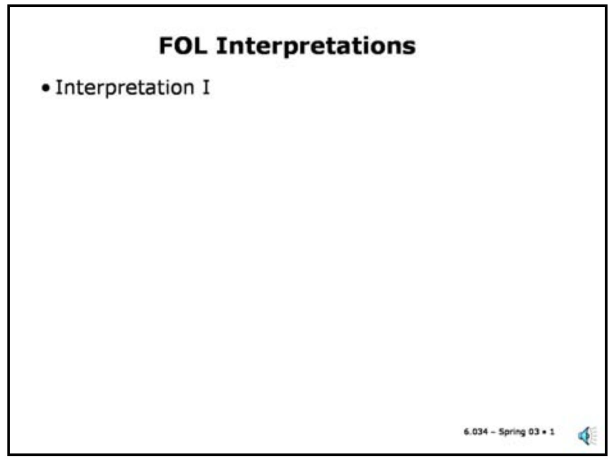
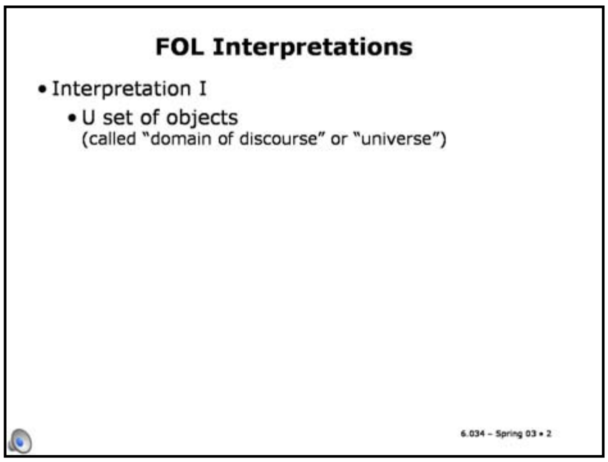
6.034 Notes: Section 9.5

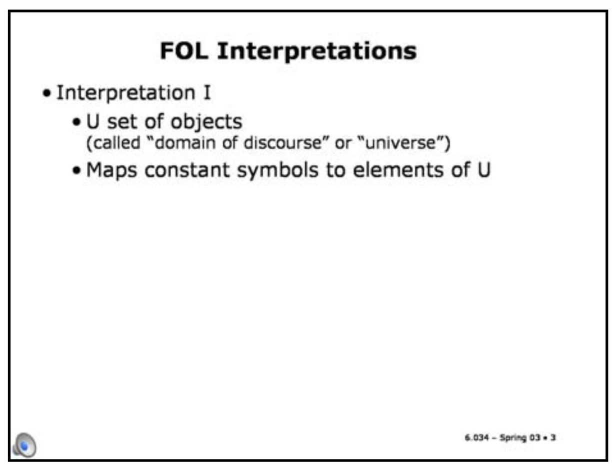
Slide 9.5.1
We're going to do the semantics informally. This isn't really going to look informal to you, but compared to the sorts of things that logicians write down, it's pretty informal. In propositional logic, an interpretation (**I**) is an assignment of truth values to sentential variables. Now an interpretation's going to be something more complicated. An interpretation is made up of a set and three mappings.

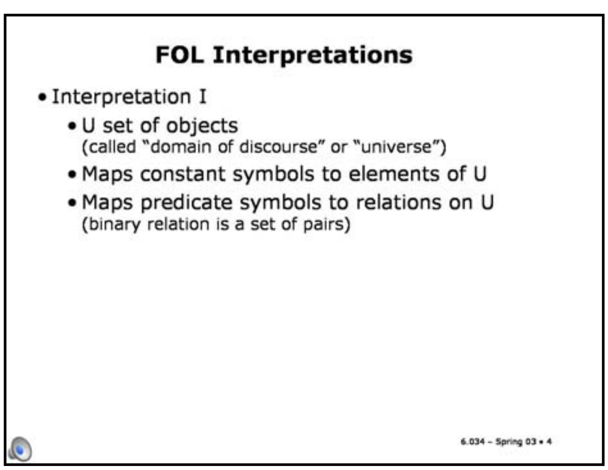




Slide 9.5.2
The set is the universe, **U**, which is a set of objects. So what's an object? Well, really, it could be this chair and that chair and these pieces of chalk or it could be all of you guys or it could be some trees out there, or it could be rather more abstract objects like meetings or points in time or numbers. An object could be anything you can think of, and the universe can be any set (finite or infinite) of objects. The universe is also sometimes called the "domain of discourse."

Slide 9.5.3
There's a mapping from constant symbols to elements of **U**, specifying how names are connected to objects in the world. So I might have the constant symbol, **Fred**, and I might have a particular person in the universe, and then the interpretation of the symbol **Fred** could be that person.

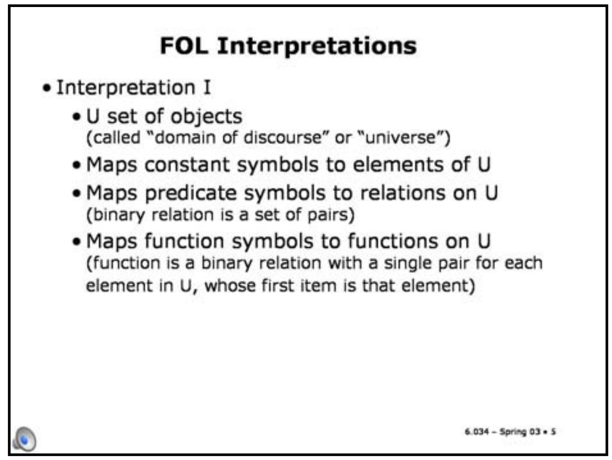


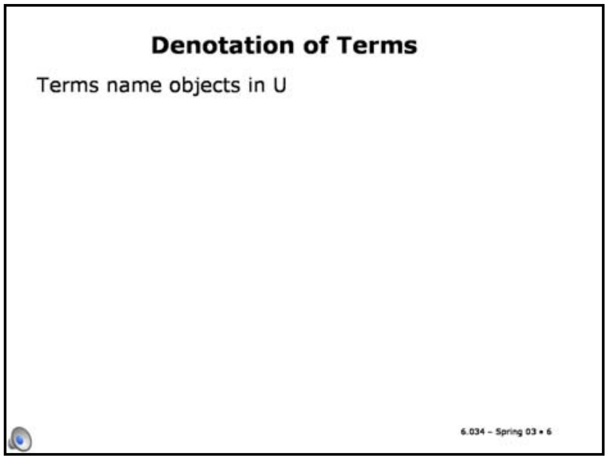


Slide 9.5.4
The next mapping is from predicate symbols to relations on **U**. An n-ary relation is a set of lists of n objects, saying which groups of things stand in that particular relation to one another. A binary relation is a set of pairs. So if I have a binary relation **brother-of** and **U** is a bunch of people, then the relation would be the set of all pairs of people such that the second is the brother of the first.

Slide 9.5.5
The last mapping is from function symbols to functions on **U**. Functions are a special kind of relation, in which, for any particular assignment of the first n-1 elements in each list, there is a single possible assignment of the last one. In the **brother-of** relation, there could be many pairs with the same first item and a different second item, but in a function, if you have the same first item then you have to have the same second item. So that means you just name the first item and then there's a unique thing that you get from applying the function. So it's OK for **mother-of** to be a function, discounting adoptions and other unusual situations. We will also, for now, assume that our functions are *total*, which means that there is an entry for every possible assignment of the first n-1 elements.

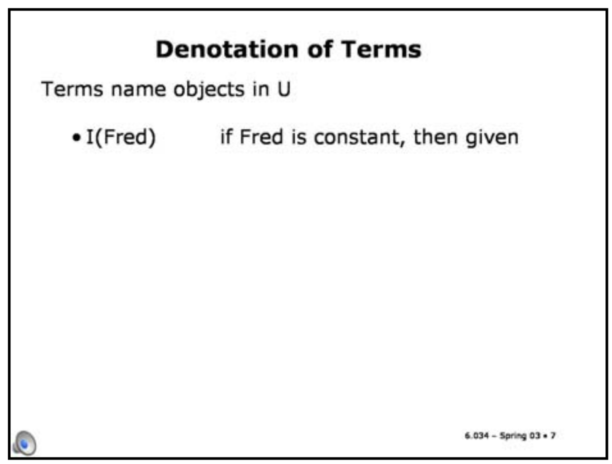
So, the last mapping is from function symbols to functions on the universe.

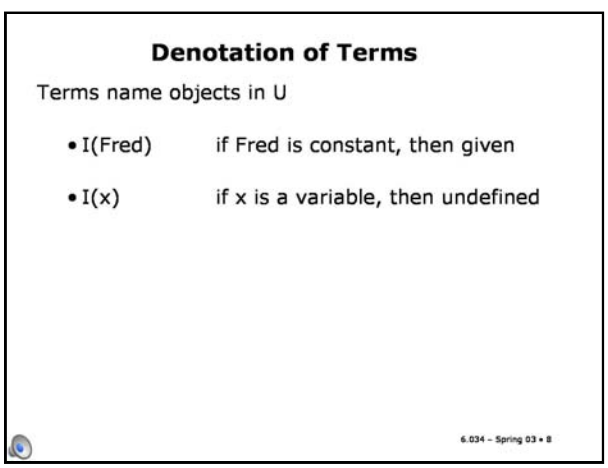




Slide 9.5.6
Before we can do the part of semantics that says what sentences are true in which interpretation, we have to talk about what terms mean. Terms name things, but we like to be fancy so we say a term denotes something, so we can talk about the denotation of a term, that is, the thing that a term names.

