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:

\[ \phi \land \psi \Rightarrow \theta, \forall x. \phi \Rightarrow \exists x. \psi \Rightarrow \theta \]

\[ \forall x. \phi \Rightarrow \exists x. \psi \Rightarrow \theta \]

\[ \phi \Rightarrow \psi \]

\[ \phi \Rightarrow \exists x. \psi \Rightarrow \theta \]

**Can you see which rules we used for the Robbie examples?**

An inference using 2 steps:  
\[ \text{Dog}(Snoopy), \ \forall x. \text{Dog}(x) \Rightarrow \text{Has-Tail}(x) \]
\[ \Rightarrow \text{Has-Tail}(Snoopy) \]

**Can you see which 2 rules have been used?**

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

\[ \text{Dog}(Snoopy) \]
\[ \Rightarrow \text{Has-Tail}(Snoopy) \]

---

We unify variable \( x \) with constant \( Snoopy \)

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

\[ \text{Dog}(Snoopy), \ \neg \text{Dog}(x) \]
\[ \Rightarrow \text{Has-Tail}(Snoopy) \]

Again, we unify as above (cf. rule 3, Has-Tail(Snoopy) 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).
Here's how we can do it (example):

Snoopy \(\text{denotes} \quad I(\text{Snoopy})\)

Dog \(\text{denotes} \quad \{\text{set of all dogs}\}\)

Has-Tail \(\text{denotes} \quad \{\text{set of all things with a tail}\}\)

The "denotes" function, say \(I\), is called an \underline{interpretation}.

We say \(\text{Dog(Snoopy)}\) is \underline{true} under interpretation \(I\) iff

\[ I(\text{Snoopy}) \in I(\text{Dog}) \]

Clearly this is the case for \(\text{Dog(Snoopy)}\).

We write (as a first approximation)

\[ \models_I \text{ Dog(Snoopy)} \]

For the given \(I\), it's also clear that

\[ \models_I \text{ Has-Tail(Snoopy)} \]

It's also the case (less obviously...) that

\[ \models_I (\forall x (\text{Dog}(x) \Rightarrow \text{Has-Tail}(x))) \]

Truth conditions for \(\forall\) -- informally:

\[ I(\text{Dog}) \subseteq I(\text{Has-Tail}), \quad \text{i.e., the set of dogs is a subset of the things with property Has-Tail.} \]

So, \[ I(\text{Snoopy}) \in I(\text{Dog}) \subseteq I(\text{Has-Tail}) \]
Another interpretation, \( I' \):

\[
\begin{align*}
\text{Snoopy} & \quad I' \quad \rightarrow \quad 4 \\
\text{Dog} & \quad I' \quad \rightarrow \quad \{4, 8, 12, 16, \ldots\} \\
\text{Has-Tail} & \quad I' \quad \rightarrow \quad \{1, 4, 6, 8, 9, 10, 12, 14, 16, \ldots\}
\end{align*}
\]

All 3 formulas are again true under \( I' \)!

Formal "soundness" (truth-preservation) claim about the original inference:

For any interpretation \( I \) of Snoopy, Dog, Has-Tail (where \( I \) interprets Snoopy as an individual and Dog, Has-Tail as sets of individuals), if

\[ \models_I \text{Dog(Snoopy)} \text{ and } \models_I (\forall x \text{ Dog}(x) \Rightarrow \text{Has-Tail}(x)) \]

then

\[ \models_I \text{Has-Tail(Snoopy)} \]

Let's get a little more precise & general.

First, before we talk about interpretations of particular symbols we should fix the domain \( \mathfrak{D} \) of all the things we may wish to talk about in our FO language.

Then we interpret individual constants \( c \) as individuals in \( \mathfrak{D} \), i.e.,

\[ \models_I (c) \in \mathfrak{D} \]

also written \( c^I \in \mathfrak{D} \)

We interpret monadic (1-place) predicate constants \( \pi \) as sets of individuals in \( \mathfrak{D} \), i.e.,

\[ \models_I (\pi) \subseteq \mathfrak{D} \]

also written \( \pi^I \subseteq \mathfrak{D} \)

For truth, we will now write

\[ \models_M \varphi \]

where \( M = (\mathfrak{D}, I) \), the model.

What is \( I(\pi) \) for 2-place, 3-place, etc. predicates?
Given this interpretation of 'Loves', and given 
\[ \text{Boy(Paris), } \forall x. \text{Boy}(x) \Rightarrow \text{Loves}(x,\text{Juliet}), \]

**Can we semantically justify the conclusion**

\[ \text{Loves}(\text{Paris,Juliet}), \text{i.e.,} \]
\[ <\text{I}(\text{Paris}), \text{I}(\text{Juliet})> \in \text{I}(\text{Loves})? \]

The first formula says \( \text{I}(\text{Paris}) \in \text{I}(\text{Boy}) \)

The second formula says that the set \( \text{I}(\text{Boy}) \)

is a subset of \( \{ b \mid <b,j> \in \text{I}(\text{Loves}) \} \)

( where \( j = \text{I}(\text{Juliet}) \)).

So,

\[ \text{I}(\text{Paris}) \in \text{I}(\text{Boy}) \subseteq \{ b \mid <b,j> \in \text{I}(\text{Loves}) \} \]

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