

Predicate Logic: Undecidability

Raffi Khatchadourian Kevin Van Valkenburgh

October 23, 2008

Undecidability

Theorem (2.22)

The *decision problem* of validity in predicate logic is undecidable in the general case, i.e., no algorithm exists which, given any *arbitrary* ϕ written in predicate logic, decides whether or not $\models \phi$.

Intuition

Definition

Let the sequence $C \stackrel{\text{def}}{=} \langle (s_1, t_1), (s_2, t_2), \dots, (s_k, t_k) \rangle$ denote an instance of the correspondence problem where for $1 \leq i \leq k$, each s_i and t_i are binary strings.

Claim

$$\models \phi \iff C \text{ has a solution}$$

Intuition

Definition

Let the sequence $C \stackrel{\text{def}}{=} \langle (s_1, t_1), (s_2, t_2), \dots, (s_k, t_k) \rangle$ denote an instance of the correspondence problem where for $1 \leq i \leq k$, each s_i and t_i are binary strings.

Claim

$$\models \phi \iff C \text{ has a solution}$$

What does that mean?

$$\forall \mathcal{M}, \ell [\mathcal{M} \models_{\ell} \phi] \\ \iff \\ \exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$$

Note that ...

Reduction **must** be in finite time & space.

What does that mean?

$$\forall \mathcal{M}, \ell [\mathcal{M} \models_{\ell} \phi] \\ \iff \\ \exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$$

Note that ...

Reduction **must** be in finite time & space.

What does that mean?

$$\begin{aligned} & \forall \mathcal{M}, l [\mathcal{M} \models_l \phi] \\ & \iff \\ & \exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}] \end{aligned}$$

Note that ...

Reduction **must** be in finite time & space.

What does that mean?

$$\begin{aligned} & \forall \mathcal{M}, \ell [\mathcal{M} \models_{\ell} \phi] \\ & \iff \\ & \exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}] \end{aligned}$$

Note that ...

Reduction **must** be in finite time & space.

Function Symbols \mathcal{F}

Constant e : Think of e as the empty string.

Function f_0 : Think of f_0 as taking a string x and returning $x0$.

Function f_1 : Think of f_1 as taking a string x and returning $x1$.

Example

- Let $b = \langle b_1, b_2, \dots, b_r \rangle$ be a string (i.e., a sequence of characters) over $\{0, 1\}$. That is, each $b_j \in \{0, 1\}$ for $1 \leq j \leq r$.
- Then, $b = f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(e))) \dots)$.
- In general, for a string d , denote $f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(d))) \dots)$ by $f_{b_1 b_2 \dots b_r}(d)$.

Function Symbols \mathcal{F}

Constant e : Think of e as the empty string.

Function f_0 : Think of f_0 as taking a string x and returning $x0$.

Function f_1 : Think of f_1 as taking a string x and returning $x1$.

Example

- Let $b = \langle b_1, b_2, \dots, b_r \rangle$ be a string (i.e., a sequence of characters) over $\{0, 1\}$. That is, each $b_j \in \{0, 1\}$ for $1 \leq j \leq r$.
- Then, $b = f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(e))) \dots)$.
- In general, for a string d , denote $f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(d))) \dots)$ by $f_{b_1 b_2 \dots b_r}(d)$.

Function Symbols \mathcal{F}

Constant e : Think of e as the empty string.

Function f_0 : Think of f_0 as taking a string x and returning $x0$.

Function f_1 : Think of f_1 as taking a string x and returning $x1$.

Example

- Let $b = \langle b_1, b_2, \dots, b_r \rangle$ be a string (i.e., a sequence of characters) over $\{0, 1\}$. That is, each $b_j \in \{0, 1\}$ for $1 \leq j \leq r$.
- Then, $b = f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(e))) \dots)$.
- In general, for a string d , denote $f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(d))) \dots)$ by $f_{b_1 b_2 \dots b_r}(d)$.

Function Symbols \mathcal{F}

Constant e : Think of e as the empty string.

Function f_0 : Think of f_0 as taking a string x and returning $x0$.

Function f_1 : Think of f_1 as taking a string x and returning $x1$.

Example

- Let $b = \langle b_1, b_2, \dots, b_r \rangle$ be a string (i.e., a sequence of characters) over $\{0, 1\}$. That is, each $b_j \in \{0, 1\}$ for $1 \leq j \leq r$.
- Then, $b = f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(e))) \dots)$.
- In general, for a string d , denote $f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(d))) \dots)$ by $f_{b_1 b_2 \dots b_r}(d)$.

Function Symbols \mathcal{F}

Constant e : Think of e as the empty string.

Function f_0 : Think of f_0 as taking a string x and returning $x0$.

Function f_1 : Think of f_1 as taking a string x and returning $x1$.

