Interpretation and Model Checking in First Order Logic

Fundamentals of Artificial Intelligence

Slides are mostly adapted from AIMA and MIT Open Courseware

FOL Interpretations

- Interpretation I
  - U set of objects
    (called "domain of discourse" or "universe")
  - Maps constant symbols to elements of U
  - Maps predicate symbols to relations on U
    (binary relation is a set of pairs)
  - Maps function symbols to functions on U
    (function is a binary relation with a single pair for each element in U, whose first item is that element)

Holds

When does a sentence hold in an interpretation?

- P is a relation symbol
- $t_1, \ldots, t_n$ are terms

holds($P(t_1, \ldots, t_n)$, I) iff $<I(t_1), \ldots, I(t_n)> \in I(P)$

Brother(Jon, Joe)??

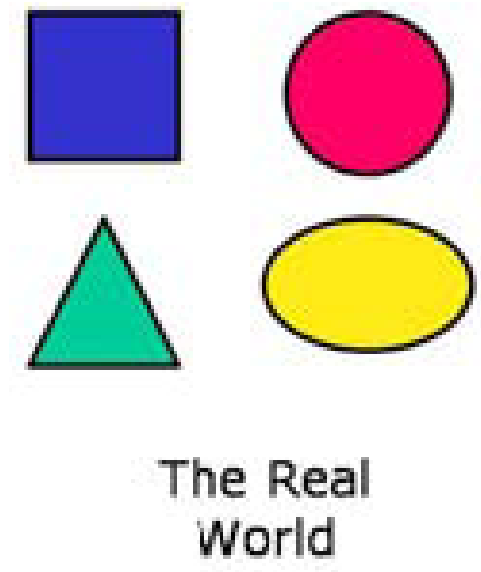
- I(Jon) =  [an element of U]
- I(Joe) =  [an element of U]
- I(Brother) = {<, >, <, >, <…,…>, … }

Extend an interpretation I to bind variable x to element $a \in U$: $I_{x/a}$

- holds($\forall x.\Phi$, I) iff holds(Φ, $I_{x/a}$) for all $a \in U$
- holds($\exists x.\Phi$, I) iff holds(Φ, $I_{x/a}$) for some $a \in U$