Slide 9.5.7
The denotations of constant symbols are given directly in the interpretation.

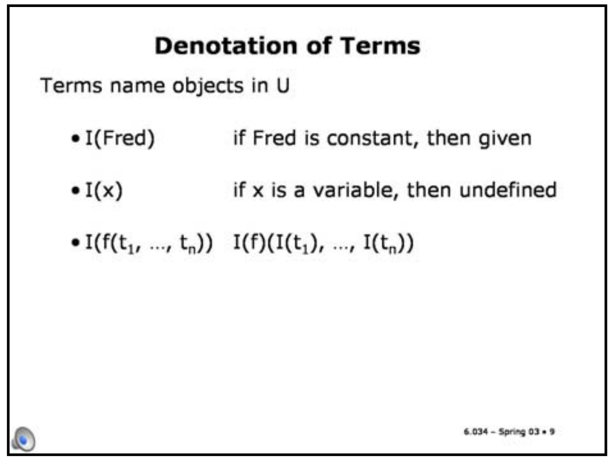


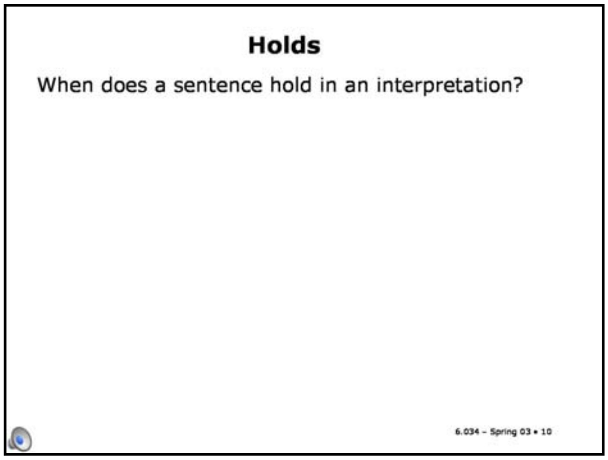


Slide 9.5.8
The denotation of a variable is undefined. What does x mean, if x is a variable? The answer is, "mu." It doesn't mean anything. That's a Zen joke. If you don't get it, don't worry about it.

Slide 9.5.9

The denotation of a complex term is defined recursively. So, to find the interpretation of a function symbol applied to some terms, first you look up the function symbol in the interpretation and get a function. (Remember that the function symbol is a syntactic thing, ink on paper, but the function it denotes is an abstract mathematical object.) Then you find the interpretations of the component terms, which will be objects in **U**. Finally, you apply the function to the objects, yielding an object in **U**. And that object is the denotation of the complex term.



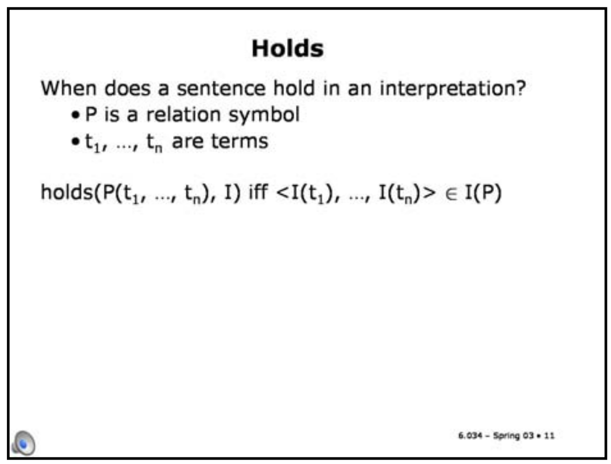


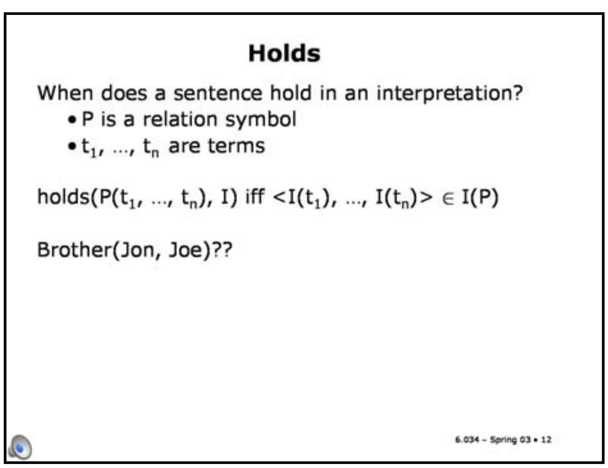
Slide 9.5.10

In the context of propositional logic, we looked at the rules of semantics, which told us how to determine whether a sentence was true in an interpretation. Now, in first-order logic, we'll add some semantic rules, for the new kinds of sentences we've introduced. One of our new kinds of sentences is a predicate symbol applied to a bunch of terms. That's a sentence, which is going to have a truth value, true or false.

Slide 9.5.11

To figure out its truth value, we first use the denotation rules to find out which objects are named by each of the terms. Then, we look up the predicate symbol in the interpretation, which gives us a mathematical relation on **U**. Finally, we look to see if the list of objects named by the terms is a member of the relation. If so, the sentence is true in the given interpretation.

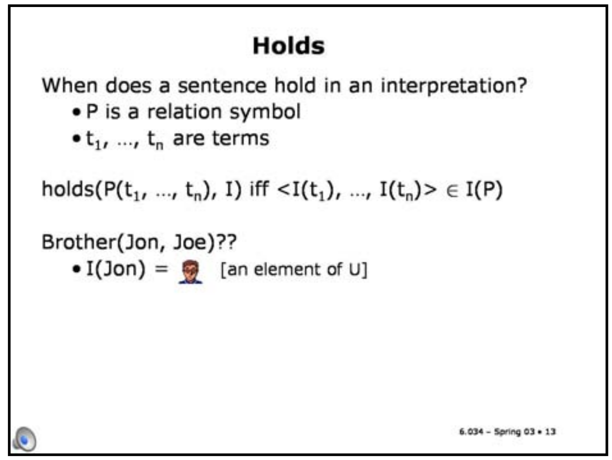




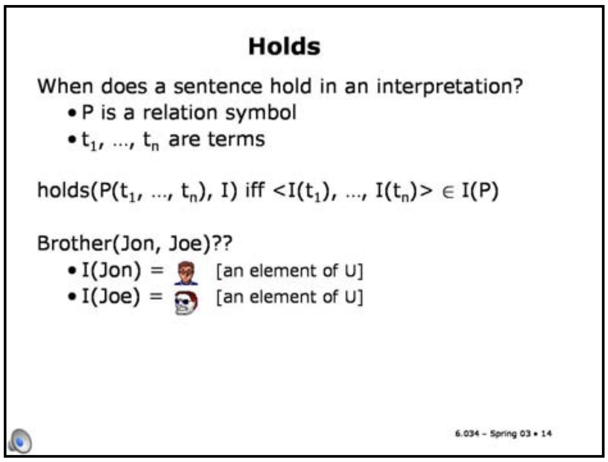
Slide 9.5.12

Let's look at an example. Imagine we want to determine whether the sentence **Brother(Jon,Joe)** is true in some interpretation.

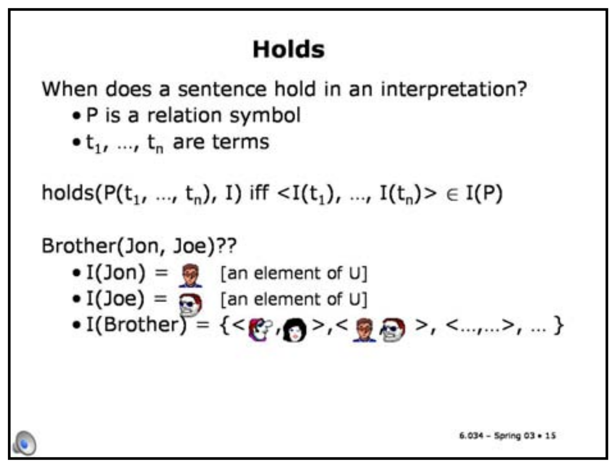
Slide 9.5.13
First, we look up the constant symbol **Jon** in the interpretation and find that it names this guy with glasses.