Example

- Let $b = \langle b_1, b_2, \dots, b_r \rangle$ be a string (i.e., a sequence of characters) over $\{0, 1\}$. That is, each $b_j \in \{0, 1\}$ for $1 \leq j \leq r$.
- Then, $b = f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(e)))) \dots$.
- In general, for a string d , denote $f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(d)))) \dots$ by $f_{b_1 b_2 \dots b_r}(d)$.

Function Symbols \mathcal{F}

Constant e : Think of e as the empty string.

Function f_0 : Think of f_0 as taking a string x and returning $x0$.

Function f_1 : Think of f_1 as taking a string x and returning $x1$.

Example

- Let $b = \langle b_1, b_2, \dots, b_r \rangle$ be a string (i.e., a sequence of characters) over $\{0, 1\}$. That is, each $b_j \in \{0, 1\}$ for $1 \leq j \leq r$.
- Then, $b = f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(e)))) \dots$.
- In general, for a string d , denote $f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(d)))) \dots$ by $f_{b_1 b_2 \dots b_r}(d)$.

Function Symbols \mathcal{F}

Constant e : Think of e as the empty string.

Function f_0 : Think of f_0 as taking a string x and returning $x0$.

Function f_1 : Think of f_1 as taking a string x and returning $x1$.

Example

- Let $b = \langle b_1, b_2, \dots, b_r \rangle$ be a string (i.e., a sequence of characters) over $\{0, 1\}$. That is, each $b_j \in \{0, 1\}$ for $1 \leq j \leq r$.
- Then, $b = f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(e)))) \dots$.
- In general, for a string d , denote $f_{b_r}(f_{b_{r-1}} \dots (f_{b_2}(f_{b_1}(d)))) \dots$ by $f_{b_1 b_2 \dots b_r}(d)$.

Predicate Symbols \mathcal{P}

Predicate P : $P(x, y)$ iff

$$\exists \langle i_1, i_2, \dots, i_m \rangle [x = s_{i_1} s_{i_2} \dots s_{i_m} \wedge y = t_{i_1} t_{i_2} \dots t_{i_m}]$$

Example

i	1	2	3
s	1	10	011
t	101	00	11

- $P(1010011, 000011) \equiv \text{true?}$
- $P(1, 00) \equiv \text{true?}$

Predicate Symbols \mathcal{P}

Predicate P : $P(x, y)$ iff

$$\exists \langle i_1, i_2, \dots, i_m \rangle [x = s_{i_1} s_{i_2} \dots s_{i_m} \wedge y = t_{i_1} t_{i_2} \dots t_{i_m}]$$

Example

i	1	2	3
s	1	10	011
t	101	00	11

- $P(1010011, 000011) \equiv \text{true?}$
- $P(1, 00) \equiv \text{true?}$

Predicate Symbols \mathcal{P}

Predicate P : $P(x, y)$ iff

$$\exists \langle i_1, i_2, \dots, i_m \rangle [x = s_{i_1} s_{i_2} \dots s_{i_m} \wedge y = t_{i_1} t_{i_2} \dots t_{i_m}]$$

Example

i	1	2	3
s	1	10	011
t	101	00	11

- $P(1010011, 000011) \equiv \text{true?}$
- $P(1, 00) \equiv \text{true?}$

Structure

Recall ...

Want to construct a formula ϕ s.t. $\models \phi \implies C$ has a solution.
That is, $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$

Let:

- $\phi \equiv \phi_1 \wedge \phi_2 \implies \phi_3$

- $\phi_1 \equiv \bigwedge_{i=1}^n P(f_{s_i}(e), f_{t_i}(e))$

- $\phi_2 \equiv \forall v \forall w [P(v, w) \implies \bigwedge_{i=1}^n P(f_{s_i}(v), f_{t_i}(w))]$

- $\phi_3 \equiv \exists z [P(z, z)]$

Structure

Recall ...

Want to construct a formula ϕ s.t. $\models \phi \implies C$ has a solution.

That is, $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$

Let:

- $\phi \equiv \phi_1 \wedge \phi_2 \implies \phi_3$

- $\phi_1 \equiv \bigwedge_{i=1}^n P(f_{s_i}(e), f_{t_i}(e))$

- $\phi_2 \equiv \forall v \forall w [P(v, w) \implies \bigwedge_{i=1}^n P(f_{s_i}(v), f_{t_i}(w))]$

- $\phi_3 \equiv \exists z [P(z, z)]$

Structure

Recall ...

Want to construct a formula ϕ s.t. $\models \phi \implies C$ has a solution.