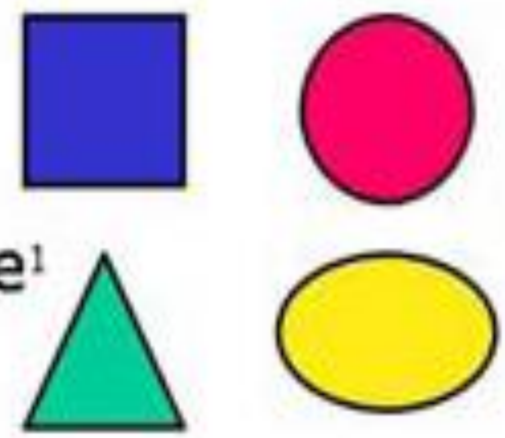
Example Domain

- U = {}



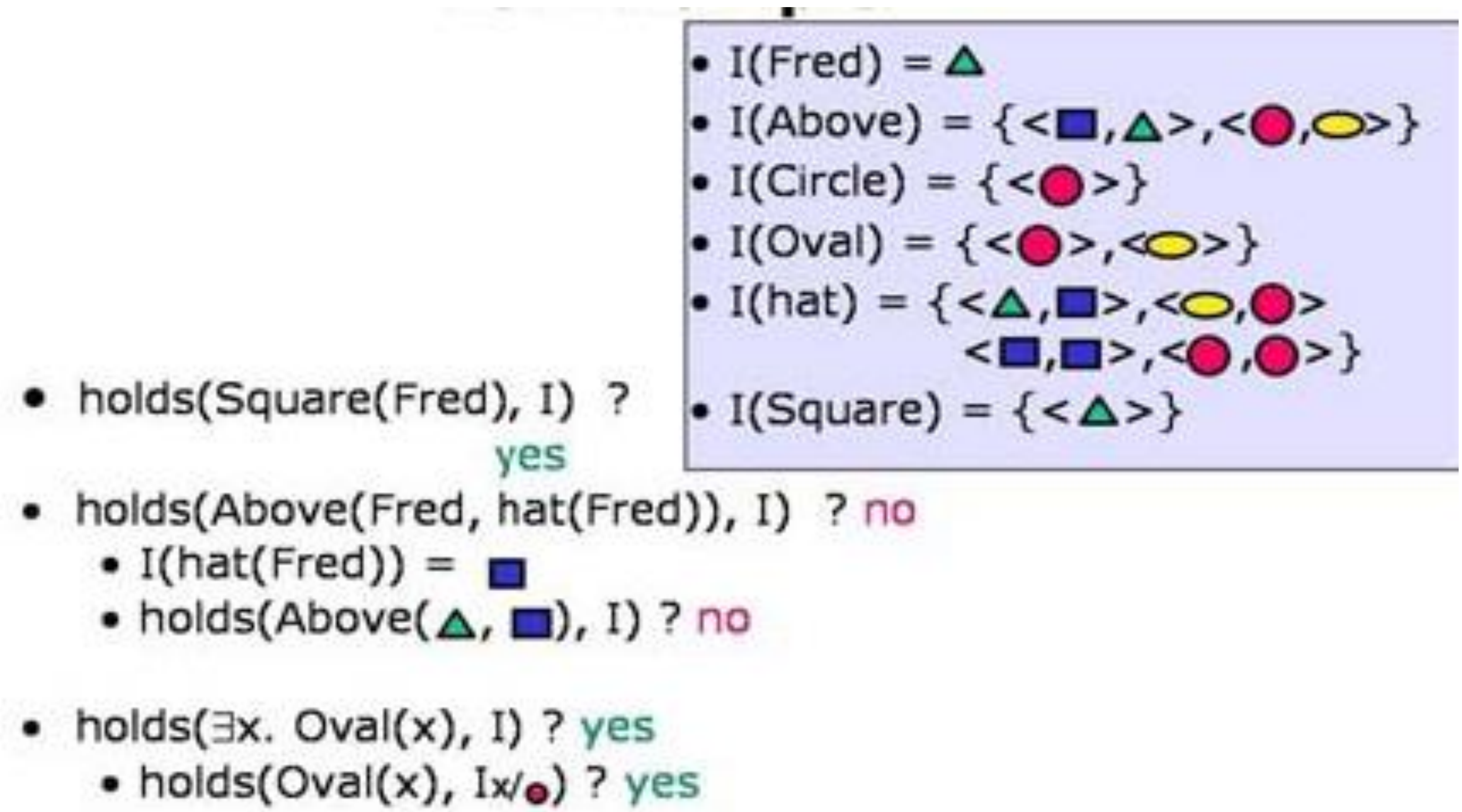
Example Domain

- U = {■, ▲, ●, ⬭}
- Constants: Fred
- Preds: Above[2], Circle[1], Oval[1], Square[1]
- Function: hat
- I(Fred) = ▲
- I(Above) = {<■, ▲>, <●, ⬭>}
- I(Circle) = {<●>}
- I(Oval) = {<●>, <⬭>}
- I(hat) = {<▲, ■>, <⬭, ●>, <■, ■>, <●, ●>}
- I(Square) = {<▲>}

