A Logical Analysis of Hanabi

Elise Perrotin

AIST, Tokyo, Japan
elise.perrotin@aist.go.jp

Abstract

The card game Hanabi has recently gained popularity as a benchmark for handling epistemic reasoning in AI systems. However it has until now mostly been approached through the lens of machine learning rather than formal logical analysis. This is mostly due to the fact that modeling Hanabi in the standard epistemic logic DEL is untractable. In this paper we take a different approach to formalizing Hanabi, using the simple epistemic logic EL-O as a starting point. We generalize common knowledge in EL-O to arbitrary groups of agents and show how to overcome some of the limitations EL-O places on agent reasoning by introducing a special reasoning action. Analyzing our formalization of Hanabi finally leads us to introduce an alternative semantics for our generalization of EL-O in which models are finite and satisfiability checking is NP-complete, and which is enough to fully describe the evolution of knowledge in a game of Hanabi.

Introduction

The importance of handling epistemic reasoning in AI systems has been receiving increasing recognition, as it is crucial in multi-agent situations with incomplete information and in particular for human-robot collaboration. The card game Hanabi has been proposed as a benchmark for advances in this field (Bard et al. 2020). Hanabi is a cooperative card game with partial information, where players can see other players' cards but not their own. Players must then coordinate through restricted means of communication to play cards in a particular order. Theory of mind is required in order to communicate efficiently and reason about why other players have made certain choices.

Most of the literature on Hanabi involves a combination of heuristics and reinforcement learning to create a system capable of mimicking theory of mind well enough to play the game (see e.g. Aru et al. (2023); Walton-Rivers, Williams, and Bartle (2019)). Attempts at formal representations, on the other hand, are typically limited to small fragments of the game (Chetcuti-Sperandio et al. 2020; Gamblin, Niveau, and Bouzid 2022). This is mostly due to the untractability of the standard Dynamic Epistemic Logic (DEL) (Van Ditmarsch, van Der Hoek, and Kooi 2007), in which must be

represented not only all worlds that agents consider possible, but also all worlds they consider that other agents might think possible, etc. This makes for very large and impractical models. In particular satisfiability in multi-agent DEL with common knowledge is known to be 2ExpTime-complete (Charrier et al. 2018).

DEL being too computationally heavy for practical applications is not a new problem, and a number of lightweight fragments have been studied, generally with tractable epistemic planning as a goal. One of these proposals is the Epistemic Logic of Observation, or EL-O (Perrotin 2021). While the standard knowledge modality is that of 'knowing that'—$K_a\varphi$ reads "agent a knows that φ is true"—EL-O uses the alternative 'knowing whether' operator, also interpreted as 'seeing': $S_a\varphi$ reads "agent a knows whether φ is true", or, put differently, "agent a sees (the truth value of) φ". The two operators are interdefinable: knowing that a formula is true amounts to seeing a formula that is true (i.e., $K_a\varphi$ is equivalent to $\varphi \wedge S_a\varphi$), and seeing a formula amounts to either knowing that it is true, or knowing that it is false (i.e., $S_a\varphi$ is equivalent to $K_a\varphi \vee K_a\neg\varphi$). The scope of this 'seeing' operator is then restricted, so that agents can see propositional formulas and whether other agents see propositional formulas, etc., but they cannot know disjunctions in general. That is, formulas of the form $S_a(\varphi \vee \psi)$ are disallowed. This restriction limits the reasoning power of agents, and in the end ensures propositional complexity results by translation into propositional logic: the EL-O satisfiability problem is NP-complete, and the extension of EL-O with STRIPS-like actions leads to PSpace-complete planning.

In this paper we explore a formalization of Hanabi in EL-O. Some of the static concepts of the game, such as knowing whether a card is of a certain color, have a natural representation in EL-O. However, two main difficulties arise. The first is that of common knowledge within groups of agents that are not the full set of agents. Common knowledge in EL-O is only defined for the full set of agents, and finite semantics for an extension of EL-O with common knowledge for any group of agents is still an open problem. However the latter is necessary in order to faithfully represent Hanabi in EL-O while maintaining finiteness of actions descriptions. Consider for example the following scenario:

Scenario 1. There are five players, and the cards have just been distributed. By chance, all 5's are already in players'

hands. Specifically, Player 1 holds two 5's and Player 2 holds the remaining three 5's. Then Players 3, 4 and 5 commonly know that all other cards, including those in their own hands and those in the stack, are not 5's. Players 1 and 2 do not share this knowledge.

To tackle this issue, we will start by defining infinite semantics for EL-O with 'any group' common knowledge and a specific class of actions. We will then show that in the case of Hanabi, modal depth can be restricted to at most 2 w.l.o.g., and hence we can return to finite semantics.

The second difficulty lies in the apparent inability of agents to reason in EL-O: as they cannot know disjunctions, agent a cannot explicitly know that a card i is either blue or red, and, upon learning that it is not blue, perform the reasoning $(K_a(Blue_i \vee Red_i) \wedge K_a \neg Blue_i) \rightarrow K_a Red_i$. In this paper we propose to delegate reasoning to a special reasoning action. Though player a cannot know that card i is either blue or red, they can know that it is neither white, green nor yellow. Upon learning that it is not blue, the following conditional effect is activated in the reasoning action:

- Condition: $\neg White_i \wedge S_a White_i \wedge \neg Green_i \wedge S_a Green_i \wedge \neg Yellow_i \wedge S_a Yellow_i \wedge \neg Blue_i \wedge S_a Blue_i$
- Effect: $S_a Red_i$ becomes true.

Here the condition states that player a knows that card i is not white ($\neg White_i \wedge S_a White_i$), green, yellow or blue. The effect is that player a now sees whether it is red (which it is).