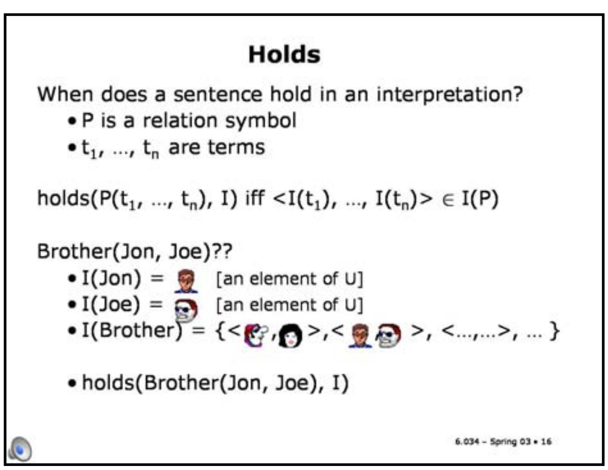
Slide 9.5.14
Then we look up **Joe** and find that it names this angry-looking guy.



Slide 9.5.15
Now we look up the predicate symbol **Brother** and find that it denotes this complicated relation.

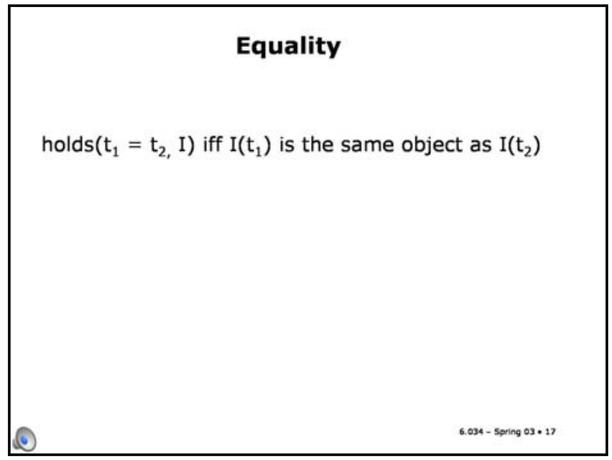


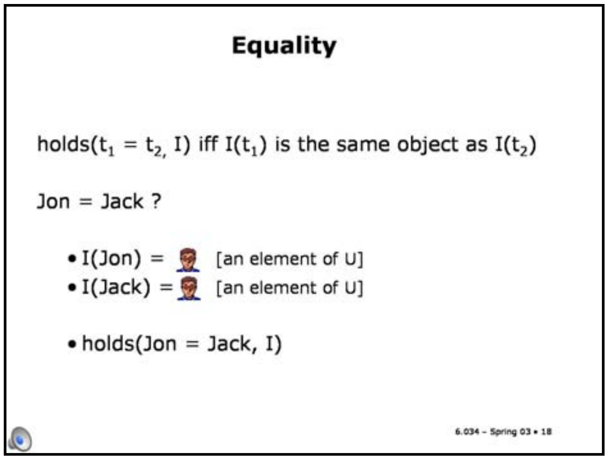
Slide 9.5.16
Finally, we look up the pair of the guy with glasses and the angry-looking guy, to see if they're in the relation. They are, so the sentence must be true in that interpretation. It's easy to think of lots of other interpretations in which it wouldn't be true (and lots of others in which it would).



Slide 9.5.17

Another new kind of sentence we introduced has the form **$term_1$ = $term_2$**. The semantics are pretty unsurprising: if the object denoted by **$term_1$** is the same as the object denoted by **$term_2$**, then the sentence holds.





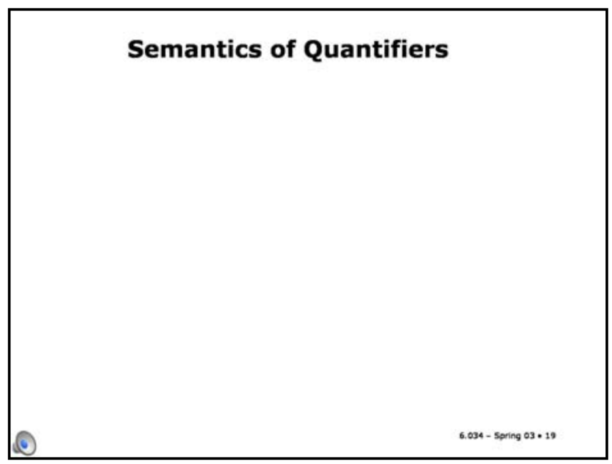
Slide 9.5.18

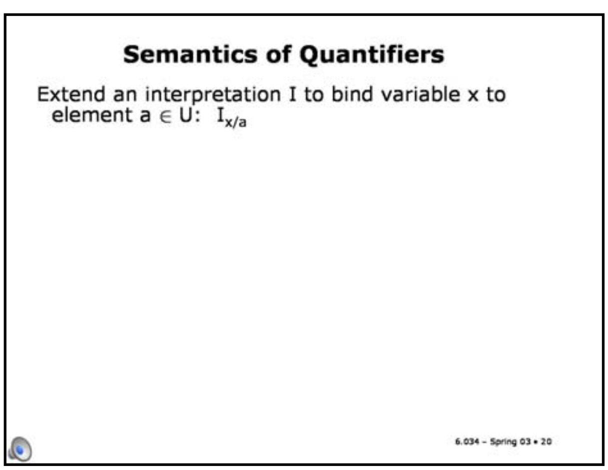
It's important to note that two different constant symbols can denote the same object in the universe; so this is not a test on equality of names.

We might have an interpretation that maps the symbols **Jon** and **Jack** both into the same guy. In that case, **Jon = Jack** holds in **I**.

Slide 9.5.19

Now we have to figure out how to tell whether sentences with quantifiers in them are true.

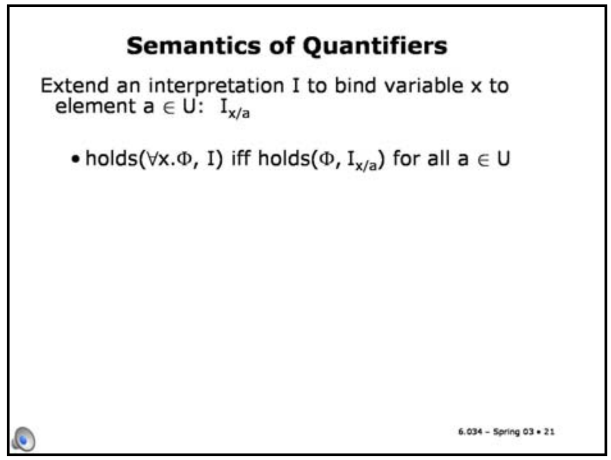


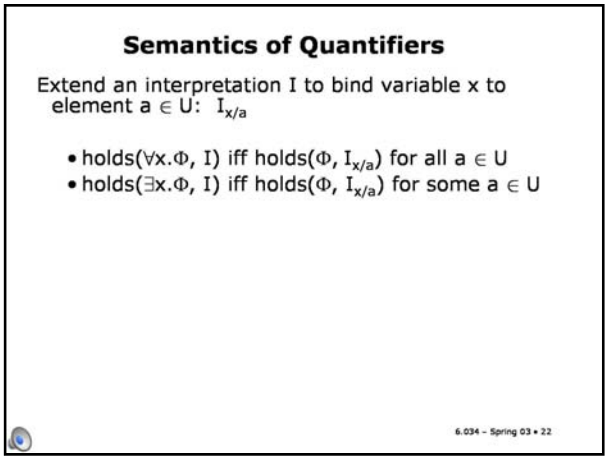


Slide 9.5.20

In order to talk about quantifiers we need the idea of extending an interpretation. We would like to be able to extend an interpretation to bind variable **x** to value **a**. We'll write that as **I** with **x** bound to **a**. Here, **x** is a variable and **a** is an object; an element of **U**. The idea is that, in order to understand whether a sentence that has variables in it is true or not, we have to make various temporary assignments to the variables and see what the truth value of the sentence is. Binding **x** to **a** is kind of like adding **x** as a constant symbol to **I**. It's kind of like temporarily binding a variable in a programming language.

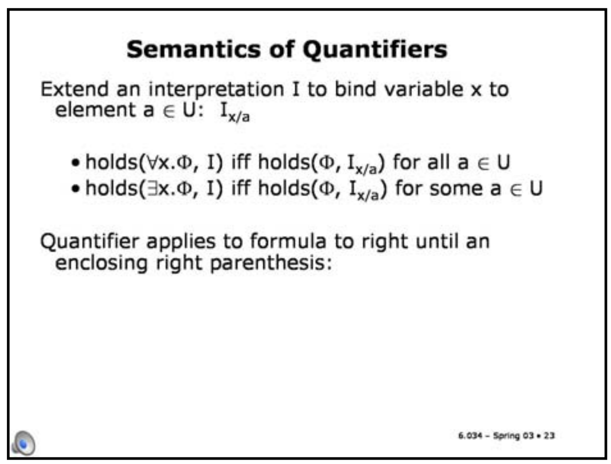
Slide 9.5.21
Now, how do we evaluate the truth under interpretation **I**, of the statement **for all x, Phi**? So how do we know if that's true? Well, it's true if and only if **Phi** is true if for every possible binding of variable **x** to thing in the world **a**. Okay? For every possible thing in the world that you could plug in for **x**, this statement's true. That's what it means to say **for all x, Phi**.