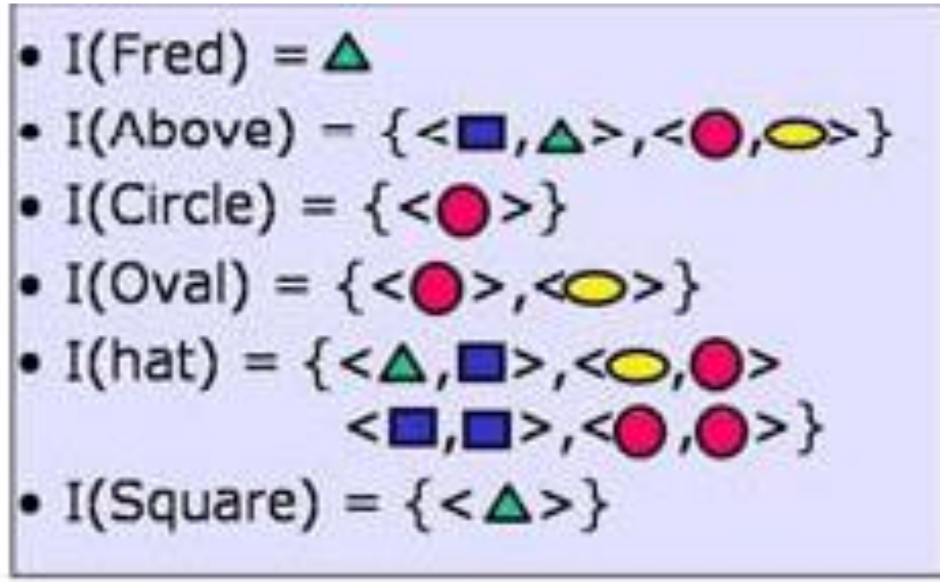


The Real World

Example Domain



Example Domain

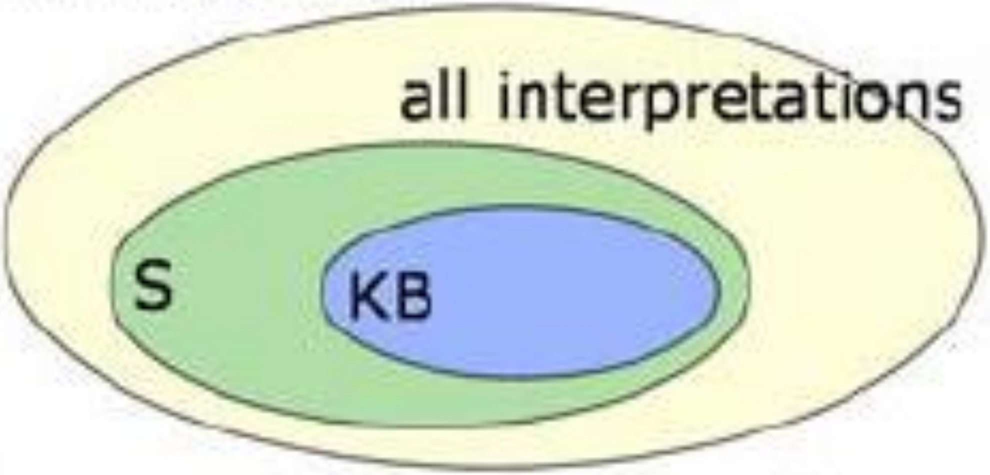


- holds(∀x. ∃y. Above(x,y) v Above(y, x), I) ? yes
  - holds(∃y. Above(x,y) v Above(y,x), Ix/△) ? yes
    holds(Above(x,y) v Above(y,x), Ix/△,y/■) ? yes
  - verify for all other values of x

- holds(∀ x. ∀ y. Above(x,y) v Above(y,x), I) ? no
  - holds(Above(x,y) v Above(y,x), Ix/■,y/●) ? no