This reasoning action is activated every time players learn something. It can be performed until the players have deduced everything they can from their knowledge, or alternatively be limited to a certain number of steps. This ties back in to the problem of logical omniscience in epistemic reasoning: the standard principles of modal logic imply that agents should know all tautologies and all possible consequences of their knowledge. This is unrealistic when modeling human knowledge, which is standardly understood as knowledge the human is (to some degree) aware of, and might take some effort to get to even if that human has all the elements in hand to get to the right conclusion. This effort is, in our case, explicitly represented by the reasoning action. As such a player might learn that a card is neither white, green, yellow nor blue, but not consciously realize that that means they know it is red; if they do realize it is red, they might not further observe that they now know the position of all red 2's in the game, and therefore the other 2 they know is in their hand cannot be red; if they do realize that, they might not realize that that other 2 must be blue, as they have formerly learned that it is neither white, green nor yellow; etc.

The paper is organized as follows. After recalling the rules of Hanabi, we present our generalization of EL-O with common knowledge among arbitrary groups of agents. We then introduce all necessary fluents for our formalization of Hanabi, the possible initial states, and the actions, including the reasoning action. An analysis of this formalization finally leads us to introduce a finite semantics simulating it.

Background: Rules of Hanabi

Hanabi is a cooperative turn-based card game for two to five players. The fifty cards each have one of five colors (blue, white, green, red, and yellow) as well as a number from 1 to 5, so that in each color there are three 1's, two each of 2's, 3's and 4's, and one 5. There are also eight "hint" tokens and three "error" tokens. The game starts with each player holding five cards, which are facing away from them so that they see the other players' cards but not their own.

The aim is to place cards on the table (referred to as "playing" the cards) in a particular order: for each color, first a 1 must be played, then a 2, and so on. If a player attempts to play a card that may not be played, an error token is received, and that card is discarded; either way, the player draws a new card after playing.

The players can only play their own cards. In order to know what to play, they can communicate hints of a particular form: they can tell another player the position of all of their cards of a certain color (e.g. "these three cards in your hand are your red cards"), or of a certain number (e.g. "you have only one 5, here"). Giving a hint uses up a hint token. To replenish hint tokens, players may discard one of the cards in their hand and draw a new one from the deck. Successfully playing a 5 also replenishes a hint token.

To summarize, at their turn each player may:

- Attempt to play a card and then draw another card (incurs an error token if unsuccessful, replenishes a hint token if successful and the played card is a 5);
- Give a hint (uses up a hint token);
- Discard a card and then draw another card (replenishes a hint token).

Playing may continue until one round after the deck becomes empty. The game is won if the 5's of each color are successfully played without receiving all of the error tokens and before running out of cards.

EL-O with Generalized Common Knowledge

We now introduce the logic EL-O extended with common knowledge for arbitrary groups of agents.

Language

Given a set $Prop$ of propositional variables and Agt of agents, the set ATM of *observability atoms* (or atoms for short) is defined as follows:

$$ATM \ni \alpha ::= p \mid JS_G \alpha$$

where p ranges over $Prop$ and G ranges over $2^{Agt} \setminus \{\emptyset\}$. That is, G is any non-empty set of agents of Agt. The formula $JS_G \alpha$ reads "all agents in G jointly see α" (that is, there is common knowledge between the agents of G of the truth value of α). The operators of individual and common knowledge in standard EL-O correspond to the particular cases where G is either a singleton or the set of all agents. For readability we write $JS\alpha$ instead of $JS_{Agt}\alpha$.

The operators JS_G are called *observability operators*. Given a group G of agents, we denote by Seq_G the set of sequences of observability operators indexed by subgroups of G, and by Seq_G^+ the set of non-empty such sequences. Formally:

$$\mathsf{Seq}_G^+ = \{JS_{G_1} \ldots JS_{G_n} \mid n \geqslant 1 \text{ and } \forall k \leqslant n, G_k \subseteq G\}$$

and $\mathsf{Seq}_G = \mathsf{Seq}_G^+ \cup \{\varnothing\}$, where $\varnothing$ denotes the empty sequence. Any observability atom can then be written as σp where $\sigma \in \mathsf{Seq}_{Agt}$. The *modal depth* of an atom σp is the number of operators in σ. For any $k \geqslant 0$, we denote by $ATM^{\leqslant k}$ the set of atoms of modal depth at most k.

Formulas are boolean combinations of atoms:

$$\mathcal{L}_{\mathsf{EL\text{-}O}} \ni \varphi ::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi$$

where α ranges over ATM. We sometimes use $K_G\alpha$ (resp. $K_G\neg\alpha$) as an abbreviation of $\alpha \wedge JS_G\alpha$ (resp. $\neg\alpha \wedge JS_G\alpha$).

Semantics

In order to define semantics, we need two notions which will allow us to ensure the right properties of knowledge: *introspective atoms* and *atomic consequence*. Introspective atoms are atoms that contain some $JS_G JS_H$ with $G \subseteq H$. Intuitively, they are atoms that should always be true: a group always knows whether a bigger group they are included in have common knowledge of the value of some α, and this property of knowledge is commonly known by all agents. The set of introspective atoms is denoted by I-ATM.

An atom β is said to be an *atomic consequence* of another atom α, and α an *atomic cause* of β, denoted by $\alpha \Rightarrow \beta$, when $\alpha = JS_G\alpha'$ and $\beta = \sigma\alpha'$ for some group G of agents, some atom α' and some non-empty sequence $\sigma \in \mathsf{Seq}_G^+$ of observability operators. Intuitively, this is a consequence relation that should always be verified: if there is joint vision of α within G then all subgroups of G should see α, all subgroups of G should see that all subgroups of G see α, etc.