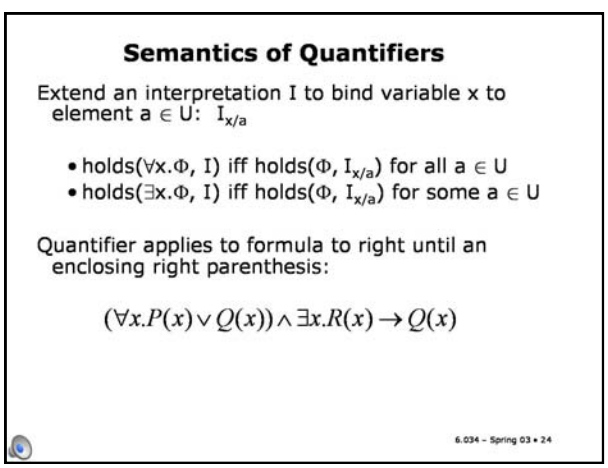




Slide 9.5.22
Similarly, to say that **there exists x such that Phi**, it means that **Phi** has to be true for some **a** in **U**. That is to say, there has to be something in the world such that if we plug that in for **x**, then Phi becomes true.

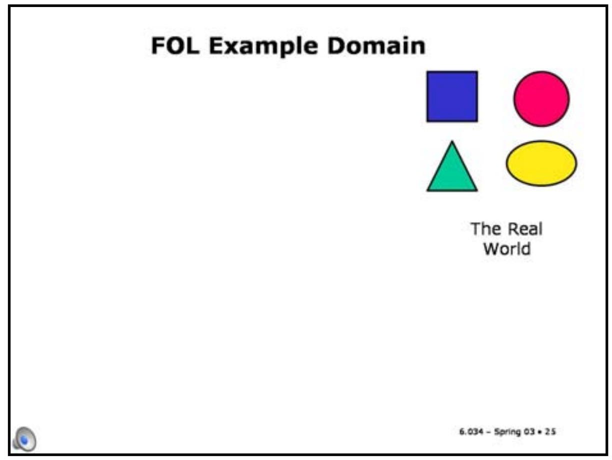
Slide 9.5.23
It's hard to understand the precedence of these operators using the usual rules. A quantifier is understood to apply to everything to its right in the formula, stopping only when it reaches an enclosing close parenthesis.

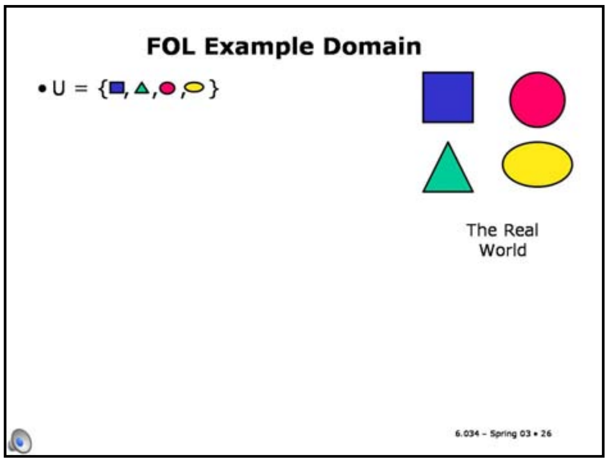




Slide 9.5.24
So in this example sentence, the **for all x** applies until the close paren after the first **Q(x)**; and the **exists x** applies to the end of the sentence.

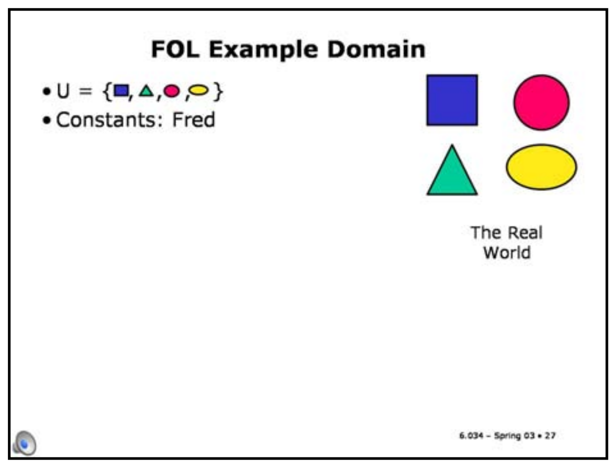
Slide 9.5.25
All right, let's work on an example. Here's a picture of our world.

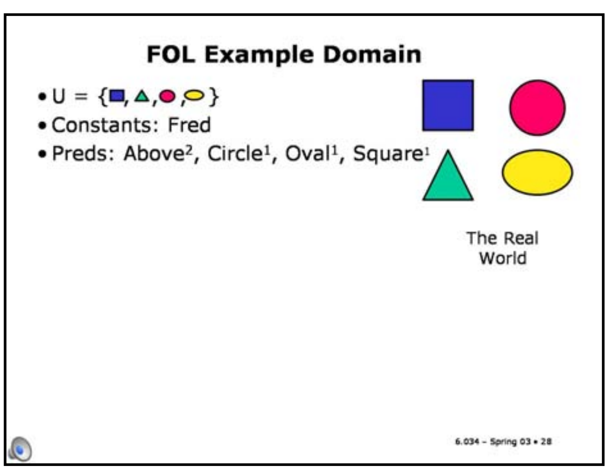




Slide 9.5.26
There are four things in our **U**. Here they are.

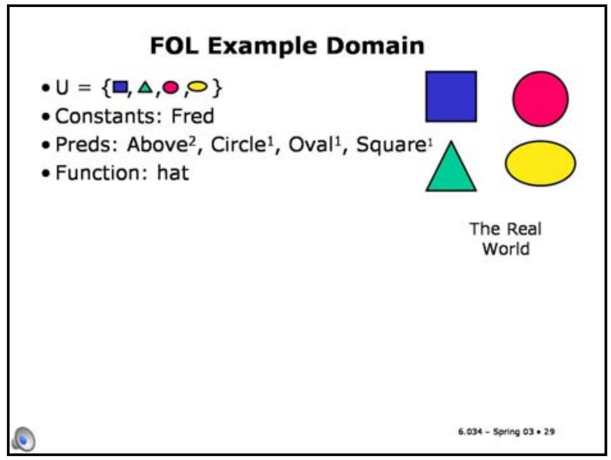
Slide 9.5.27
We have one constant symbol, Fred.

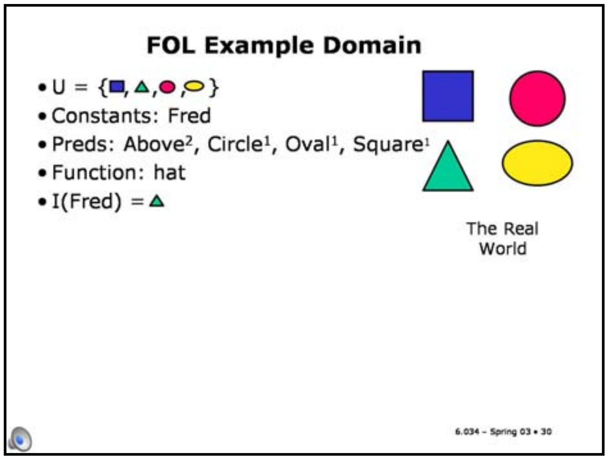




Slide 9.5.28
We have four predicates: **Above**, **Circle**, **Oval**, **Square**. The numbers above them indicate their arity, or the number of arguments they take. Now these particular predicate names suggest a particular interpretation. The fact that I used this word, "circle", makes you guess that probably the interpretation of circle is going to be true for the red object. But of course it needn't be. The fact that those marks on the page are like an English word that we think means something about the shape of an object, that doesn't matter. The syntax is just some words that we write down on our page. But it helps us understand what's going on. It's just like using reasonable variable names in a program that you might write. When you call a variable "the number of times I've been through this loop," that doesn't mean that the computer knows what that means. It's the same thing here.

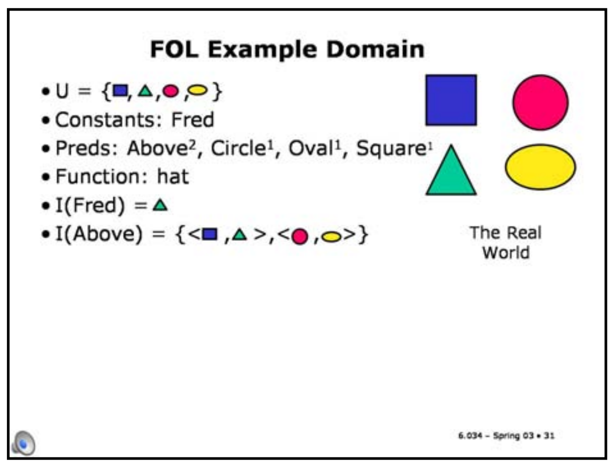
Slide 9.5.29
And we have one function symbol, called **hat**, that takes a single argument.

