## CSC 191/291

## Some Rules of Reasoning

The reasoning example we saw for "Robbie's self-model" showed particular logical reasoning examples (Robbie is not a person, Robbie is located somewhere, there is a dog), but without explicitly listing the mechanical rules we can use for such reasoning. Here are 4 very common inference rules, followed by another example:

C is some constant

An inference using 2 steps: <u>Dog(Snoopy)</u>,  $\forall x. Dog(x) \Rightarrow Has-Tail(x)$ Has-Tail(Snoopy)

Can you see which 2 rules have been used?

Horn clause form would use just one step (free variables are implicitly universal):

Dog(Snoopy) Dog(x)  $\Rightarrow$  Has-Tail(x) ------ We <u>unify</u> variable x with constant Snoopy Has-Tail(Snoopy)

Similarly, using clause form, a single resolution (cancellation) step suffices:

<u>Dog(Snoopy),</u> ¬Dog(x) v Has-Tail(x) Has-Tail(Snoopy) Again, we unify as above (cf. rule 3, and the "Robbie" QA example)

These are *deductive rules* (as such completely reliable -- "sound"). But keep in mind that not all reasoning is deduction: We saw various ways of "jumping to conclusions" – in unsound, but generally useful ways; and recall special "analogue methods", for example for "conjuring up" images in the mind (and in computers, we can use computer graphics to store and manipulated such images).



Faulty self-modeling ...

## SEMANTICS OF FOL

Consider the inference Dog (Snoopy), (∀x Dog(x) ⇒ Has-Tail(x)) Has-Tail (Snoopy)

It seems that if the premises are true, so is the conclusion. Why? More generally, how can we distinguish such "truth-preserving" inferences from others??

We need a theory of truth for FOL sentences !

This only makes sense if FOL sentences are taken to be assertions about some domain.

This in turn requires that terms like <u>Dog</u>, <u>Snoopy</u>, <u>Has-Tail</u>,... refer to (denote) something in the domain.

Here's how we can do it (example): denotes Snoopy denotes Has-Tail (set of all things with a tail)

The "denotes" function, say I, is called an interpretation. We say Dog(Snoopy) is true under interpretation I iff I(Snoopy) & I(Dog) Clearly this is the case for Dog(Snoopy). We write (as a first approximation) FT Dog (Snoopy) . For the given I, it's also clear that Fr Has-Tail (Snoopy) It's also the case (less obviously ... ) that FI (Vx (Dog(x) => Has-Tail (x)) Truth conditions for  $\forall$  -- informally:  $I(Dog) \subseteq I(Has-Tail)$ , i.e., the set of dogs is a subset of the things with property Has-Tail. So,  $I(Snoopy) \in I(Dog) \subseteq I(Has-Tail)$ 

<b>SC</b> :	LIN /444
-------------	----------

Another interpretation, I': Snoopy I' + Dog I' {4,8,12,16,...} Has-Tail - I' {1, 4, 6, 8, 9, 10, 12, 14, 15] All 3 formulas are again true under I'! Formal "soundness" (truth-preservation) claim about the original inference: For any interpretation I of <u>Snoopy</u>, Dog, <u>Has-Tail</u> (where I interprets <u>Snoopy</u> as an individual and Dog, <u>Has-Tail</u> as sets of individuals), if FIDog(Snoopy) and FI(∀x Dog(x) ⇒ Has. Tail(x)) then FI Has-Tail (Snoopy)

Let's get a little more precise & general. First, before we talk about interpretations of particular symbols we should fix the domain 9 of all the things we may wish to talk about in our FO Language. Then we interpret individual constants c as individuals in D, i.e.,  $I(\tau) \in \mathcal{D}$  also written  $\tau^{T} \in \mathcal{D}$ We interpret monadic (1-place) predicate constants or as sets of individuals in S, i.e.,  $I(\pi) \leq \mathcal{D}$  also written  $\pi^{T} \in \mathcal{D}$ For truth, we will now write Fre φ, where M = (D, I), the model. what is I(TT) for 2-place, 3-place, etc. prechicates?



Given this interpretation of 'Loves', and given Boy(Paris),  $\forall x. Boy(x) \Rightarrow Loves(x, Juliet),$ Can we semantically justify the conclusion Loves(Paris, Juliet), i.e., <I(Paris), I(Juliet)> ∈ I(Loves)? The first formula says  $I(Paris) \in I(Boy)$ ; The second formula says that the set I(Boy) is a subset of  $\{b \mid \langle b, j \rangle \in I(Loves)\}$ (where j = A = I(Juliet)). So,  $I(Paris) \in I(Boy) \subseteq \{b \mid \langle b, j \rangle \in I(Loves)\}$ 

To specify truth conditions for ' $\forall$ ' formally requires an inductive truth definition – covered in CSC 244.