Interpretability of games

A *game* (with two-players **A** and **B**, turn-based, with perfect information, without draws) is a rooted tree $\mathcal{G}$ (usually infinite). We write $\mathcal{G}^A$ for the set of nodes of even depth, including the root (it is **A**'s turn to play), and $\mathcal{G}^B$ for the set of nodes of odd depth (it is **B**'s turn to play). We moreover require that every leaf belong to $\mathcal{G}^A$ (otherwise, add an irrelevant single child to the corresponding node of $\mathcal{G}^B$), and we exclude the "trivial game" with a single node (i.e., the game where **A** immediately loses before even playing). A *play* is a branch of $\mathcal{G}$, either ending with a leaf (in which case **B** is declared to be the winner) or infinite (in which case **A** is declared to be the winner).[1] If x is a node of a tree, we write $\mathrm{Ch}(x)$ for the set of (immediate) children of x.

1. Interpretations

The idea of an interpretation[2] (turn-for-turn[3], from the perspective of **A**) is for **A** to translate states of a game into those of another game, so that they can pretend that they are playing the other game (but still perhaps win the original game!). **B** does not have to cooperate, so the translation must deal with whatever moves **B** decides to play.

Definition 1.1. A subtree $\mathcal{G}_* \subseteq \mathcal{G}$ is a *subgame of $\mathcal{G}$ obtained by restricting only the allowed moves of **A*** if it is a subtree with the same root as $\mathcal{G}$, whose leaves are exactly the leaves of $\mathcal{G}$ belonging to $\mathcal{G}_*$ (so that **A** is losing in $\mathcal{G}_*$ only if they are also losing in $\mathcal{G}$), and any $x \in \mathcal{G}^B$ has the same children in $\mathcal{G}$ and in $\mathcal{G}_*$.

Definition 1.2. Consider two games $\mathcal{G}$ and $\mathcal{H}$. We define an *interpretation* of $\mathcal{H}$ in $\mathcal{G}$ as a tuple $(\mathcal{G}_*, f, f^*)$ where $\mathcal{G}_*$ is a subgame[4] of $\mathcal{G}$ obtained by restricting only the allowed moves of **A**, the *translation map* $f : \mathcal{G}_* \to \mathcal{H}$ is a map from the nodes of $\mathcal{G}_*$ to those of $\mathcal{H}$, and for each $x \in \mathcal{G}_*^A$, the *reverse translation map* f_x^* is a map $\mathrm{Ch}(f(x)) \to \mathrm{Ch}(x) \cap \mathcal{G}_*$ (each legal move of **A** in the interpreted game is reverse translated into a legal move in the original game), such that:

- f maps the root to the root, maps $\mathcal{G}_*^A$ to $\mathcal{H}^A$ and $\mathcal{G}_*^B$ to $\mathcal{H}^B$, and maps leaves to leaves (a loss is translated into a loss)
- for any $x \in \mathcal{G}_*^A$, the map $f \circ f_x^*$ is the identity of $\mathrm{Ch}(f(x))$
- for any $x \in \mathcal{G}_*^B$, we have $f(\mathrm{Ch}(x)) \subseteq \mathrm{Ch}(f(x))$ (each legal move of **B** in the original game is translated into a legal move in the interpreted game)

Example 1.3. Each game $\mathcal{G}$ interprets itself trivially via the *identity interpretation* $\left(\mathcal{G}, \mathrm{id}, \left(\mathrm{id}_{\mathrm{Ch}(x)}\right)_{x \in \mathcal{G}}\right)$. More generally, any isomorphism of games induces an interpretation.

[1]The goal for **A** is thus to ensure infinite play. For instance, any finite game without draws can be transformed into such a game by giving "useless" legal moves to each player once **A** has won.

[2]Perhaps words like *simulation*, *emulation*, or *reduction* make more sense, but my starting point was an analogy with interpretability of first order theories. This analogy works as follows: if one sees proving a given statement as some sort of one-player game, so that strategies correspond to proofs, then the fact that a theory is interpretable in another means that it suffices to prove a statement ("play the game") in the theory which is interpretable (e.g., establishing an arithmetic statement in ZFC by instead proving it in PA).

[3]Instead, one could play the interpreted game via *sequences* of moves in the interpreting game. Of course, one runs into the issue that we need to account for **B**'s reactions, hence it is more like a "short-term strategy" than an actual sequence of moves. Moreover, one must ensure that the final result of this short-term strategy translates into a single state of the interpreted game independently of **B**'s play.