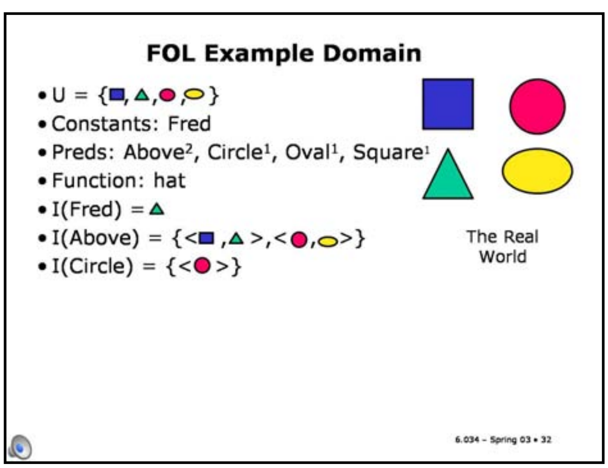




Slide 9.5.30
Now we can talk about a particular interpretation, **I**. We'll define **I** so that **I(Fred)** is the triangle.

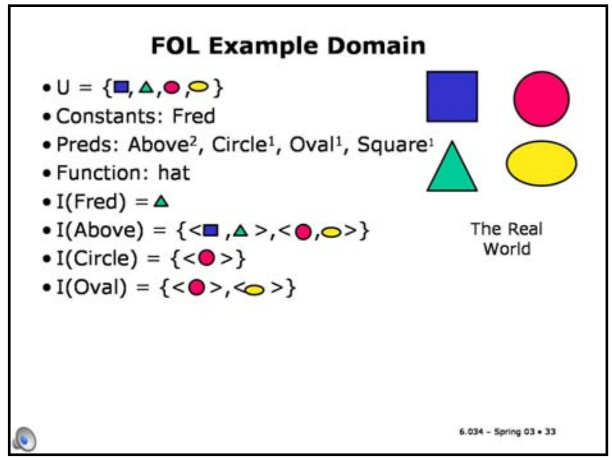
Slide 9.5.31
Now, what kind of a thing is **I(Above)**? Well, **Above** is a predicate symbol, and the interpretation of a predicate symbol is a relation, so **I(Above)** is a relation. Here's the particular relation we define it to be; it's a set of pairs, because **Above** has arity 2. It contains every pair of objects for which we want the relation **Above** to be true.

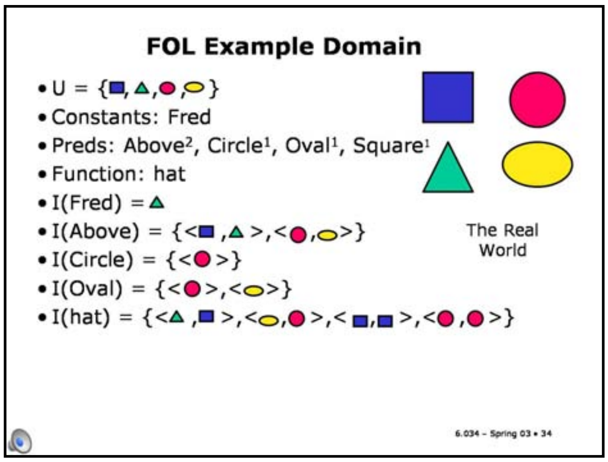




Slide 9.5.32
The interpretation of **Circle** is a unary relation. As you might expect in this world, it's the singleton set, whose element is a one-tuple containing the circle. (Of course, it doesn't have to be!).

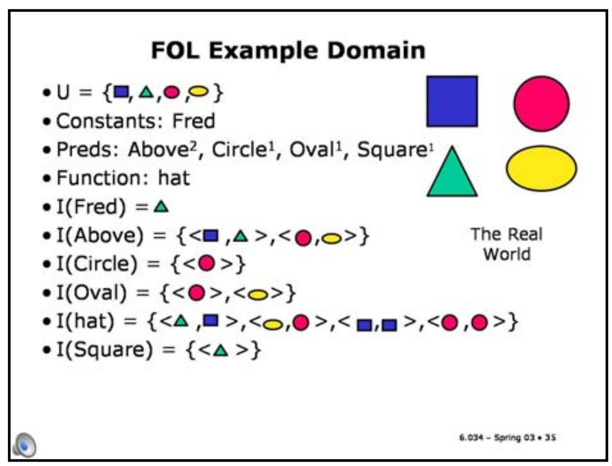
Slide 9.5.33
We'll interpret the predicate **Oval** to be true of both the oval object and the round one (circles are a special case of ovals, after all).

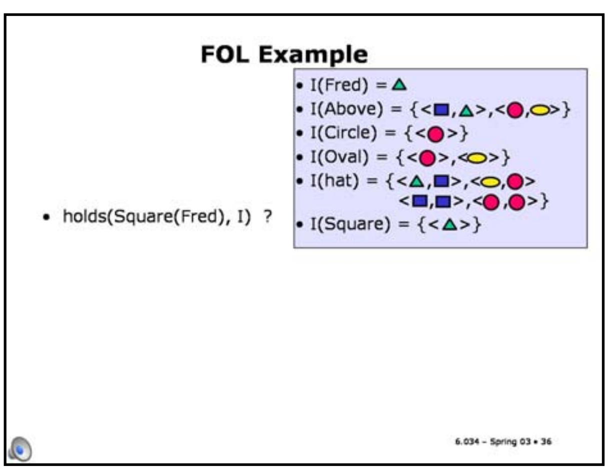




Slide 9.5.34
And we'll say that the hat of the triangle is the square and the hat of the oval is the circle. If we stopped at this point, we would have a function, but it wouldn't be total (it wouldn't have an entry for every possible first argument). So, we'll make it total by saying that the square's hat is the square and the circle's hat is the circle.

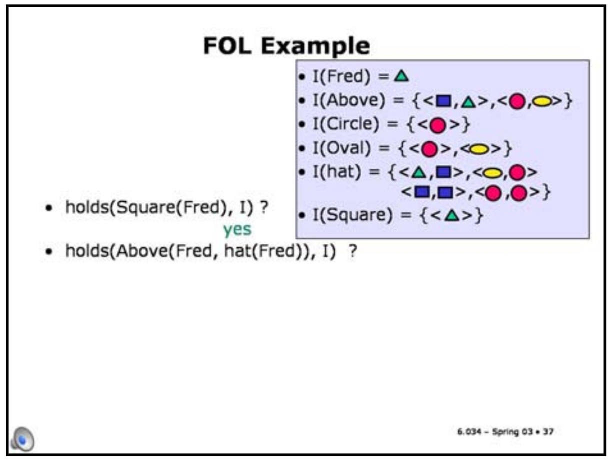
Slide 9.5.35
Finally, just to cause trouble, we'll interpret the predicate **Square** to be true of the triangular object.





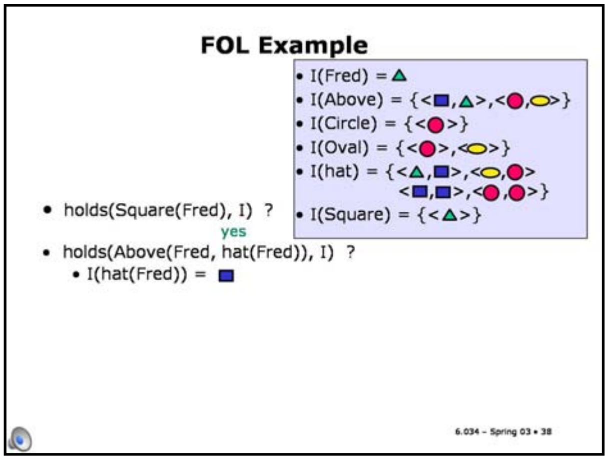
Slide 9.5.36
Now, let's find the truth values of some sentences in this interpretation. What about **Square(Fred)**, is that true in this interpretation? Yes. We look to see that **Fred** denotes the triangle, and then we look for the triangle in the relation denoted by square, and we find it there. So the sentence is true.

Slide 9.5.37
What about this one? Is **Fred** above its hat?

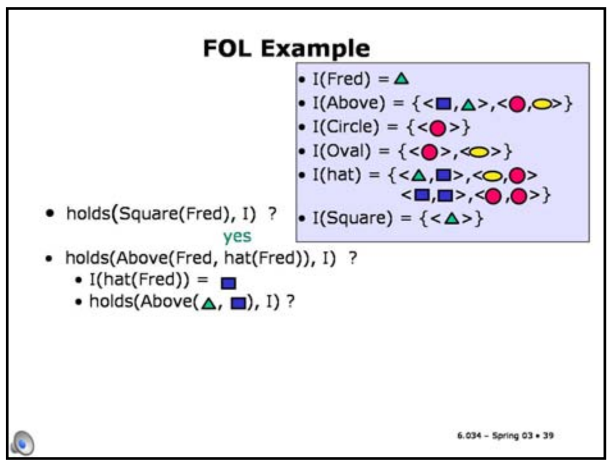


Slide 9.5.38
Let's start by asking the question, what's the denotation of the term, **hat(Fred)**?