While the above is enough to move forward with semantics in standard EL-O, we here need to go a step further and consider that it should also be commonly known that atomic consequence is verified. That is, if $\alpha \Rightarrow \beta$ then if a group G of agents knows *that* α is true, then they should also know that β is true; and if they know that β is false then they should also know that α is false. Moreover, if another group H of agents knows that G knows that α is true, then they should know that G knows that β is true, and so on. This is trivial in standard EL-O as the only non-trivial case of atomic consequence is $JS\alpha \Rightarrow \sigma\alpha$, where σ is any sequence of observability operators. Then if agents know that $JS\alpha$ is true then any consequence follows from $JS\alpha$ itself, and agents always know whether $JS\alpha$ is true.

To formalize this, what we want is to consider valuations $s \subseteq ATM$ such that for any atoms α and β such that $\alpha \Rightarrow \beta$ and any sequence $G_1 \dots G_n$ of groups of agents with $n \geqslant 0$, if $K_{G_1} \dots K_{G_n}\alpha$ is true in s then $K_{G_1} \dots K_{G_n}\beta$ is true in s, and if $K_{G_1} \dots K_{G_n}\neg\beta$ is true in s then $K_{G_1} \dots K_{G_n}\neg\alpha$ is true in s. Writing this with observability operators we get the two following conditions:

1. If $\sigma\alpha \in s$ for all $\sigma \sqsubseteq JS_{G_1} \dots JS_{G_n}$, then $JS_{G_1} \dots JS_{G_n}\beta \in s$;
2. If $n \geqslant 1$, $\beta \notin s$, and $\sigma\beta \in s$ for all non-empty $\sigma \sqsubseteq JS_{G_1} \dots JS_{G_n}$, then $JS_{G_1} \dots JS_{G_n}\alpha \in s$;

where $\sqsubseteq$ is the subsequence relation. For example, the set of subsequences of $JS_{G_1} JS_{G_2} JS_{G_3}$ is $\{\varnothing, JS_{G_1}, JS_{G_2}, JS_{G_3}, JS_{G_1} JS_{G_2}, JS_{G_1} JS_{G_3}, JS_{G_2} JS_{G_3}, JS_{G_1} JS_{G_2} JS_{G_3}\}$.

We call item 1 above *positive generalized atomic consequence* and item 2 *negative generalized atomic consequence*. Note that standard atomic consequence corresponds to the case $n = 0$ in the former.

Negative generalized atomic consequence is the main source of difficulty when generalizing EL-O common knowledge to subgroups of agents. In standard EL-O, as the generalization of atomic consequence is trivial as noted above, any set of atoms can be transformed into a valuation satisfying properties of S5 knowledge by taking its closure under atomic consequence and adding in introspective atoms. Semantics can then be defined over arbitrary sets of atoms by considering that an atom is true in a set if it is true in the corresponding valuation. From there it has been shown that standard EL-O has the finite model property: any EL-O-satisfiable formula is satisfied in a finite state. However, in the generalized setting negative atomic consequence depends not only on which atoms *are* in the state, but on which atoms *aren't*, making defining a valuation from an arbitrary set of atoms more difficult, and still an open problem.

We do not aim to resolve this problem in this paper, but will instead show that it can be sidestepped in the case of Hanabi. We first finish with semantics by defining *states* as sets of atoms that satisfy both positive and negative generalized atomic consequence, and that include all introspective atoms. We call $STATES$ the set of all states. Semantics are then standard propositional semantics: for any state $s \in STATES$ and formula $\varphi \in \mathcal{L}_{\mathsf{EL\text{-}O}}$, $s \models \varphi$ iff $s \models^{\mathsf{CPC}} \varphi$, where $\models^{\mathsf{CPC}}$ is propositional entailment. We write $\varphi \models \psi$ to denote that any model of φ is a model of ψ, that is, for any $s \in STATES$, if $s \models \varphi$ then $s \models \psi$.

Actions

EL-O actions are essentially STRIPS actions adapted to the EL-O syntax and semantics. We here define a specific class of actions that behave well in our generalized setting and are enough to formalize Hanabi.

Syntax Actions have preconditions describing when that action can take place, and conditional effects describing the exact effects of an action depending on additional conditions. For example, playing a card will lead to different results depending on whether that card is one that is allowed to be played (in which case it is placed on the table) or not (in which case it is discarded and an error token is received).

Formally, an EL-O *action description* (or *action* for short) a is a pair $\big(pre(\mathsf{a}), \mathit{eff}(\mathsf{a})\big)$ where $pre(\mathsf{a}) \in \mathcal{L}_{\mathsf{EL\text{-}O}}$ is the *precondition* of a and $\mathit{eff}(\mathsf{a}) \subseteq \mathcal{L}_{\mathsf{EL\text{-}O}} \times 2^{ATM} \times 2^{Prop}$ is the set of its *conditional effects*. For any conditional effect $ce = (cnd(ce), \mathit{ceff}^+(ce), \mathit{ceff}^-(ce))$ in $\mathit{eff}(\mathsf{a})$, we call $cnd(ce)$ the *condition* of ce, $\mathit{ceff}^+(ce)$ its *positive effects* (the atoms it makes true) and $\mathit{ceff}^-(ce)$ its *negative effects* (the atoms it makes false). Note that we only allow propositional negative effects here. That is, we do not allow knowledge to be lost.

We make two more requirements on actions. The first is that they should be *consistent*: for any $ce, ce' \in \mathit{eff}(\mathsf{a})$, if $\mathit{ceff}^+(ce) \cap \mathit{ceff}^-(ce') \neq \emptyset$ then $pre(\mathsf{a}) \wedge cnd(ce) \wedge cnd(ce')$ must be EL-O-unsatisfiable. This ensures that an action does not make an atom both true and false at the same time.

