \sigma \models K_z \alpha$ .

(3) Finally, we prove that (K-Elimination) preserves the validity of formulas. Assume that  $\varphi \rightarrow (K_z \alpha \rightarrow \beta)$  is valid on any  $k$ -sight models based on  $(\mathbf{D}, \mathbf{I})$ . W.l.o.g., let  $z := x$  and  $z' := y$ . Also, suppose that for some  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  and  $\sigma \in \Sigma$ , it holds that  $M, \sigma \models \varphi \wedge \alpha \wedge \neg \beta$ . Now, define  $M' = (\mathbf{D}, \mathbf{I}, \Sigma, \sim')$  such that for any  $\sigma_1, \sigma_2 \in \Sigma$ ,  $\sigma_1 \sim'_x \sigma_2$  iff  $\sigma_1 = \sigma_2$ , and for the other variable  $y$ ,  $\sigma_1 \sim'_y \sigma_2$  iff  $\sigma_1 \sim_y \sigma_2$ . Now, by the form of  $\varphi$ , it is easy to see that  $M', \sigma \models \varphi$ . Also, by Fact 4,  $M', \sigma \models \alpha \wedge \neg \beta$ . So,  $M', \sigma \models K_x \alpha \wedge \neg \beta$ , a contradiction.  $\square$

## Appendix B Proof for Theorem 3

*Proof.* Based on Fact 7, it suffices to show the validity of the new axioms (R1)-(R10). Let  $M = (\mathbf{D}, \mathbf{I}, \Sigma, \sim)$  be a  $k$ -sight model and  $\sigma_1 \in \Sigma$ . The validity of (R2), (R3), (R4) and (R7) is easy to see. In what follows, we consider (R1), (R5), (R8) and (R10), and the key ideas of the proofs for (R6) and (R9) are similar to those of (R10) and (R5).

(1) We consider (R1). We have the following:

$$\begin{aligned} M, \sigma_1 \models [z]\alpha & \quad \text{iff} \quad \text{For all } \sigma_2 \in \mathbf{R}^z(\sigma_1), (\mathbf{D}, \mathbf{I}, \Sigma', \sim'), \sigma_2 \models \alpha \\ & \quad \text{iff} \quad M, \sigma_1 \models \bigwedge_{\mathcal{T} \subseteq \text{Cons}} (Rz = \mathcal{T} \rightarrow \bigwedge_{c \in \mathcal{T}} \alpha[c/z]) \end{aligned}$$

The last equivalence holds by Fact 5, as needed.

(2) We consider the formula (R5). W.l.o.g., let  $z := x$  and  $z' := y$ . Assume that  $\mathcal{T}, \mathcal{T}_1 \subseteq \text{Cons}$  such that  $M, \sigma_1 \models K_x y = \mathcal{T} \wedge Rx = \mathcal{T}_1$ .

First, let  $M, \sigma_1 \models [x]K_x y$ . Then, for all  $\sigma_2 \in \mathbf{R}^x(\sigma_1)$ ,  $M', \sigma_2 \models K_x y$ . Suppose towards a contradiction that the following is not the case:

$$K_x y \vee \bigwedge_{a \in \mathcal{T}_1} (\mathbf{D}^k a y \vee \bigwedge_{b \in \mathcal{T}} (b \neq y \rightarrow \mathbf{D}^k a b))$$

Since  $M, \sigma_1 \models \neg K_x y$ , the  $\mathcal{T}$  cannot be a set of constants that refer to the same vertex. Also, there are constants  $a \in \mathcal{T}_1$  and  $b \in \mathcal{T}$  s.t.  $M, \sigma_1 \models \neg \mathbf{D}^k a y \wedge b \neq y \wedge \neg \mathbf{D}^k a b$ . Let  $\sigma := \sigma_1[y := b]$ . Since  $b \in \mathcal{T}$ , we have  $\sigma \in \Sigma$  and  $\sigma \sim_x \sigma_1$ , and so  $\sigma \in \Sigma|\sigma_1$ . Moreover, let  $M, \sigma_1 \models y \equiv c$  for some  $c \in \text{Cons}$ . Such a constant always exists (the axiom (At-Some-Where)), and due to the fact  $\sim_x$  is an equivalence relation,  $c \in \mathcal{T}$ . Assume that  $x$  moves to  $a$  and we write  $\sigma_2$  for the new situation (i.e.,  $\sigma_2 := \sigma_1[x := a]$ ). Consider  $\sigma' := \sigma[x := a]$ . Obviously,  $\sigma' \in \mathbf{R}^x(\Sigma|\sigma_1)$ , and it follows from  $M, \sigma_1 \models \neg \mathbf{D}^k a y$  that  $\sigma'(y) \notin \mathbf{D}^k(\sigma_2)$ . So,  $\sigma' \in \Sigma'$ . Now,  $M', \sigma \models \langle K_x \rangle y \equiv b \wedge y \equiv c \wedge b \neq c$ , a contradiction.