[4]The reason for not taking all of $\mathcal{G}$ is that we do not require that we have a translation of the states which we do not intend to reach. (For instance, maybe it is not always possible to translate **B**'s moves, but **A** keeps playing a subgame where this is possible.)

Example 1.4. If $\mathcal{G}_*$ is a subgame of $\mathcal{G}$ obtained by restricting only the allowed moves of **A**, then $\mathcal{G}_*$ is interpreted in $\mathcal{G}$ via $\left(\mathcal{G}_*, \mathrm{id}, \left(\mathrm{id}_{\mathrm{Ch}(x)}\right)_{x \in \mathcal{G}_*^A}\right)$.

Example 1.5. If $\mathcal{H}$ is a game obtained from $\mathcal{G}$ by extending only the allowed moves of **B** (i.e., $\mathcal{G} \subseteq \mathcal{H}$ and any $x \in \mathcal{G}^A$ has the same children in $\mathcal{G}$ and in $\mathcal{H}$) and without changing the leaves, then $\mathcal{H}$ is interpreted in $\mathcal{G}$ via $(\mathcal{G}, f, f^*)$, where f is the inclusion $\mathcal{G} \to \mathcal{H}$, and $f_x^* : \mathrm{Ch}(f(x)) \to \mathrm{Ch}(x)$ is the identity map for each $x \in \mathcal{G}^A$ (we have $\mathrm{Ch}(f(x)) = \mathrm{Ch}(x)$ by hypothesis).

Example 1.4 and Example 1.5 have an intuitive explanation: if they want to do so, **A** can be play "pessimistically", assuming that they have less allowed moves than they actually do, and assuming that **B** has more moves than they actually do. Indeed, if **A** finds a way to win even under these pessimistic assumptions (which both work against them), then they have in particular found a way to win in the real game. This principle is formalized in what follows.

2. Strategies and interpretations

Let $\mathcal{G}$ be a game. A *strategy* of $\mathcal{G}$ (for **A**) is a partial map $\sigma : \mathcal{G}^A \to \mathcal{G}^B$ such that $\sigma(x) \in \mathrm{Ch}(x)$ whenever it is defined. We say that σ is a *winning strategy* if the value of $\sigma(x_n)$ is defined (in particular, x_n is not a leaf) for any finite sequence $x_1, x_2, ..., x_n \in \mathcal{G}^A$ where x_1 is the root of $\mathcal{G}$ and $x_{i+1} \in \mathrm{Ch}(\sigma(x_i))$ for all $1 \leq i < n$.

Let $(\mathcal{G}_*, f, f^*)$ be an interpretation of a game $\mathcal{H}$ in a game $\mathcal{G}$, and let $\sigma : \mathcal{H}^A \to \mathcal{H}^B$ be a strategy of $\mathcal{H}$. For any $x \in \mathcal{G}^A$ such that $\sigma(f(x))$ is defined, we define $(f^*\sigma)(x) := f_x^*(\sigma(f(x))) \in \mathrm{Ch}(x) \cap \mathcal{G}_*$. This defines a strategy $f^*\sigma$ of $\mathcal{G}$ (a partial map $\mathcal{G}^A \to \mathcal{G}^B$), which we call the *pullback of* σ *by the interpretation.*

Proposition 2.1. The pullback of a winning strategy σ by an interpretation $(\mathcal{G}_*, f, f^*)$ is a winning strategy. In particular, if $\mathcal{H}$ is interpreted in $\mathcal{G}$ and admits a winning strategy, then so does $\mathcal{G}$.

Proof: By definition of an interpretation, we have $f((f^*\sigma)(x)) = \sigma(f(x))$ for any $x \in \mathcal{G}^A$ such that $\sigma(f(x))$ is defined. Consider a finite sequence $x_1, x_2, ..., x_n \in \mathcal{G}^A$, where x_1 is the root of $\mathcal{G}$ and $x_{i+1} \in \mathrm{Ch}((f^*\sigma)(x_i))$ for all $1 \leq i < n$. Then, $f(x_1), f(x_2), ..., f(x_n) \in \mathcal{H}^A$ is a finite sequence for $\mathcal{H}$ where $f(x_1)$ is the root of $\mathcal{H}$ and $f(x_{i+1}) \in \mathrm{Ch}(\sigma(f(x_i)))$ by definition of an interpretation and of $f^*\sigma$. Since σ is winning, $\sigma(f(x_n))$ is defined, and thus $(f^*\sigma)(x_n)$ is also defined, so $f^*\sigma$ is a winning strategy. □