That is, $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$

Let:

- $\phi \equiv \phi_1 \wedge \phi_2 \implies \phi_3$

- $\phi_1 \equiv \bigwedge_{i=1}^n P(f_{s_i}(e), f_{t_i}(e))$

- $\phi_2 \equiv \forall v \forall w [P(v, w) \implies \bigwedge_{i=1}^n P(f_{s_i}(v), f_{t_i}(w))]$

- $\phi_3 \equiv \exists z [P(z, z)]$

Structure

Recall ...

Want to construct a formula ϕ s.t. $\models \phi \implies C$ has a solution.

That is, $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$

Let:

- $\phi \equiv \phi_1 \wedge \phi_2 \implies \phi_3$

- $\phi_1 \equiv \bigwedge_{i=1}^n P(f_{s_i}(e), f_{t_i}(e))$

- $\phi_2 \equiv \forall v \forall w [P(v, w) \implies \bigwedge_{i=1}^n P(f_{s_i}(v), f_{t_i}(w))]$

- $\phi_3 \equiv \exists z [P(z, z)]$

Selecting a Concrete Model

- Let $A = \{x \mid x \text{ is a finite string over } \{0, 1\}\}$.
 - A is the set of all finite binary strings.
 - Includes the **empty** string ϵ .

Selecting a Concrete Model

- Let $A = \{x \mid x \text{ is a finite string over } \{0, 1\}\}$.
 - A is the set of all finite binary strings.
 - Includes the **empty** string ϵ .

Selecting a Concrete Model

- Let $A = \{x \mid x \text{ is a finite string over } \{0, 1\}\}$.
 - A is the set of all finite binary strings.
 - Includes the **empty** string ϵ .

Interpreting Functions and Predicates

$$e^{\mathcal{M}} \stackrel{\text{def}}{=} \epsilon$$

$$f_0^{\mathcal{M}}(x) \stackrel{\text{def}}{=} x0$$

$$f_1^{\mathcal{M}}(x) \stackrel{\text{def}}{=} x1$$

$$P^{\mathcal{M}} \stackrel{\text{def}}{=} \{(x, y) \mid \exists \langle i_1, i_2, \dots, i_m \rangle [x = s_{i_1} s_{i_2} \dots s_{i_m} \wedge y = t_{i_1} t_{i_2} \dots t_{i_m}]\}$$

Interpreting Functions and Predicates

$$e^{\mathcal{M}} \stackrel{\text{def}}{=} \epsilon$$

$$f_0^{\mathcal{M}}(x) \stackrel{\text{def}}{=} x0$$

$$f_1^{\mathcal{M}}(x) \stackrel{\text{def}}{=} x1$$

$$P^{\mathcal{M}} \stackrel{\text{def}}{=} \{(x, y) \mid \exists \langle i_1, i_2, \dots, i_m \rangle [x = s_{i_1} s_{i_2} \dots s_{i_m} \wedge y = t_{i_1} t_{i_2} \dots t_{i_m}]\}$$

Interpreting Functions and Predicates

$$e^{\mathcal{M}} \stackrel{\text{def}}{=} \epsilon$$

$$f_0^{\mathcal{M}}(x) \stackrel{\text{def}}{=} x0$$

$$f_1^{\mathcal{M}}(x) \stackrel{\text{def}}{=} x1$$

$$P^{\mathcal{M}} \stackrel{\text{def}}{=} \{(x, y) \mid \exists \langle i_1, i_2, \dots, i_m \rangle [x = s_{i_1} s_{i_2} \dots s_{i_m} \wedge y = t_{i_1} t_{i_2} \dots t_{i_m}]\}$$

Interpreting Functions and Predicates

$$e^{\mathcal{M}} \stackrel{\text{def}}{=} \epsilon$$

$$f_0^{\mathcal{M}}(x) \stackrel{\text{def}}{=} x0$$

$$f_1^{\mathcal{M}}(x) \stackrel{\text{def}}{=} x1$$

$$P^{\mathcal{M}} \stackrel{\text{def}}{=} \{(x, y) \mid \exists \langle i_1, i_2, \dots, i_m \rangle [x = s_{i_1} s_{i_2} \dots s_{i_m} \wedge y = t_{i_1} t_{i_2} \dots t_{i_m}]\}$$



Assume $\forall \mathcal{M}, \ell[\mathcal{M} \models_{\ell} \phi]$. We show that

$\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$.

Recall ...

- $\phi \equiv \phi_1 \wedge \phi_2 \implies \phi_3$

- $\phi_2 \equiv \forall v \forall w [P(v, w) \implies \bigwedge_{i=1}^n P(f_{s_i}(v), f_{t_i}(w))]$

Claim: $\mathcal{M} \models \phi_2$

- Suppose $P(v, w)$.