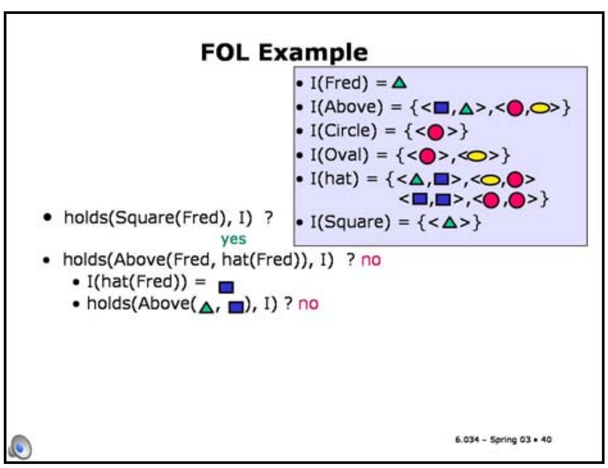
It's the square, right? We look up **Fred**, and get the triangle. Then we look in the **hat** function, and, sure enough, there's a pair with triangle first and square second. So **hat(Fred)** is a square.



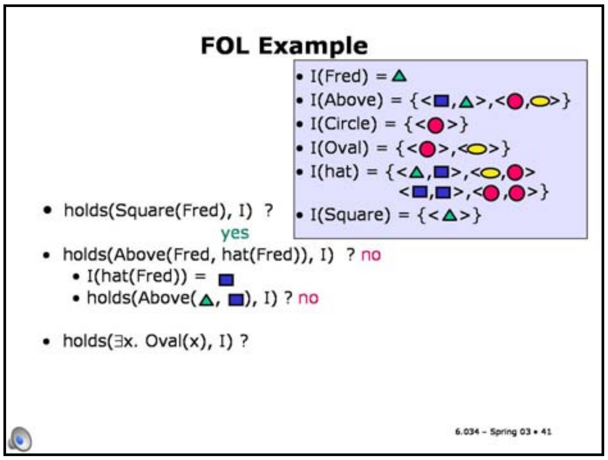
Slide 9.5.39
Now the question is: does the **Above** relation hold of the triangle and the square? We look this pair up in the relation denoted by **Above**, and we can't find it. So the **Above** relation doesn't hold of these objects.

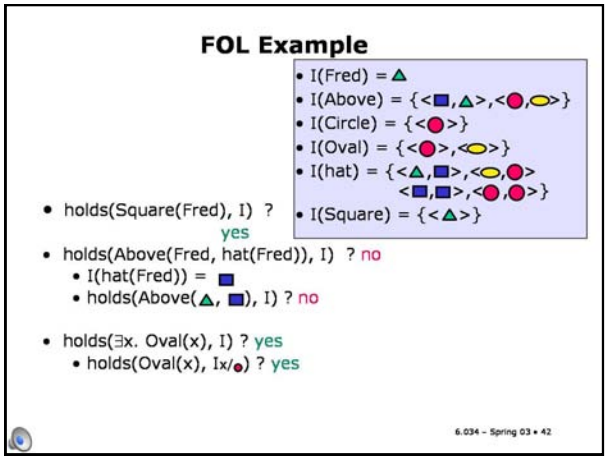


Slide 9.5.40
And our original sentence is false.



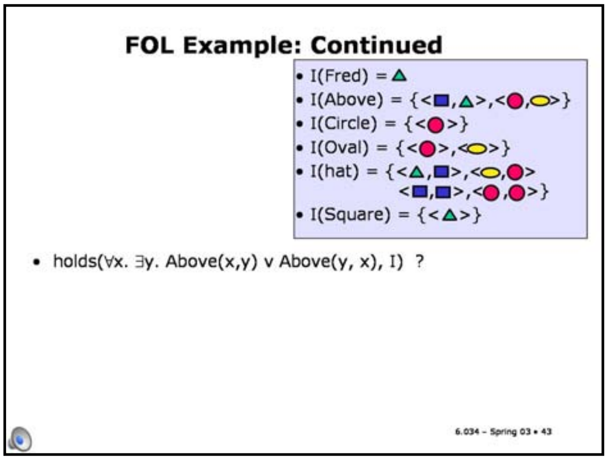
Slide 9.5.41
Okay. What about this sentence: there exists an **x** such that **Oval(x)**. Is there a thing that is an oval? Yes. So how do we show that carefully?

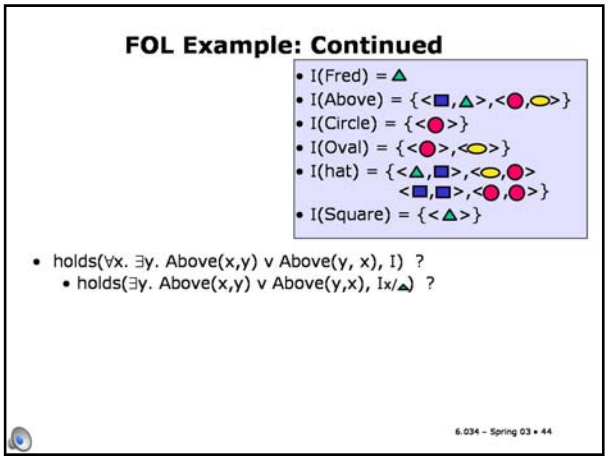




Slide 9.5.42
We say that there's an extension of this interpretation where we take **x** and substitute in for it, the circle. Temporarily, I say that **I(x)** is a circle. And now I ask, in that new interpretation, is it true that **Oval(x)**. So I look up **x** and I get the circle. I look up **Oval** and I get the relation with the circle and the oval, and so the answer's yes.

Slide 9.5.43
Here's a more complicated question in the same domain and interpretation. Is the sentence: **For all x there exists a y such that either x is Above y or y is Above x** true in **I**?

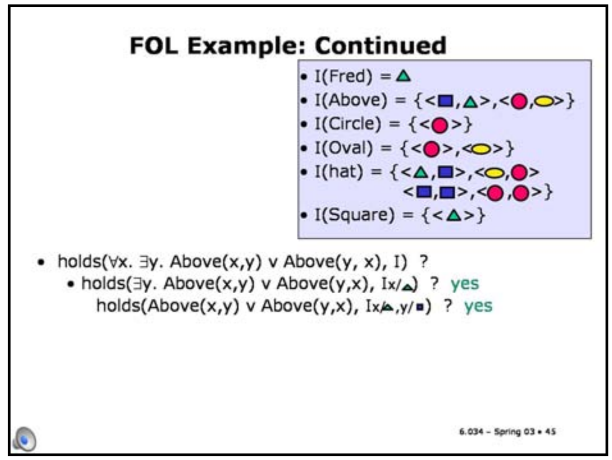


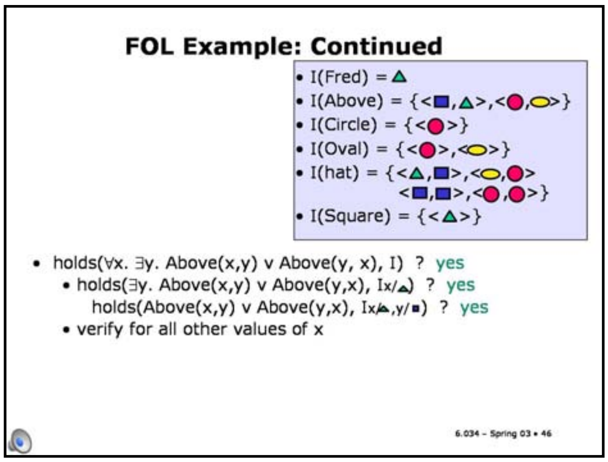


Slide 9.5.44
We can tell whether this is true by going through every possible object in the universe and binding it to the variable **x**, and then seeing whether the rest of the sentence is true. So, for example, we might put in the triangle for **x**, just to start with.

Slide 9.5.45

Now, having made that binding, we have to ask whether the sentence "There exists a y such that either x is above y or y is above x" true in the new interpretation. Existentials are easier than universals; we just have to come up with one y that makes the sentence true. And we can; if we bind **y** to the square, then that makes **Above(y,x)** true, which makes the disjunction true. So, we've proved this existential statement is true.



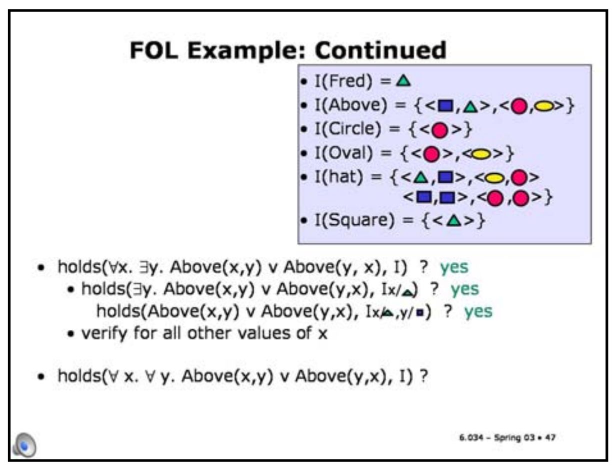


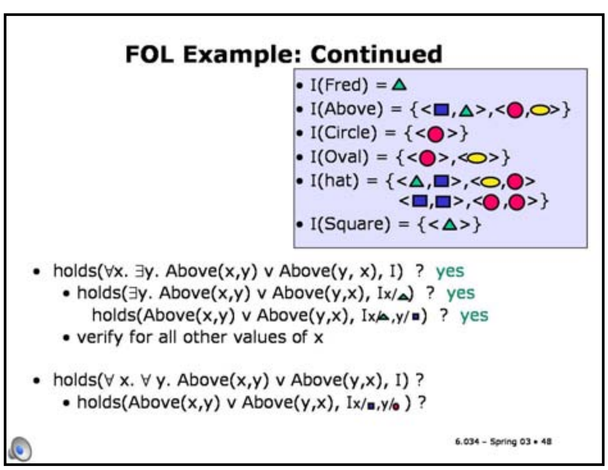
Slide 9.5.46

If we can do that for every other binding of **x**, then the whole universal sentence is true. You can verify that it is, in fact, true, by finding the truth value of the sentence with the other objects substituted in for **x**.