Entailment in First Order Logic

- KB entails S: for every interpretation I, if KB holds in I, then S holds in I



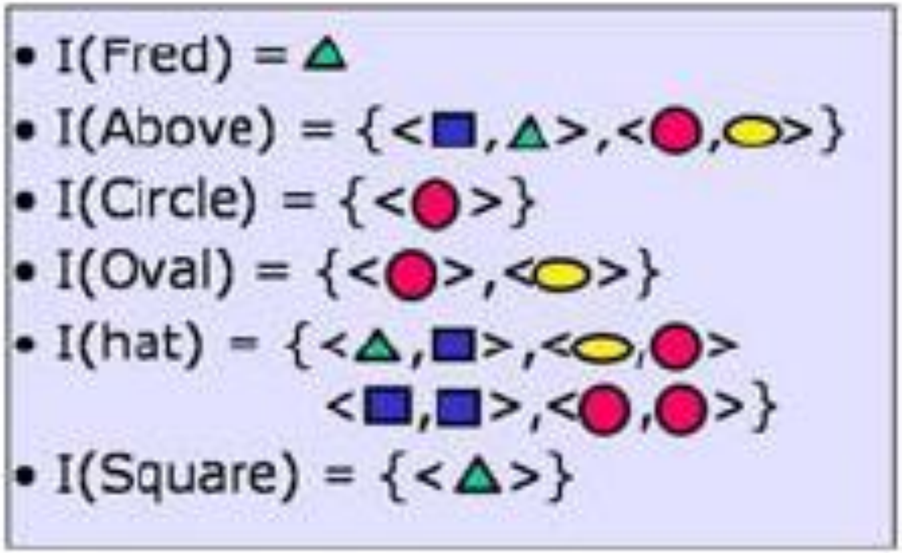
- Computing entailment is impossible in general, because there are infinitely many possible interpretations
- Even computing holds is impossible for interpretations with infinite universes

Intended Interpretations

$$KB : (\forall x.\,\mathrm{Circle}(x) \rightarrow \mathrm{Oval}(x)) \wedge (\forall x.\,\mathrm{Square}(x) \rightarrow \neg \mathrm{Oval}(x))$$

$$S : \forall x.\,\mathrm{Square}(x) \rightarrow \neg \mathrm{Oval}(x)$$

- We know holds(KB, I)
- We wonder whether holds(S, I)
- We could ask: Does KB entail S?
- Or we could just try to check whether holds(S, I)



An Infinite Interpretation

$$KB: (\forall x.\text{Circle}(x) \rightarrow \text{Oval}(x)) \wedge (\forall x.\text{Square}(x) \rightarrow \neg\text{Oval}(x))$$

$$S: \forall x.\text{Square}(x) \rightarrow \neg\text{Oval}(x)$$

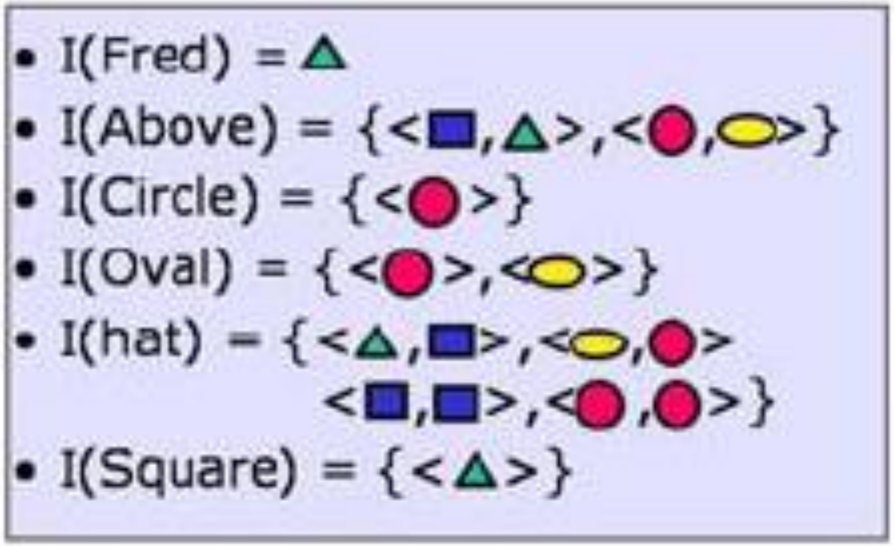
- Does KB hold in I_1?
- Yes, but can't answer via enumerating U
- S also holds in I_1
- No way to verify mechanically