The last requirement ensures that groups can only learn about whether a group G knows whether α if G is also learning the value of α. Formally, we say that an action a is *knowledge-positive* if for all $ce \in \mathit{eff}(\mathsf{a})$ and $JS_H JS_G \alpha \in \mathit{ceff}^+(ce)$, there exists $ce' \in \mathit{eff}(\mathsf{a})$ such that $JS_G \alpha \in \mathit{ceff}^+(ce')$ and $\mathit{pre}(\mathsf{a}) \wedge \mathit{cnd}(ce) \models \mathit{cnd}(ce')$.

Semantics Let $\mathsf{a} = \big(\mathit{pre}(\mathsf{a}), \mathit{eff}(\mathsf{a})\big)$ be a consistent and knowledge-positive action description. The function τ_a is a partial function over $STATES$ such that for any state $s \in STATES$, $\tau_\mathsf{a}(s)$ is defined if $s \models \mathit{pre}(\mathsf{a})$; in that case we say that a is *applicable at* s. When a is applicable at s, $\tau_\mathsf{a}(s)$ is defined by removing from s all negative effects and adding all positive effects of applicable conditional effects, and taking the closure of the resulting set under positive generalized atomic consequence. Formally, we define

$$\tau_\mathsf{a}^0(s) = \left(s \setminus \bigcup_{\substack{ce \in \mathit{eff}(\mathsf{a}) \\ s \models \mathit{cnd}(ce)}} \mathit{ceff}^-(ce)\right) \cup \bigcup_{\substack{ce \in \mathit{eff}(\mathsf{a}) \\ s \models \mathit{cnd}(ce)}} \mathit{ceff}^+(ce)$$

and for all $k \geqslant 0$,

$$\tau_\mathsf{a}^{k+1}(s) = \tau_\mathsf{a}^k(s) \cup \{JS_{G_1} \ldots JS_{G_n} \beta \mid \exists \alpha \Rightarrow \beta,$$
$$\forall \sigma \sqsubseteq JS_{G_1} \ldots JS_{G_n}, \sigma \alpha \in \tau_\mathsf{a}^k(s)\}.$$

We then define $\tau_\mathsf{a}(s) = \bigcup_{k \geqslant 0} \tau_\mathsf{a}^k(s)$.

Proposition 1. *Let* a *be a consistent knowledge-positive action. For any* $s \in STATES$, *if* $\tau_\mathsf{a}(s)$ *is defined then* $\tau_\mathsf{a}(s) \in STATES$.

In the next section we present a formalization of Hanabi in this generalized EL-O framework.

A Formalization of Hanabi

Fluents and Initial States

Initially we consider that the deck of cards has been shuffled, and number the cards as they appear in that deck, from 1 to 50. For $i \leqslant 50$, each card i has a color ($Yellow_i$, $Blue_i$, Red_i, $White_i$, $Green_i$) and a number ($1_i, 2_i, 3_i, 4_i, 5_i$) such that there is the right number of each card. The set of all colors is $\mathbf{Colors} = \{Yellow, Blue, Red, White, Green\}$. The set of all card attributes is $\mathbf{A} = \mathbf{Colors} \cup [\![1, 5]\!]$. For any $c \in \mathbf{Colors}$ and $n \in [\![1, 5]\!]$, we call $\mathsf{nb}(c, n)$ the number of cards that have the color c and the number n, and for any $A \in \mathbf{A}$, we call $\mathsf{nb}(A)$ the number of cards that have the attribute A.

The set of players is denoted by $Agt = [\![1, \mathsf{np}]\!]$. Initially each player has five cards; a card i being in player a's hand is represented by $Belongs_{i,a}$. Initially player 1 has cards 1 to 5, player 2 has cards 6 to 10, etc.; that is, $Belongs_{i,1}$ is true iff $1 \leqslant i \leqslant 5$, $Belongs_{i,2}$ is true iff $6 \leqslant i \leqslant 10$, etc. We denote by $\mathsf{Init}(a)$ the set of indexes of cards initially in the hand of player a. That is, $\mathsf{Init}(a) =]\!]5(a-1), 5a]\!]$. We generalize this to groups of agents by defining $\mathsf{Init}(G) = \bigcup_{a \in G} \mathsf{Init}(a)$.

Card i being on the table (i.e., having been successfully played at an earlier point in the game) is represented by the fluent $OnTable_i$. There are no cards on the table initially, and the rest of the cards are in the stack. The fluent $Next_i$

expresses the fact that the stack starts at card i: initially only $Next_{5\mathsf{np}+1}$ is true. $Next_{51}$ indicates that the stack is empty.

The fluent $Turn_a$ expresses the fact that it is player a's turn to play. $Turn_{\mathsf{np}+1}$ refers to $Turn_1$. Player a needing to draw a card is represented by $MustDraw_a$. Players needing to reason is represented by $Reason$.

We also need to keep track of error tokens and hint tokens. For $0 \leqslant k \leqslant 3$, $ErrToken_k$ represents the fact that k error tokens have been received, and for $0 \leqslant k \leqslant 8$, $HintToken_k$ represents the fact that k hint tokens are available. In the initial state, only $ErrToken_0$ and $HintToken_8$ are true.

Finally, let us consider the observations of agents at each state. It is common knowledge that any group $G \subseteq Agt$ has common knowledge of all cards in the hands of players outside of G. It is tempting to go further say that it is common knowledge what each player knows and doesn't know at the very beginning of the game. However, recall Scenario 1: players 3, 4 and 5 know from the start whether any card is a 5, but players 1 and 2 are not aware that they know that. In our setting, we could technically say that it is initially common knowledge that players have no knowledge outside of what they directly see, and that this only changes when the players first reason about what they see. For simplicity, we prefer to constrain initial states so that they contain the fluent $Reason$, that is, so that a game must start with agents reasoning; and we consider that agents do not start with higher order knowledge of what they don't directly see.