Slide 9.5.47

Okay. Here's our last example in this domain. What about the sentence: "for all x, for all y, x is above y or y is above x"? Is it true in interpretation **I**?



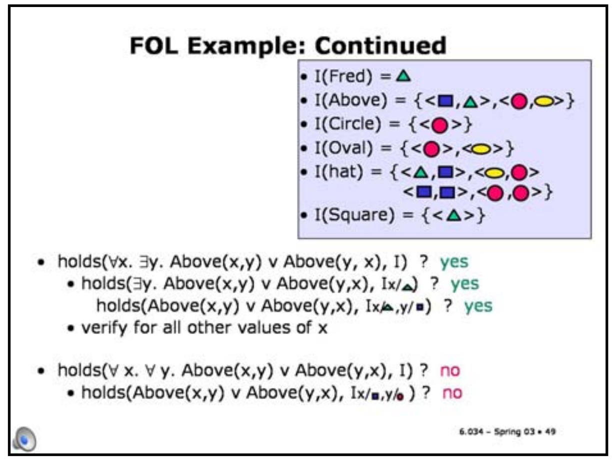


Slide 9.5.48

If it's going to be true, then it has to be true for every possible instantiation of **x** and y to elements of **U**. So, what, in particular, about the case when **x** is the square and **y** is the circle?

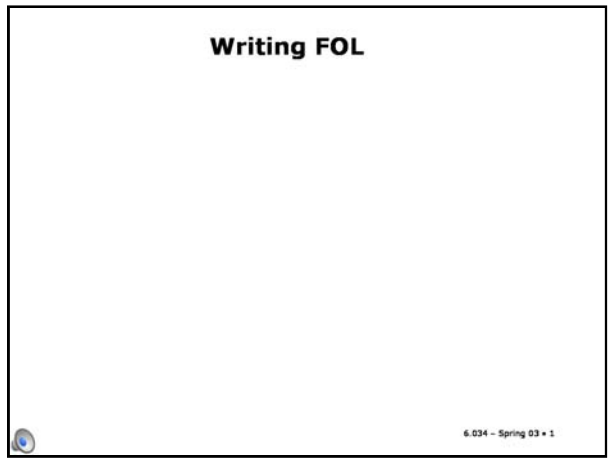
Slide 9.5.49
We can't find either the pair (square, circle), or the pair (circle, square) in the above relation, so this statement isn't true.

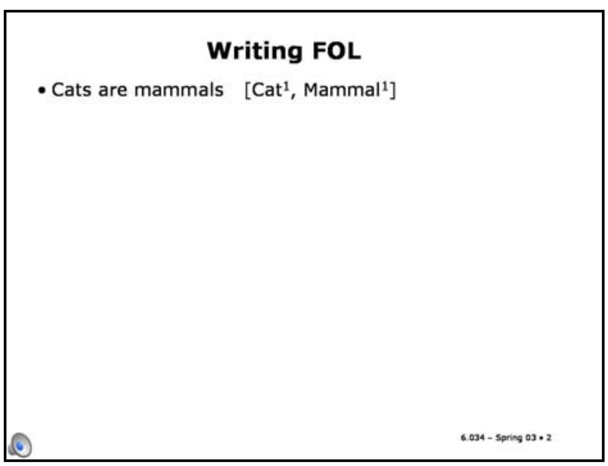
And, therefore, neither is the universally quantified statement.



6.034 Notes: Section 9.6

Slide 9.6.1
Now we're going to see how first-order logic can be used to formalize a variety of real-world concepts and situations. In this batch of problems, you should try to think of the answer before you go on to see it.

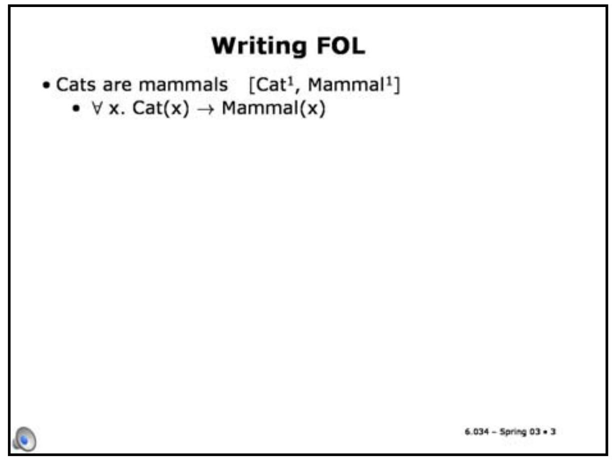


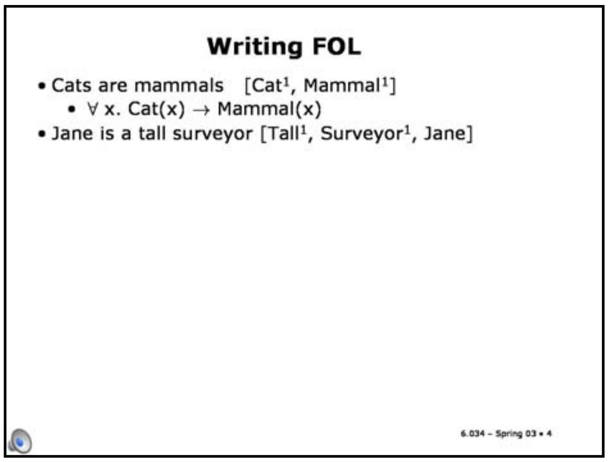


Slide 9.6.2
How would you use first-order logic to say "Cats are mammals"? (You can use a unary predicate **cat** and another unary predicate **mammal**).

Slide 9.6.3

For all x, Cat(x) implies Mammal(x). This is saying that every individual in the cat relation is also in the mammal relation. Or that cats are a subset of mammals.



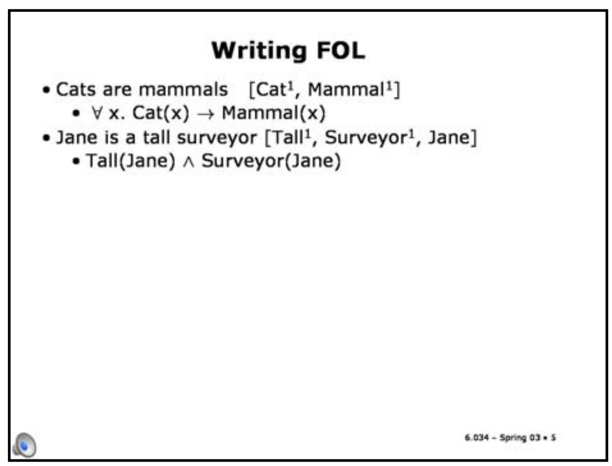


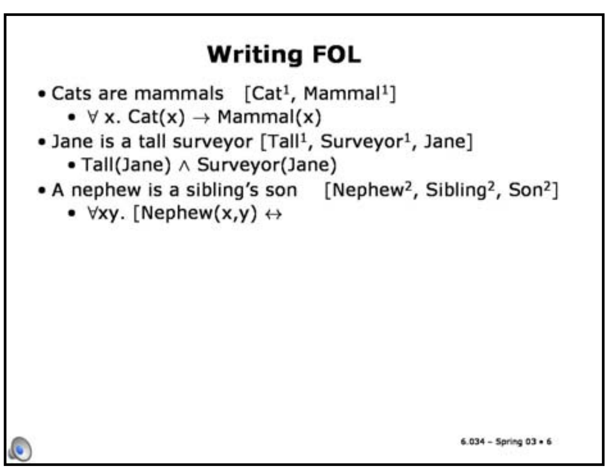
Slide 9.6.4

All right. Let's let **Jane** be a constant, **Tall** and **Surveyor** can be unary predicates. How can we say Jane is a tall surveyor?

Slide 9.6.5

Surveyor(Jane) and Tall(Jane).

