Tarski’s Definition of Truth

In first-order logic with equality, we always have that \( \equiv^A = \{ \langle a, a \rangle \mid a \in |A| \} \) where \( |A| = U_A \) is the universe of \( A \). That is, the equality predicate symbol, “=”, must always be interpreted as true equality:

\[
A \models t_1 = t_2 \iff t_1^A = t_2^A.
\]

For \( A \in \text{STRUC}[\Sigma], t \in \text{term}(\Sigma), \varphi \in L(\Sigma) \) we give the following inductive definitions of \( t^A \) and \( A(\varphi) \):

**term base:** for \( x_i \in \text{var}, x_i^A \) is already given. (Each structure has a default value for each variable.)

**term inductive:** \((f_i(t_1, \ldots, t_{r_i}))^A = f_i^A(t_1^A, \ldots, t_{r_i}^A)\)

**truth base:** \( A(R_i(t_1, \ldots, t_{a_i})) = 1 \) if \( \langle t_1^A, \ldots, t_{a_i}^A \rangle \in R_i^A \) then 1 else 0

**truth inductive:**

1. \( A(\neg \alpha) = 1 - A(\alpha) \)
2. \( A(\alpha \lor \beta) = \max(A(\alpha), A(\beta)) \)
3. \( A(\exists x_i(\alpha)) = \max_{a \in |A|}(\langle A, x_i/a \rangle(\alpha)) \)

\((A, x_i/a)\) is the same structure as \( A \) with the single exception that \( x_i^{(A, x_i/a)} = a \), i.e., the default value of \( x_i \) in \((A, x_i/a)\) is \( a \in |A| \).
Game-Theoretic Definition of Truth

The truth of a first-order formula corresponds to a two-person game: $G(\mathcal{A}, \varphi)$ is the game on structure $\mathcal{A}$, formula $\varphi$. Assume that $\varphi$ is in **negation normal form**, i.e., the quantifiers are $\forall, \exists$, the propositional connectives are $\land, \lor, \neg$ and all $\neg$'s have been pushed inside as far as possible using the de Morgan laws, so the only occurrences of $\neg$'s are directly in front of atomic formulas. The truth game has two players named Dumbledore ($D$) and Gandalf ($G$). D is trying to prove that $\mathcal{A} \models \varphi$ and G is trying to prove that $\mathcal{A} \not\models \varphi$.

In $G(\mathcal{A}, \varphi)$,

**game base:** If $\varphi$ is atomic, then if $\mathcal{A} \models \varphi$ then $D$ wins, else $G$ wins

**game inductive:**

1. If $\varphi = \alpha \lor \beta$, then $D$ chooses one of the disjuncts: $\psi \in \{\alpha, \beta\}$ and the next position is $G(\mathcal{A}, \psi)$.
2. If $\varphi = \alpha \land \beta$, then $G$ chooses one of the conjuncts: $\psi \in \{\alpha, \beta\}$ and the next position is $G(\mathcal{A}, \psi)$.
3. If $\varphi = \exists x_i(\psi)$, then $D$ chooses an element $e \in |\mathcal{A}|$ and the next position is $G((\mathcal{A}, x_i/e), \psi)$.
4. If $\varphi = \forall x_i(\psi)$, then $G$ chooses an element $a \in |\mathcal{A}|$ and the next position is $G((\mathcal{A}, x_i/a), \psi)$.

**Theorem:** For any vocabulary $\Sigma$, formula $\varphi \in L(\Sigma)$ in negation normal form, and structure $\mathcal{A} \in \text{STRUC}[\Sigma]$, Tarski’s definition of truth, and the game theoretic definition of truth are equivalent, i.e,

$$\mathcal{A} \models \varphi \iff D \text{ has a winning strategy for } G(\mathcal{A}, \varphi) \text{ and},$$
$$\mathcal{A} \not\models \varphi \iff G \text{ has a winning strategy for } G(\mathcal{A}, \varphi).$$

**Proof:** This can be proved by induction on $\varphi$. It would be a good exercise for you to fill in the details. $\square$