For the converse, assume that  $M, \sigma_1 \not\models [x]K_x y$ , i.e.,  $M', \sigma_2 \models \neg K_x y$  for some  $\sigma_2 \in \mathbf{R}^x(\sigma_1)$ . Then there is a  $\sigma' \in \Sigma'$  s.t.  $\sigma_2(x) = \sigma'(x)$  and  $\sigma_2(y) \neq \sigma'(y)$ . We consider the previous stage  $\sigma := \sigma'[x := \sigma_3(x)]$  of  $\sigma'$ , where  $\sigma_3 \in \Sigma|\sigma_1$  (so  $\sigma$  is  $\sigma_3$ ). Now, at least one of  $\sigma \sim_x \sigma_1$  and  $\sigma \sim_y \sigma_1$  must be the case. Notice that  $\sigma(y) = \sigma'(y)$  and  $\sigma_1(y) = \sigma_2(y)$ . Since  $\sigma_2(y) \neq \sigma'(y)$ , we have  $\sigma \not\sim_y \sigma_1$  (due to the  $0 \leq k$ -sight ability), which then gives us  $\sigma \sim_x \sigma_1$  and so  $M, \sigma_1 \not\models K_x y$ . Recall the assumption

that  $M, \sigma_1 \models K_x y = \mathcal{T} \wedge Rx = \mathcal{T}_1$ . Moreover, assume that  $a$  is a constant with  $M', \sigma_2 \models x \equiv a$ . By (At-Some-Where), such a constant always exists, and so  $a \in A$ . Also, let  $b$  be a constant such that  $M, \sigma \models y \equiv b$ . Then, we can see that  $b \in \mathcal{T}$ . Now, we have  $M, \sigma_1 \models \neg D^k a y \wedge b \not\equiv y \wedge \neg D^k a b$ , as needed.

(3) We move to (R8). Note that for any  $\sigma_2 \in R^z(\sigma_1)$ ,  $\sigma_1(z') = \sigma_2(z')$ , where  $z'$  is the variable distinct from  $z$ . By Fact 4,  $M, \sigma_1 \models \alpha$  iff  $M', \sigma_2 \models \alpha$ . Since  $\alpha$  does not contain  $z'$ ,  $M', \sigma_2 \models \alpha$  iff  $M', \sigma_2 \models K_z \alpha$ . Now,  $M, \sigma_1 \models [z]K_z \alpha$  iff  $M, \sigma_1 \models \alpha$ .

(4) Let us consider (R9). Let  $z := x$  and  $z' := y$ . Also, let  $\mathcal{T}, \mathcal{T}_1 \subseteq Cons$  such that  $M, \sigma_1 \models Rx = \mathcal{T} \wedge K_x y = \mathcal{T}_1$ .

(4.1) For the direction from left to right, assume that  $M, \sigma_1 \models [x]K_x \alpha(y)$ , i.e., for any  $\sigma_2 \in R^x(\sigma_1)$ ,  $M', \sigma_2 \models K_x \alpha(y)$ . Suppose that the following is not the case:

$$(K_x y \wedge \bigwedge_{c \in \mathcal{T}} \alpha[c/x]) \vee \\ (\neg K_x y \wedge \bigwedge_{t \in \mathcal{T}} ((D^k t y \wedge \alpha[t/x]) \vee (\neg D^k t y \wedge \bigwedge_{c \in \mathcal{T}_1} (\neg D^k t c \rightarrow \alpha[t/x][c/y])))$$

(4.1.1) We first assume that  $M, \sigma_1 \models K_x y$ . Then, by the first disjunct, there is some  $c \in \mathcal{T}$  such that  $M, \sigma_1 \not\models \alpha[c/x]$ . But by Fact 5, we arrive at a contradiction.