$U_1 = \{1, 2, 3, \ldots\}$
$I_1(\text{circle}) = \{4, 8, 12, 16, \ldots\}$
$I_1(\text{oval}) = \{2, 4, 6, 8, \ldots\}$
$I_1(\text{square}) = \{1, 3, 5, 7, \ldots\}$

An Argument for Entailment

$$KB : (\forall x.\,\mathrm{Circle}(x) \rightarrow \mathrm{Oval}(x)) \wedge (\forall x.\,\mathrm{Square}(x) \rightarrow \neg \mathrm{Oval}(x))$$

$$S_1 : \forall x, y.\ \mathrm{Circle}(x) \wedge \mathrm{Oval}(y) \wedge \neg \mathrm{Circle}(y) \rightarrow \mathrm{Above}(x, y)$$



$U_1 = \{1, 2, 3, \ldots\}$
$I_1(\mathrm{Circle}) = \{4, 8, 12, 16, \ldots\}$
$I_1(\mathrm{Oval}) = \{2, 4, 6, 8, \ldots\}$
$I_1(\mathrm{Square}) = \{1, 3, 5, 7, \ldots\}$
$I_1(\mathrm{Above}) = >$

- holds(KB, I)
- holds(S_1, I)

- holds(KB, I_1)
- fails(S_1, I_1)

KB doesn't entail S_1!

Proof and Entailment

- Entailment captures general notion of "follows from"
- Can't evaluate it directly by enumerating interpretations
- So, we'll do proofs
- In FOL, if S is entailed by KB, then there is a finite proof of S from KB

Axiomatization

- What if we have a particular interpretation, I, in mind, and want to test whether holds(S, I)?
- Write down a set of sentences, called *axioms*, that will serve as our KB
- We would like KB to hold in I, and as few other interpretations as possible
- No matter what,
  - If holds(KB, I) and KB entails S,
  - then holds(S, I)
- If your axioms are weak, it might be that
  - holds(KB, I) and holds(S, I), but
  - KB doesn't entail S

Axiomatization Example

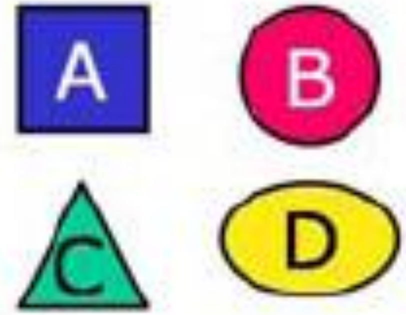
KB_2

$Above(A, C)$

$Above(B, D)$

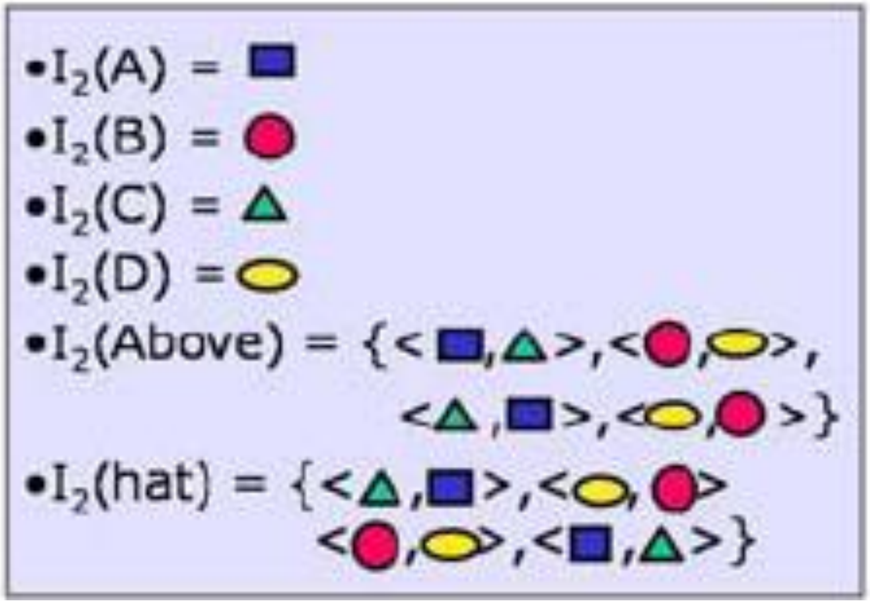
$\forall x, y.\ Above(x, y) \rightarrow hat(y) = x$