3. The category of games and interpretations

We can compose interpretations: if $(\mathcal{G}_*, f, f^*)$ is an interpretation of $\mathcal{H}$ in $\mathcal{G}$ and $(\mathcal{H}_*, g, g^*)$ is an interpretation of $\mathcal{I}$ in $\mathcal{H}$, then $\left(f^{-1}(\mathcal{H}_*), g \circ f, (f_x^* \circ g_x^*)_{x \in \mathcal{G}^B \cap f^{-1}(\mathcal{H}_*)}\right)$ is an interpretation of $\mathcal{I}$ in $\mathcal{G}$. Hence, there is a category Interp of games, where a morphism $\mathcal{G} \to \mathcal{H}$ is an interpretation of $\mathcal{H}$ in $\mathcal{G}$, and the identity morphisms are given by the identity interpretations.

Proposition 3.1. Let $\mathcal{G}$ and $\mathcal{H}$ be two games. Assume that they are isomorphic in Interp, i.e., that there are two interpretations $(\mathcal{G}, f, f^*) : \mathcal{G} \to \mathcal{H}$ and $(\mathcal{H}, g, g^*) : \mathcal{H} \to \mathcal{G}$ whose compositions (in both directions) are the respective identity interpretations. Then, $\mathcal{G}$ and $\mathcal{H}$ are isomorphic as games.

Proof: First, we must have $f^{-1}(\mathcal{H}_*) = \mathcal{G}$ and $g^{-1}(\mathcal{G}_*) = \mathcal{H}$, which implies $\mathcal{G} = \mathcal{G}_*$ and $\mathcal{H} = \mathcal{H}_*$. Since $f \circ g = g \circ f = \mathrm{id}$, the maps f and g are inverse bijections between the nodes of $\mathcal{G}$ and those of $\mathcal{H}$. It suffices to show that f and g are morphisms of trees, i.e., that $f(y)$ is a child of $f(x)$

whenever y is a child of x, and similarly for g. As the cases of f and g are symmetric, we focus on f. If $x \in \mathcal{G}^B$, then $f(\mathrm{Ch}(x)) \subseteq \mathrm{Ch}(f(x))$ by definition of interpretations. We now assume that $x \in \mathcal{G}^A$. Let $x' = f(x) \in \mathcal{H}^A$, so that $x = g(x')$, and then by definition of an interpretation we have $g(g^*_{x'}(y)) = y$ for any $y \in \mathrm{Ch}(x)$, meaning that $f(y) = g^*_{x'}(y)$ belongs to $\mathrm{Ch}(x') = \mathrm{Ch}(f(x))$. □

For example, Proposition 2.1 implies that the map that takes a game to the set of its winning strategies defines a contravariant functor from Interp to Set, i.e., a presheaf on Interp.

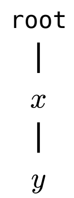
4. The interpretability preorder

If there is an interpretation of $\mathcal{H}$ in $\mathcal{G}$, we say that $\mathcal{H}$ is *interpretable* in $\mathcal{G}$, and we write $\mathcal{H} \preceq \mathcal{G}$: this defines a partial preorder on games.

Example 4.1. Consider any game $\mathcal{G}$ with no leaves (i.e., a game where **A** always wins). In particular, there exists an infinite branch $\mathcal{H} \subseteq \mathcal{G}$. Consider the map $f : \mathcal{G} \to \mathcal{H}$ taking any node to the unique node of $\mathcal{H}$ with the same depth. If $x \in \mathcal{G}$ has depth i, then $f(\mathrm{Ch}(x))$ and $\mathrm{Ch}(f(x))$ both consist of the unique element y of $\mathcal{H}$ of depth $i + 1$. (In particular, a reverse translation map f^*_x is given by any choice of a child of x, whose image by f will automatically coincide with y.) Hence $\mathcal{H} \preceq \mathcal{G}$.

4.1. Minimal games

We say that a game is *minimal* (for $\preceq$) if $\mathcal{G}$ is interpretable in any game interpretable in $\mathcal{G}$. In what follows, we denote by $\mathcal{L}$ the "losing game" where **A** and **B** each play a forced move, then **A** loses:



Proposition 4.2. $\mathcal{L}$ is interpretable in any game $\mathcal{G}$.