(4.1.2) Assume that  $M, \sigma_1 \not\models K_x y$ . Then, there is a constant  $c \in \mathcal{T}$  such that exactly one of the following is the case:

(a)  $D^k c y$  and  $\neg \alpha[c/x]$

(b)  $\neg D^k c y$ , and there is some  $c_1 \in \mathcal{T}_1$  such that  $\neg D^k c c_1$  and  $\neg \alpha[c/x][c_1/y]$ .

We fix an assignment  $\sigma' := \sigma_1[x := \mathbf{I}(c)]$ . When (a) is the case, by definition, we have  $\Sigma' = \{\sigma'\}$  in the pointed model  $M', \sigma'$ . Then, it follows from  $\neg \alpha[c/x]$  that  $M', \sigma' \not\models \alpha$ , which gives us  $M', \sigma' \not\models K_x \alpha$ , a contradiction. Let us assume that (b) is the case. Since there is some  $c_1 \in \mathcal{T}_1$  with  $\neg D^k c c_1$ , for the assignment  $\sigma_3 := \sigma_1[y := \mathbf{I}(c_1)]$  and  $\sigma'_3 := \sigma_3[x := \mathbf{I}(c)]$ , we have  $\sigma_3 \sim_x \sigma_1$  and  $\sigma'_3 \in \Sigma'$ . It holds that  $\sigma' \sim'_x \sigma'_3$ . Also,  $M', \sigma'_3 \models \neg \alpha[c/x][c_1/y]$ , and so  $M', \sigma' \not\models K_x \alpha[c/x]$ , which gives us  $M', \sigma' \not\models K_x \alpha$ .

(4.2) We consider the other direction. Suppose that  $M, \sigma_1 \not\models [x]K_x \alpha(y)$ , i.e., there is some  $\sigma_2 \in R^x(\sigma_1)$  with  $M', \sigma_2 \not\models K_x \alpha(y)$ . We discuss different cases.

(4.2.1) We assume that  $M, \sigma_1 \models K_x y$ . Then,  $M', \sigma_2 \models K_x y$  (Fact 6). Given this, it follows from  $M', \sigma_2 \not\models K_x \alpha(y)$  that  $M', \sigma_2 \not\models \alpha(y)$ . By Fact 5,  $M, \sigma_1 \not\models \bigwedge_{c \in \mathcal{T}} \alpha[c/x]$ .

(4.2.2) Assume  $M, \sigma_1 \not\models K_x y$ . Then, there is an assignment  $\sigma \in \Sigma$  s.t.  $\sigma_1 \sim_x \sigma$  and  $\sigma_1(y) \neq \sigma(y)$ . Let  $\mathbf{I}(c_1) = \sigma(y)$  and  $\mathbf{I}(c) = \sigma_2(x)$ . Clearly,  $c_1 \in \mathcal{T}_1$  and  $c \in \mathcal{T}$ . There are two different cases. We first suppose that  $\sigma_2(x) \in \mathbb{D}^k(\mathbf{I}(c_1))$ . Then, the  $\Sigma'$  in  $M'$  is  $\{\sigma_2\}$ , and the fact  $M', \sigma_2 \not\models K_x \alpha(y)$  indicates  $M', \sigma_2 \not\models \alpha$ , and so,  $M', \sigma_2 \not\models \alpha[c/x]$ . Next, assume  $\sigma_2(x) \notin \mathbb{D}^k(\mathbf{I}(c_1))$ . Since  $M', \sigma_2 \not\models K_x \alpha(y)$ , there is  $\sigma'_3 \in \Sigma'$  s.t.  $\sigma_2 \sim_x \sigma'_3$  and  $M', \sigma'_3 \not\models \alpha(y)$ . Let  $c_2, c_3 \in Cons$  s.t.  $M', \sigma'_3 \models c_2 \equiv y \wedge c_3 \equiv x$ . Clearly,  $c_2 \in \mathcal{T}_1$  and  $c_3 \in \mathcal{T}$ . It follows from  $\sigma'_3 \in \Sigma'$  and  $\sigma_2(x) \notin \mathbb{D}^k(\mathbf{I}(c_1))$  that  $\sigma'_3(x) \notin \mathbb{D}^k(\sigma'_3(y))$ , i.e.,  $M', \sigma'_3 \models \neg D^k c_3 c_2$ . Since  $M', \sigma'_3 \not\models \alpha(y)$ ,  $M', \sigma'_3 \not\models \alpha[c_3/x][c_2/y]$ .

Putting (4.2.1) and (4.2.2) together, we can get the negation of the right part of the equivalence (R9), as needed.  $\square$