- Implies that

$$\exists \langle i_1, i_2, \dots, i_m \rangle [v = s_{i_1} s_{i_2} \dots s_{i_m} \wedge w = t_{i_1} t_{i_2} \dots t_{i_m}].$$

- Show for each $j = 1 \dots n$, $P(s_{i_1} s_{i_2} \dots s_{i_m} s_j, t_{i_1} t_{i_2} \dots t_{i_m} t_j)$.

- $\langle i_1, i_2, \dots, i_m, j \rangle$



Assume $\forall \mathcal{M}, \ell[\mathcal{M} \models_{\ell} \phi]$. We show that
 $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$.

Recall ...

- $\phi \equiv \phi_1 \wedge \phi_2 \implies \phi_3$
- $\phi_1 \equiv \bigwedge_{i=1}^n P(f_{s_i}(e), f_{t_i}(e))$

Claim: $\mathcal{M} \models \phi_1$

Why?



Assume $\forall \mathcal{M}, \ell[\mathcal{M} \models_{\ell} \phi]$. We show that
 $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$.

Recall ...

- $\phi \equiv \phi_1 \wedge \phi_2 \implies \phi_3$
- $\phi_3 \equiv \exists z [P(z, z)]$

Claim: $\mathcal{M} \models \phi_3$

- Follows from $\mathcal{M} \models \phi_1 \wedge \mathcal{M} \models \phi_2$.
- Implies that there exists a solution to C .



Assume $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$. We show that $\forall \mathcal{M}, \ell [\mathcal{M} \models_{\ell} \phi]$.

- Recall that, by definition, \mathcal{M} so that we can check ϕ .
- Let \mathcal{M}' be such a model that includes
 - any constant $e^{\mathcal{M}'}$
 - two unary functions $f_0^{\mathcal{M}'}, f_1^{\mathcal{M}'}$
 - a binary predicate $P^{\mathcal{M}'}$
- We show that $\mathcal{M}' \models \phi$



Assume $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$. We show that $\forall \mathcal{M}, \ell [\mathcal{M} \models_{\ell} \phi]$.

Recall ...

- $\phi \equiv \phi_1 \wedge \phi_2 \implies \phi_3$

Cases:

- $\mathcal{M}' \not\models \phi_1 \vee \mathcal{M}' \not\models \phi_2$. Done! Why?
- $\mathcal{M}' \not\models \phi_1 \wedge \mathcal{M}' \not\models \phi_2$.



Assume $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$. We show that $\forall \mathcal{M}, \ell [\mathcal{M} \models_{\ell} \phi]$.

Interpreting Binary Strings:

$$\text{interpret}(\epsilon) \stackrel{\text{def}}{=} e^{\mathcal{M}'}$$

$$\text{interpret}(x0) \stackrel{\text{def}}{=} f_0^{\mathcal{M}'}(\text{interpret}(x))$$

$$\text{interpret}(x1) \stackrel{\text{def}}{=} f_1^{\mathcal{M}'}(\text{interpret}(x))$$



Assume $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$. We show that $\forall \mathcal{M}, \ell [\mathcal{M} \models_{\ell} \phi]$.

Notice

$\text{interpret}(\langle b_1, b_2, \dots, b_r \rangle) = f_{b_r}^{\mathcal{M}'}(f_{b_{r-1}}^{\mathcal{M}'} \dots (f_{b_2}^{\mathcal{M}'}(f_{b_1}^{\mathcal{M}'}(e^{\mathcal{M}'})) \dots))$ is just the meaning of $f_s(e)$ in A' where $s = \langle b_1, b_2, \dots, b_r \rangle$.

Recall ...

- $\phi_1 \equiv \bigwedge_{i=1}^n P(f_{s_i}(e), f_{t_i}(e))$

Thus ...

$\mathcal{M}' \models \phi_1 \implies P^{\mathcal{M}'}(\text{interpret}(ss_i), \text{interpret}(tt_i))$ for $i = 1, 2, \dots, n$.



Assume $\exists \langle i_1, i_2, \dots, i_n \rangle [n \geq 1 \wedge s_{i_1} s_{i_2} \dots s_{i_n} = t_{i_1} t_{i_2} \dots t_{i_n}]$. We show that $\forall \mathcal{M}, \ell [\mathcal{M} \models_{\ell} \phi]$.

Starting with $(s, t) = (s_{i_1}, t_{i_1})$, can repeatedly obtain

$$P^{\mathcal{M}'}(\text{interpret}(s_{i_1} s_{i_2} \dots s_{i_n}), \text{interpret}(t_{i_1} t_{i_2} \dots t_{i_n})). \quad (1)$$

In Conclusion ...

Since $s_{i_1} s_{i_2} \dots s_{i_n}$ and $t_{i_1} t_{i_2} \dots t_{i_n}$ form a solution of C , they are the same elements in A' . (1) verifies $\exists z [P(z, z)]$ in \mathcal{M}' and thus $\mathcal{M}' \models \phi_3$.