Formally, we define the set $INIT$ of initial states of Hanabi as the set of sets of the form

$\mathsf{I\text{-}ATM} \cup \mathsf{CardAtb} \cup \{Belongs_{i,a} \mid a \in Agt, i \in \mathsf{Init}(a)\} \cup$

$\{Next_{5\mathsf{np}+1}, Turn_1, ErrToken_0, HintToken_8, Reason\} \cup$

$\{\sigma JS_G A_i \mid \sigma \in \mathsf{Seq}_{Agt}, G \subseteq Agt, A \in \mathbf{A}, i \in \mathsf{Init}(Agt \setminus G)\},$

where $\mathsf{CardAtb} \subseteq \{A_i \mid A \in \mathbf{A}, i \leqslant 50\}$ is such that for all $i \leqslant 50$, there are exactly one $c \in \mathbf{Colors}$ and one $n \in [\![1, 5]\!]$ such that $c_i \in \mathsf{CardAtb}$ and $n_i \in \mathsf{CardAtb}$, and for any color $c \in \mathbf{Colors}$ and $k \in [\![1, 5]\!]$, there are exactly $\mathsf{nb}(c, n)$ indices $i \leqslant 50$ such that both c_i and n_i are in $\mathsf{CardAtb}$.

Proposition 2. $INIT \subseteq STATES$.

Actions

We now describe how to represent actions in Hanabi as EL-O actions. These actions are: giving a hint to another player; playing, discarding, and drawing a card; and finally the special reasoning action.

Giving a Hint Player a announces to player b (where $b \neq a$) all of b's cards that have a given attribute A. In order for a to give a hint to b:

• It must be player a's turn;

• No reasoning must need to be done;

• There must be at least one hint token available.

  When a gives b a hint about attribute A:

1. Players gain common knowledge of which of b's cards have the attribute A and which don't;

2. There is now one less hint token available;

3. It is no longer a's turn;

4. Players must now reason on their new knowledge.

Formally, the precondition of GiveHint(a, b, A) is:

$$Turn_a \wedge \neg Reason \wedge \bigvee_{k>0} HintToken_k.$$

The conditional effects are:

1. For all $i \leqslant 50$:
   - Condition: $Belongs_{i,b}$
   - Positive effect: JSA_i

2. For all $k \in [\![1, 8]\!]$:
   - Condition: $HintToken_k$
   - Positive effect: $HintToken_{k-1}$
   - Negative effect: $HintToken_k$

3. Next player's turn and reasoning:
   - Condition: $\top$
   - Positive effect: $Reason$, $Turn_{a+1}$
   - Negative effect: $Turn_a$

Playing a card In order for a player to play a card, it must be their turn, no reasoning must need to be done, and the card must belong to them. The player loses the card, and common knowledge is gained of all attributes of the card. The player must then draw a card. In case of error, an error token is received; otherwise the card is placed on the table.

Formally, the precondition of Play(i, a) is:

$$Turn_a \wedge \neg Reason \wedge Belongs_{i,a}.$$

To describe the conditional effects, let us first write the condition for playing a card i to be successful:

$$PlaySuccess_i = \neg \bigvee_{j \neq i} (OnTable_j \wedge SameCard_{i,j})$$
$$\wedge \bigvee_{j \neq i} (OnTable_j \wedge Successor_{i,j})$$

where $SameCard_{i,j}$ and $Successor_{i,j}$ are defined as abbreviations of formulas expressing respectively that i and j have the same color and number, and that i and j have the same color with the number of i being the successor of that of j. The conditional effects of Play(i, a) are then:

1. General:
   - Condition: $\top$
   - Positive effects: JSA_i for all $A \in \mathbf{A}$
   - Negative effects: $Belongs_{i,a}$, $Turn_a$

2. Success case:
   - Condition: $PlaySuccess_i$
   - Positive effect: $OnTable_i$, $MustDraw_a$

3. Case of a 5 being successfully played, for $k < 8$:
   - Condition: $HintToken_k \wedge 5_i \wedge PlaySuccess_i$
   - Positive effect: $HintToken_{k+1}$
   - Negative effect: $HintToken_k$

4. Regular error case, for $k < 2$:
   - Condition: $\neg PlaySuccess_i \wedge ErrToken_k$
   - Positive effect: $ErrToken_{k+1}$, $MustDraw_a$
   - Negative effect: $ErrToken_k$

5. Final error case:
   - Condition: $\neg PlaySuccess_i \wedge ErrToken_2$
   - Positive effect: $ErrToken_3$
   - Negative effect: $ErrToken_2$

In case of error when two error tokens had already been drawn, this action ensures that no further action can be executed, as all $Turn_b$ and $MustDraw_b$ become false for $b \in Agt$ and falseness of $Reason$ is maintained.

Discarding a card Discarding a card is similar to playing a card, except there is no success or error, and the card simply stops belonging to the player.

Formally, the precondition of Discard(i, a) is:

$$Turn_a \wedge \neg Reason \wedge Belongs_{i,a}.$$

There is one conditional effect:

- Condition: $\top$
- Positive effects: $MustDraw_a$, JSA_i for all $A \in \mathbf{A}$
- Negative effects: $Belongs_{i,a}$, $Turn_a$

Drawing a card In order for a player to draw a card, they must have just put down or discarded a card, and the stack of cards must not be empty. The card on top of the stack is then attributed to the player, and all other players see the value of the card.