Proof: Define a map $\mathcal{G} \to \mathcal{L}$ as follows: the root is mapped to the root, all nodes in $\mathcal{G}^B$ are mapped to x, and all nodes in $\mathcal{G}^A$ besides the root are mapped to y.

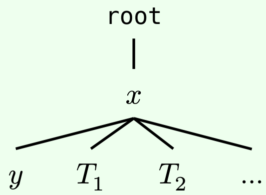
Let $v \in \mathcal{G}^B$. Then, $f(\mathrm{Ch}(v)) \subseteq f(\mathcal{G}^A \setminus \texttt{root}) = \{y\}$, and $\mathrm{Ch}(f(v)) = \mathrm{Ch}(x) = \{y\}$, so $f(\mathrm{Ch}(v)) \subseteq \mathrm{Ch}(f(v))$.

Now, let v be the root of $\mathcal{G}$. Then, we can pick any reverse translation map $f^*_{\texttt{root}}$ mapping x to any child of the root of $\mathcal{G}$, and then $f \circ f^*_{\texttt{root}} = \mathrm{id}$ is automatically true.

Finally, if $v \in \mathcal{G}^A \setminus \texttt{root}$, then $\mathrm{Ch}(f(x)) = \emptyset$, so the corresponding translation map is the trivial map and $f \circ f^*_v = \mathrm{id}$ is vacuously true. □

As a consequence, a game is minimal if and only if it is interpretable in $\mathcal{L}$.

Proposition 4.3. The games which are interpretable in $\mathcal{L}$ (and, hence, the minimal games) are exactly the games of the following form (the first move of **A** is forced, and then **B** has the possibility to win in one):



where $\{T_1, T_2, ...\}$ is a set of games (possibly empty). Equivalently, these are the games obtained from $\mathcal{L}$ by extending only the allowed moves of **B**.

Proof: The conditions that an interpretation $(\mathcal{L}_*, f, f^*)$ of $\mathcal{G}$ in $\mathcal{L}$ must satisfy are:

- $\mathcal{L}_* = \mathcal{L}$ (there are no proper subgames of $\mathcal{L}$ obtained by restricting the allowed moves of **A**)
- f maps the root of $\mathcal{L}$ to the root of $\mathcal{G}$, x to some $f(x) \in \mathcal{G}^B$, and y to some leaf $f(y) \in \mathcal{H}^A$.
- the reverse translation map $f^*_{\texttt{root}}$ is constant, equal to x, so the root of $\mathcal{G}$ must have $f(x)$ as its single child.
- we must have $f(y) \in \mathrm{Ch}(f(x))$, so the leaf $f(y)$ is a child of $f(x)$. □

This intuitively makes sense: a minimal game is a game where **A** is "as pessimistic as possible", which indeed corresponds to there being an immediate way for **B** to win. Similarly, if we classify games which are minimal among the games for which **B** does not have a winning strategy, these would certainly be games for which any single mistake of **A** leads to **B** winning in one move.

4.2. Classification of maximal games

A game $\mathcal{G}$ is *maximal* (for $\preceq$) if any game in which $\mathcal{G}$ is interpretable is itself interpretable in $\mathcal{G}$.

[**TODO:** Up to mutual interpretability, the only maximal game is the following game $\mathcal{W}$: **A** always has infinitely many moves, **B** always has a single move, and there are no leaves (**A** always wins). We shall in fact show that $\mathcal{W}$ interprets any game. Indeed: take a game $\mathcal{G}$, it is interpreted in a game where **B** has a single move by Example 1.5, so we can assume that this is the case for $\mathcal{G}$. Now, $\mathcal{G}$ can be embedded in $\mathcal{W}$, so we fix such an embedding. We let $\mathcal{G}'$ be obtained by replacing each leaf of $\mathcal{G}$ by a copy of $\mathcal{W}$. $\mathcal{G}'$ is obtained from $\mathcal{W}$ by restricting only the moves of **A**, so it suffices now to show that $\mathcal{G}'$ interprets $\mathcal{G}$. For this, define the map f extending the identity of $\mathcal{G}$ by mapping each remaining node of $\mathcal{G}'$ to either the leaf above it (if it is in ${\mathcal{G}'}^A$), or the parent of that leaf (if it is in ${\mathcal{G}'}^B$)

Intuitively: the most optimistic that **A** can be is to assume that they can play whatever and still win.]