$\forall x.\ (\neg \exists y.\ Above(y, x)) \rightarrow hat(x) = x$

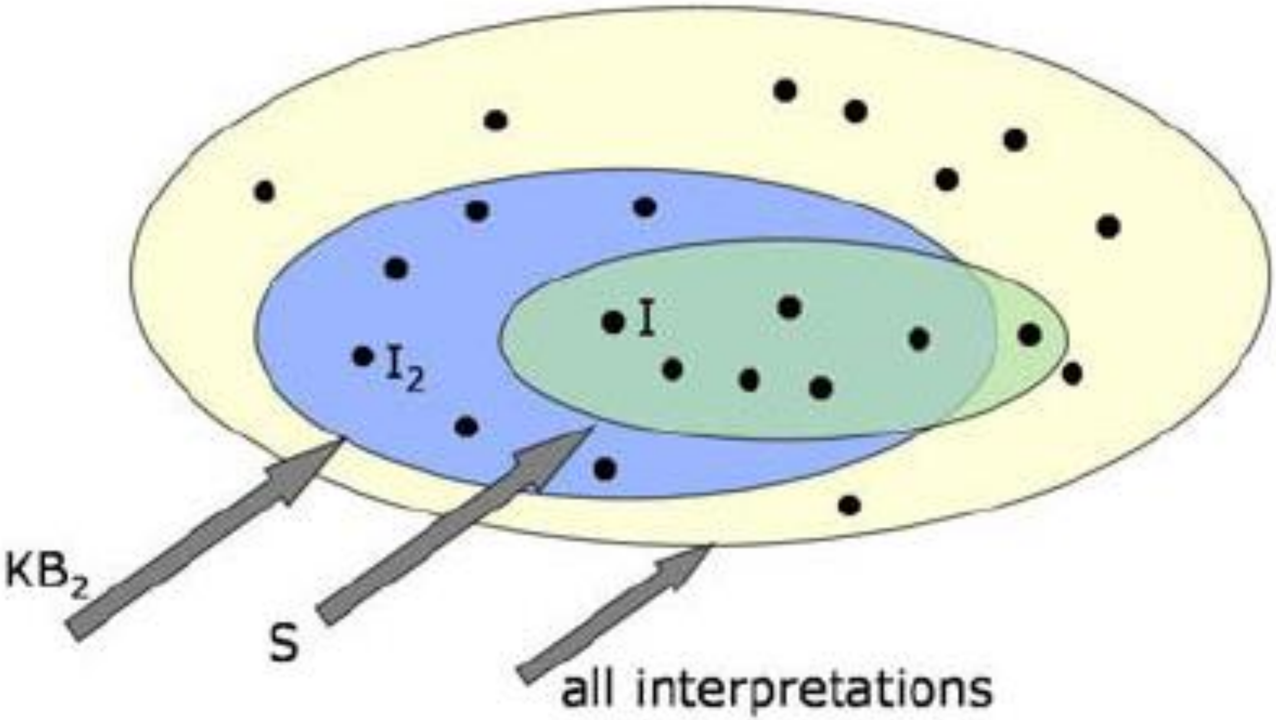


S: $hat(A) = A$

- holds(KB_2, I_2)
- fails(S, I_2)
- KB_2 doesn't entail S



KB2 is a Weakling!