Formally, the precondition of Draw(a) is:

$$MustDraw_a \wedge \neg Next_{51}.$$

The conditional effects are:

1. For all $i \leqslant 50$:
   - Condition: $Next_i$
   - Positive effects: $Belongs_{i,a}$, $Next_{i+1}$, $JS_{Agt \setminus a}A_i$ and $JSJS_{Agt \setminus a}A_i$ for all $A \in \mathbf{A}$
   - Negative effect: $Next_i$

2. Next player's turn and reasoning:
   - Condition: $\top$
   - Positive effect: $Reason$, $Turn_{a+1}$
   - Negative effect: $MustDraw_a$

Reasoning Players may go through a number of reasoning steps in order to make deductions from their new knowledge. Here we suppose that reasoning is done until there is nothing left to learn; that is, we define the reasoning action so that it is repeated until it no longer has any effect on the state. Alternatively, one could adjust the conditions so that only a given number of reasoning steps can be performed.

The reasoning goes as follows:

1. If e.g. all yellow 3s and 4s are now jointly seen by some group G, then group G also gains common knowledge that all other cards that are jointly observed not to be 1's, 2's or 5's are not yellow;

2. If e.g. all yellow and green 4's are now jointly seen by some group G, then group G also gains common knowledge that all other cards that are jointly observed not to be red, white or blue are not 4's;

3. If any cards are now jointly known by some group G to not be four different colors, they are now jointly known by group G to be the fifth color;

4. If any cards are now jointly known by some group G to not be four different numbers then they are now jointly known by group G to be the fifth number;

5. If all cards of a given attribute are jointly seen by some group G then it is now also jointly seen by G that no other card has this attribute.

We then check whether we can move on to the next turn:

6. If no reasoning has been performed during this action (i.e. all other conditional effects did not modify the current state), it becomes the next player's turn.

Formally, the precondition of Reason is $Reason$. In order to describe its conditional effects, let us first introduce an abbreviation. The following expresses that group G sees all cards that have color c and number n, and none of those cards are indexed by i:

$$KAll(G, c, n, i) = \bigvee_{\substack{I \subset [\![1,50]\!] \\ |I| = \mathsf{nb}(c,n) \\ i \notin I}} \bigwedge_{\substack{j \in I \\ A \in \{c,n\}}} (A_j \wedge JS_G A_j)$$

The conditional effects of Reason are then:

1. For all $G \subseteq Agt$, $c \in \mathbf{Colors}$ and $i \leqslant 50$:
   - Condition:
   $$\bigvee_{N \subsetneq [\![1,5]\!]} (\bigwedge_{n \in N} KAll(G, c, n, i) \wedge \bigwedge_{n \in [\![1,5]\!] \setminus N} (\neg n_i \wedge JS_G n_i))$$
   - Positive effect: $JS_G c_i$

2. For all $G \subseteq Agt$, $n \leqslant 5$ and $i \leqslant 50$:
   - Condition:
   $$\bigvee_{C \subsetneq \mathbf{Colors}} (\bigwedge_{c \in C} KAll(G, c, n, i) \wedge \bigwedge_{\substack{c \in \\ \mathbf{Colors} \setminus C}} (\neg c_i \wedge JS_G c_i))$$
   - Positive effect: $JS_G n_i$

3. For all $i \leqslant 50$ and $c \in \mathbf{Colors}$:
   - Condition: $\bigwedge_{\gamma \in \mathbf{Colors} \setminus c} (\neg \gamma_i \wedge JS_G \gamma_i)$
   - Positive effect: $JS_G c_i$

4. For all $i \leqslant 50$ and $n \in [\![1,5]\!]$:
   - Condition: $\bigwedge_{\nu \in [\![1,5]\!] \setminus \{n\}} (\neg \nu_i \wedge JS_G \nu_i)$
   - Positive effect: $JS_G n_i$

5. For any $A \in \mathbf{A}$ and any $I \subseteq [\![1, 50]\!]$ of indexes such that $|I| = \mathsf{nb}(A)$:
   - Condition: $\bigwedge_{i \in I} (A_i \wedge JS_G A_i)$
   - Positive effects: $JS_G A_i$ for all $i \notin I$

6. Reasoning test:
   - Condition:
   $$\bigwedge_{ce \in reff} \neg \big(cnd(ce) \wedge \bigvee_{\alpha \in ceff^+(ce)} \neg \alpha \big)$$
   where $reff$ denotes the set of all conditional effects described in items 1-5
   - Negative effect: $Reason$

We call **HAct** the set of all actions described in this section. Let us remark that all actions of **HAct** are indeed consistent and knowledge-positive. In the following section we will explore their properties and see that things can actually be done in a much simpler framework.

We do not handle here the last round of the game when the stack becomes empty. This is easily doable by adding some extra fluents signalling this last round and adapting some of the conditional effects of actions. However we choose to leave it aside for simplicity as it has no effect on the evolution of knowledge, which is the focus of this paper.

From Infinite to Finite Semantics

We define $H\text{-}STATES$ as the set of states reachable from $INIT$ via the actions of **HAct**, including $INIT$ itself.

Proposition 3. *Let $s \in H\text{-}STATES$. For all $\alpha \in s$ of modal depth at least 1, one of the following hold:*

1. $\alpha \in \mathsf{I\text{-}ATM}$;
2. α *is of the form* $\sigma JS_G A_i$ *where* $\sigma \in \mathsf{Seq}_{Agt}$ *and both* $JS JS_G A_i \in s$ *and* $JS_G A_i \in s$;
3. *There exist* $G \subseteq Agt$ *and* $ce \in eff(\mathsf{Reason})$ *such that* $s \models cnd(ce)$, $ceff^+(ce) = \{JS_G A_i\}$, α *is of the form* σA_i *for some* $\sigma \in \mathsf{Seq}_G$, *and* $\sigma' A_i \in s$ *for all* $\sigma' \in \mathsf{Seq}_G$.

It follows that atoms of modal depth strictly larger than 2 do not carry any significant information in Hanabi.

Proposition 4. *Let* $\alpha = JS_{G_1} \ldots JS_{G_n} JS_G A_i$ *be such that* $\alpha \notin \mathsf{I\text{-}ATM}$ *and let* $s \in H\text{-}STATES$. *Then* $\alpha \in s$ *iff* $JS_{G_1 \cup \cdots \cup G_n} JS_G Ai \in s$.

Proposition 4 entails that Hanabi states can be encoded in a compact way, only considering atoms of modal depth at most 2. Any formula φ can be evaluated in these states by first translating it to a formula $tr(\varphi)$ that is equivalent to φ in $H\text{-}STATES$, defined as follows:

$$tr(p) = p$$
$$tr(JS_G p) = JS_G p$$
$$tr(JS_{G_1}...JS_{G_n} JS_G p) = \begin{cases} \top \text{ if } JS_{G_1}...JS_{G_n} JS_G p \in \mathsf{I\text{-}ATM} \\ JS_{G_1 \cup \cdots \cup G_n} JS_G p \text{ otherwise} \end{cases}$$
$$tr(\neg \varphi) = \neg tr(\varphi)$$
$$tr(\varphi \wedge \psi) = tr(\varphi) \wedge tr(\psi)$$

We know by Proposition 4 that for any $\varphi \in \mathcal{L}_{\mathsf{EL\text{-}O}}$ and any $s \in H\text{-}STATES$, $s \models \varphi$ iff $s \models tr(\varphi)$. As atoms in $tr(\varphi)$ are all of depth at most 2, it follows that:

Proposition 5. *For any* $s \in H\text{-}STATES$ *and* $\varphi \in \mathcal{L}_{\mathsf{EL\text{-}O}}$, $s \models \varphi$ *iff* $s \models tr(\varphi)$ *iff* $s \cap ATM^{\leqslant 2} \models^{\mathsf{CPC}} tr(\varphi)$.

The point here is obviously not to fully compute states of H-$STATES$ so that we can then take their restriction to $ATM^{\leqslant 2}$, but to work with subsets of $ATM^{\leqslant 2}$ from the start. For this we at least need a new semantics for actions that restricts changes to modal depth at most 2. We actually show that things can be done in a more compact way. Let us first observe a few more properties of Hanabi states.

Proposition 6. *All states $s \in H$-$STATES$ satisfy the following properties:*

H1. If $JS_G JS_H p \in s \setminus$ I-ATM then $JS_H p \in s$;
H2. If $JS_G JS_H p \in s$ then $JS_{G \cup H} JS_H p \in s$;
H3. If $JS_G JS_H \in s \setminus$ I-ATM then $JS_{G'} JS_{H'} \in s$ for all $G' \subseteq G$ and $H' \subseteq H$;
H4. If $JS_G p \in s$ then $JS_{G'} p \in s$ for all $G' \subseteq G$;
H5. If $JS_G p \in s$ then $JS_{G'} JS_{G''} p \in s$ for all $G', G'' \subseteq G$.

The properties stated above are instrumental in going from a compact representation of Hanabi states to their original counterparts. Let us first show how to go from a set of atoms satisfying these properties to an EL-O state. We define the following expansion function: for any $s \in ATM^{\leqslant 2}$,

$$exp_1(s) = s \cup \text{I-ATM} \cup \{\sigma\alpha \mid \exists JS_G \alpha \in s, \sigma \in \mathsf{Seq}_G\}.$$

Proposition 7. *Let $s \subseteq ATM^{\leqslant 2}$ be a set of atoms satisfying properties H1–H5. Then $exp_1(s) \in STATES$ and for all $\varphi \in \mathcal{L}_{\text{EL-O}}$, $s \models^{\text{CPC}} tr(\varphi)$ iff $exp_1(s) \models \varphi$.*

Second, we show how to get a set satisfying the above properties from any arbitrary subset of $ATM^{\leqslant 2}$ with a second expansion function. For any $s \subseteq ATM^{\leqslant 2}$, we define:

$$exp_2(s) = s \cup \{JS_G p \mid \exists G' \supseteq G, JS_{G'} p \in s\}$$
$$\cup \{JS_H p \mid \exists H' \supseteq H, \exists G' \subseteq Agt, JS_{G'} JS_{H'} p \in s\}$$
$$\cup \{JS_G JS_H p \mid \exists G' \supseteq G \cup H, JS_{G'} p \in s\}$$
$$\cup \{JS_G JS_H p \mid \exists H' \supseteq H, \exists G' \subseteq Agt, G \subseteq G' \cup H'$$
$$\text{and } JS_{G'} JS_{H'} p \in s\}.$$

Proposition 8. *For any $s \subseteq ATM^{\leqslant 2}$, $exp_2(s)$ satisfies properties H1–H5.*

We can now define truth of $\mathcal{L}_{\text{EL-O}}$ formulas in any set $s \subseteq ATM^{\leqslant 2}$ in the context of Hanabi:

$$s \models^{\text{H}} \varphi \text{ if } exp_2(s) \models^{\text{CPC}} tr(\varphi).$$

We call H-EL-O the corresponding logic. It follows from Propositions 7 and 8 that for all $s \subseteq ATM^{\leqslant 2}$, $s \models^{\text{H}} \varphi$ iff $exp_1(exp_2(s)) \models \varphi$. Note that it is a standard property of classical propositional calculus that $exp_2(s) \models^{\text{CPC}} tr(\varphi)$ iff $exp_2(s) \cap ATM(tr(\varphi)) \models^{\text{CPC}} tr(\varphi)$, where $ATM(tr(\varphi))$ denotes the set of all atoms appearing in $tr(\varphi)$; hence we do not need to compute $exp_2(s)$ in its entirety in order to evaluate truth value of a formula φ.