Axiomatization Example: Another Try

KB$_3$

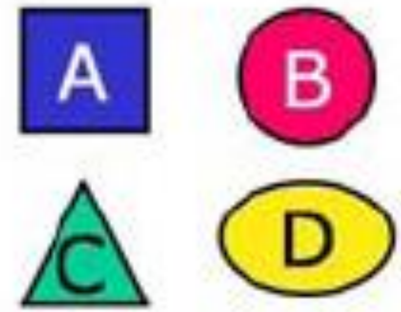
$\text{Above}(A, C)$

$\text{Above}(B, D)$

$\forall x, y.\ \text{Above}(x, y) \rightarrow \text{hat}(y) = x$

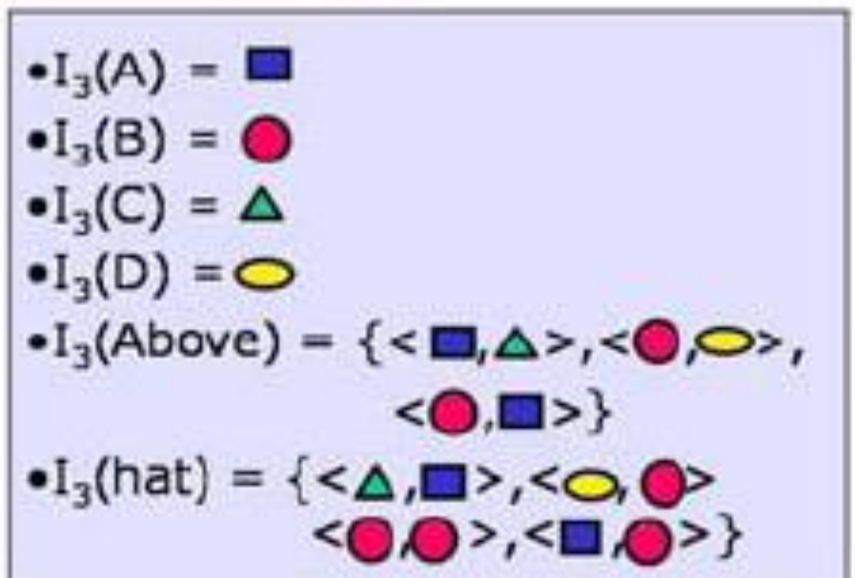
$\forall x.\ (\neg\exists y.\ \text{Above}(y, x)) \rightarrow \text{hat}(x) = x$

$\forall x, y.\ \text{Above}(x, y) \rightarrow \neg\text{Above}(y, x)$



S: $\text{hat}(A) = A$

- fails(KB$_3$, I$_2$)
- holds(KB$_3$, I$_3$)
- fails(S, I$_3$)
- KB$_3$ doesn't entail S



Axiomatization Example: One last time

KB_4

$Above(A, C)$

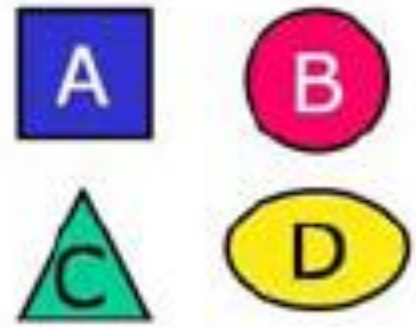
$Above(B, D)$

$\neg \exists x.\ Above(x, A)$

$\neg \exists x.\ Above(x, B)$

$\forall x, y.\ Above(x, y) \rightarrow hat(y) = x$

$\forall x.\ (\neg \exists y.\ Above(y, x)) \rightarrow hat(x) = x$



S: $hat(A) = A$

- fails(KB_4, I_3)
- KB_4 entails S

We'll prove S from KB_4 later.