Finally, we give a new set of initial Hanabi states and a new semantics for actions, and show that the original semantics are correctly simulated. The set C-$INIT$ is defined as the set of sets of atoms of the form

$\mathsf{CardAtb} \cup \{Belongs_{i,a} \mid a \in Agt, i \in \mathsf{Init}(a)\} \cup$

$\{Next_{5np+1}, Turn_1 \, ErrToken_0, HintToken_8, Reason\} \cup$

$\{JS \, JS_{Agt \setminus \{a\}} A_i \mid a \in Agt, A \in \mathbf{A}, i \in \mathsf{Init}(a)\}.$

Proposition 9. $INIT = \{exp_1(exp_2(s)) \mid s \in C$-$INIT\}.$

Given an action a, the function τ_{a}^H is a partial function over subsets of $ATM^{\leqslant 2}$. If $s \subseteq ATM^{\leqslant 2}$, $\tau_{\mathsf{a}}^H(s)$ is defined if $s \models^{\text{H}} pre(\mathsf{a})$, and in that case

$$\tau_{\mathsf{a}}^H(s) = \left(s \setminus \bigcup_{\substack{ce \in eff(\mathsf{a}) \\ s \models^{\text{H}} cnd(ce)}} ceff^-(ce) \right) \cup \bigcup_{\substack{ce \in eff(\mathsf{a}) \\ s \models^{\text{H}} cnd(ce)}} ceff^+(ce).$$

Proposition 10. *For any $s \subseteq ATM^{\leqslant 2}$ and $\mathsf{a} \in \mathbf{HAct}$, $\tau_{\mathsf{a}}(exp_1(exp_2(s)))$ is defined iff $\tau_{\mathsf{a}}^H(s)$ is defined, and in that case if $exp_1(exp_2(s)) \in H$-$STATES$ then $\tau_{\mathsf{a}}(exp_1(exp_2(s))) = exp_1(exp_2(\tau_{\mathsf{a}}^H s))$.*

It follows that starting from C-$INIT$ and using τ_{a}^H for action semantics and $\models^{\text{H}}$ for static semantics is enough to completely model the evolution of knowledge in a game of Hanabi.

Let us finish with a note on complexity in H-EL-O. First, the size of $tr(\varphi)$ is linear in that of φ. Second, as we have already noted, $exp_2(s) \models^{\text{CPC}} tr(\varphi)$ iff $exp_2(s) \cap ATM(tr(\varphi)) \models^{\text{CPC}} tr(\varphi)$. Third, classical propositional calculus corresponds to the fragment of H-EL-O with no observability operators. It follows that:

Proposition 11. *The H-EL-O satisfiability problem is* NP-*complete.*

Conclusion

In this paper we have extended the epistemic logic EL-O with infinite semantics accounting for common knowledge within arbitrary groups of agents. We have then proposed a formalization of Hanabi in this generalized setting, which relies on a special reasoning action to make up for the limited reasoning capabilities of agents in EL-O. Finally, we have shown that Hanabi can actually be captured in a much simpler, finite fragment of the full framework. We thus obtain a formalization of the evolution of knowledge in Hanabi that is both lightweight and complete. In particular we have shown formal grounds for restriction to modal depth 2, a heuristic featured in the literature on Hanabi.

There are many directions in which this work remains to be extended. On the theoretical side, we must still formally show that the general semantics proposed in this paper correspond to a fragment of the standard epistemic logic S5, and that the proposed action descriptions correctly simulate evolution of knowledge as has been represented with DEL in the literature. Moreover, while modeling knowledge in Hanabi gives us a solid background from which to start, we have yet said nothing about how players choose which action is best. This in particular requires agents to reason about the outcomes of actions, and cannot be done with the centralized planning approach that has been previously used with EL-O. The dynamic extension of EL-O proposed in (Herzig, Maris, and Perrotin 2021) could be useful for this. Finally, it remains to explore practical applicability of this method of keeping track of knowledge in actual playing systems.

Acknowledgments

References

Aru, J.; Labash, A.; Corcoll, O.; and Vicente, R. 2023. Mind the gap: Challenges of deep learning approaches to theory of mind. *Artificial Intelligence Review*, 56(9): 9141–9156.

Bard, N.; Foerster, J. N.; Chandar, S.; Burch, N.; Lanctot, M.; Song, H. F.; Parisotto, E.; Dumoulin, V.; Moitra, S.; Hughes, E.; et al. 2020. The Hanabi challenge: A new frontier for AI research. *Artificial Intelligence*, 280: 103216.

Charrier, T.; Schwarzentruber, F.; Bezhanishvili, G.; D'Agostino, G.; Metcalfe, G.; and Studer, T. 2018. Complexity of Dynamic Epistemic Logic with Common Knowledge. *Advances in Modal Logic*, 12: 27–31.

Chetcuti-Sperandio, N.; Goudyme, A.; Lagrue, S.; and de Lima, T. 2020. First Steps for Determining Agent Intention in Dynamic Epistemic Logic. In *12th International Conference on Agents and Artificial Intelligence (ICAART 2020)*, 725–733.

Gamblin, S.; Niveau, A.; and Bouzid, M. 2022. A symbolic representation for probabilistic dynamic epistemic logic. In *21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022)*.

Herzig, A.; Maris, F.; and Perrotin, E. 2021. A Dynamic Epistemic Logic with Finite Iteration and Parallel Composition. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*, volume 18, 676–680.

Perrotin, E. 2021. *Lightweight approaches to reasoning about knowledge and belief*. Ph.D. thesis, Université Paul Sabatier-Toulouse III.

Van Ditmarsch, H.; van Der Hoek, W.; and Kooi, B. 2007. *Dynamic epistemic logic*, volume 337. Springer Science & Business Media.

Walton-Rivers, J.; Williams, P. R.; and Bartle, R. 2019. The 2018 Hanabi competition. In *2019 IEEE Conference on Games (CoG)*, 1–8. IEEE.