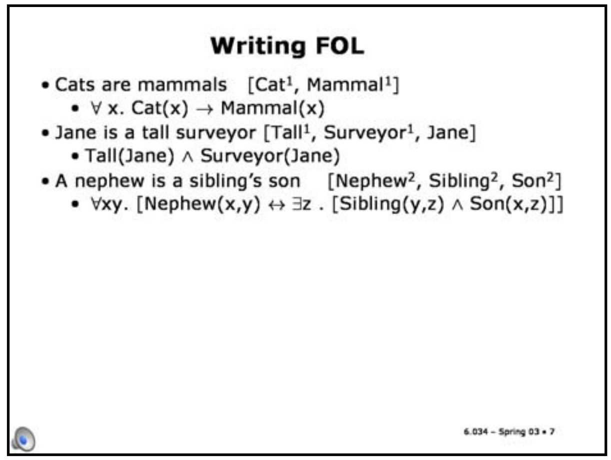


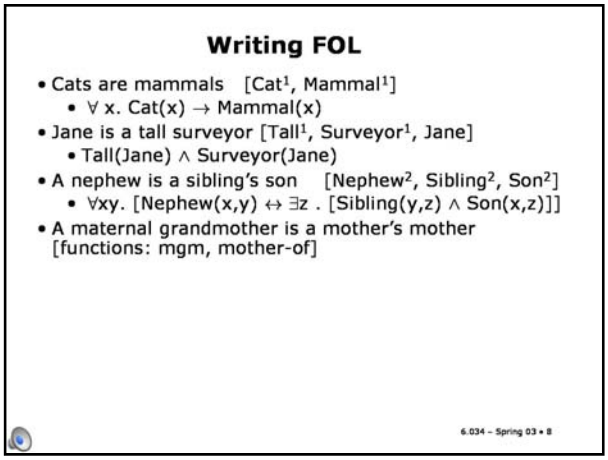


Slide 9.6.6

A nephew is a sibling's son. **Nephew**, **Sibling**, and **Son** are all binary relations. I'll start you off and say **for all x and y, x is the nephew of y if and only if** something. In English, what relationship has to hold between x and y for x to be a nephew of y? There has to be another person z who is a sibling of y and x has to be the son of z.

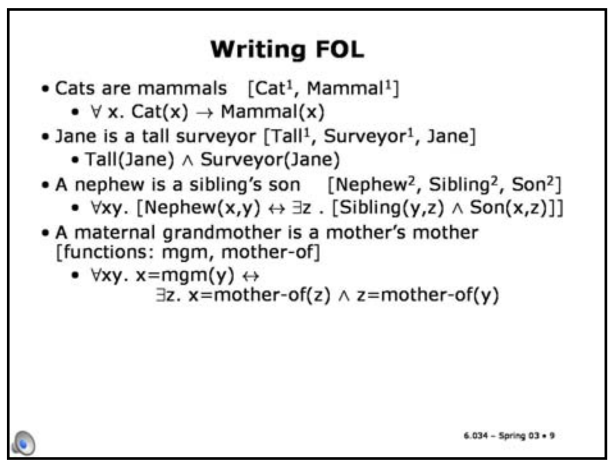
Slide 9.6.7
So, the answer is, "for all x and y, x is the nephew of y if and only if there exists a z such that y is a sibling of z and x is a son of z.

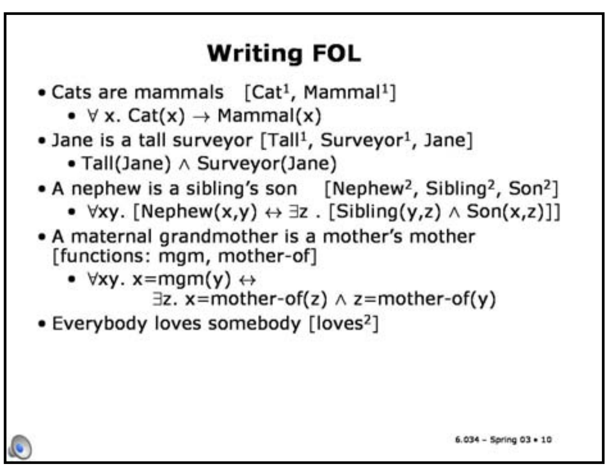




Slide 9.6.8
When you have relationships that are functional, like "mother of", and "maternal grandmother of", then you can write expressions using functions and equality. So, what's the logical way of saying that someone's maternal grandmother is their mother's mother? Use **mgm**, standing for maternal grandmother, and **mother-of**, each of which is a function of a single argument.

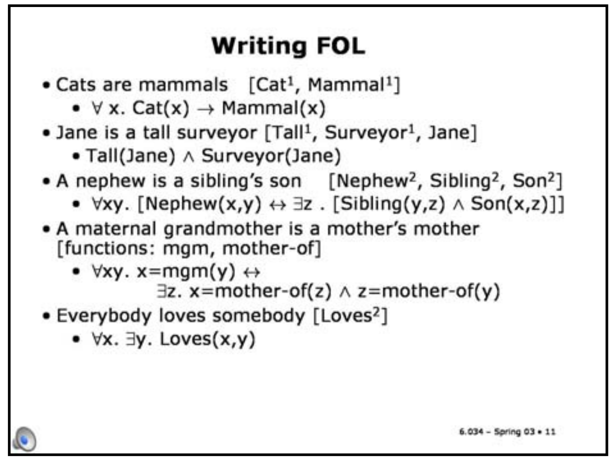
Slide 9.6.9
We can say that, "for all x and y, x is the maternal grandmother of y if and only if there exists a z such that x is the mother of z, and z is the mother of y".

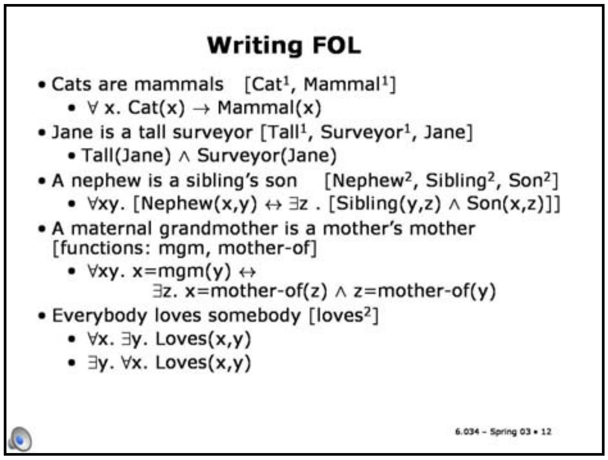




Slide 9.6.10
Using a binary predicate **Loves(x,y)**, how can you say that everybody loves somebody?

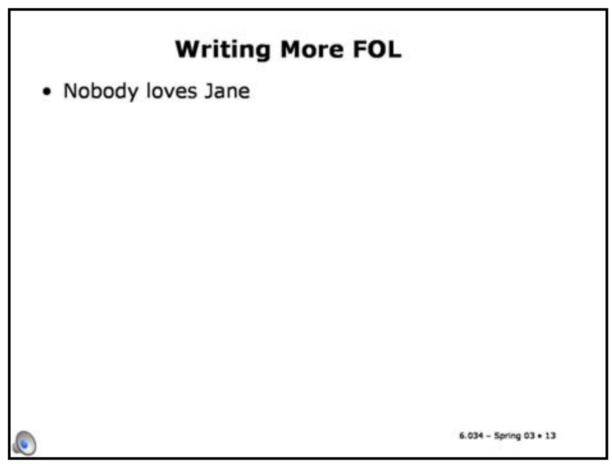
Slide 9.6.11
This one's fun, because there are really two answers. The usual answer is **for all x, there exists a y such that Loves(x,y)**. So, for each person, there is someone that they love. The loved one can be different for each lover. The other interpretation is that there is a particular person that everybody loves. How would we say that?

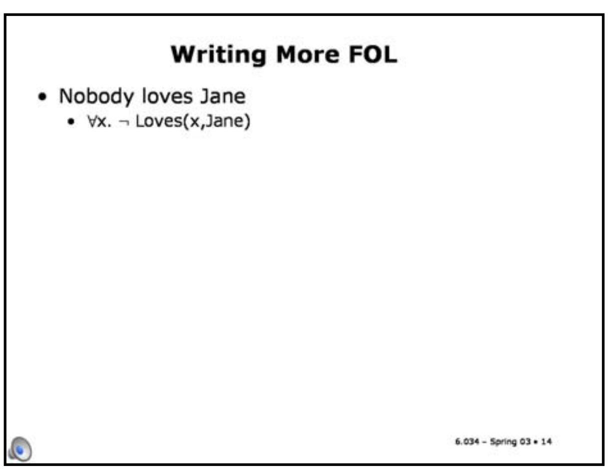




Slide 9.6.12
There exists a y such that for all x, Loves(x,y). So, just by changing the order of the quantifiers, we get a very different meaning.

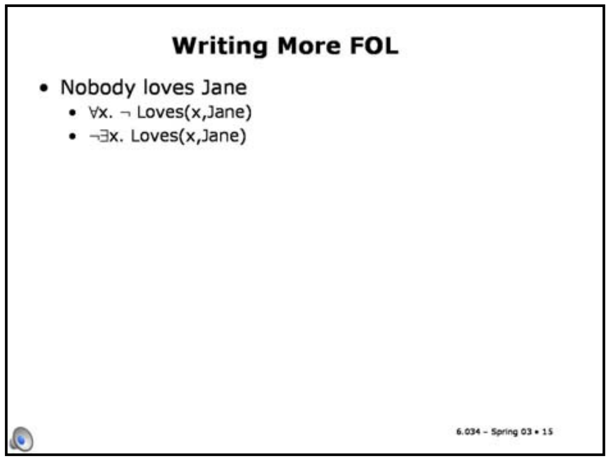
Slide 9.6.13
Let's say nobody loves Jane. Poor Jane. How can we say that?

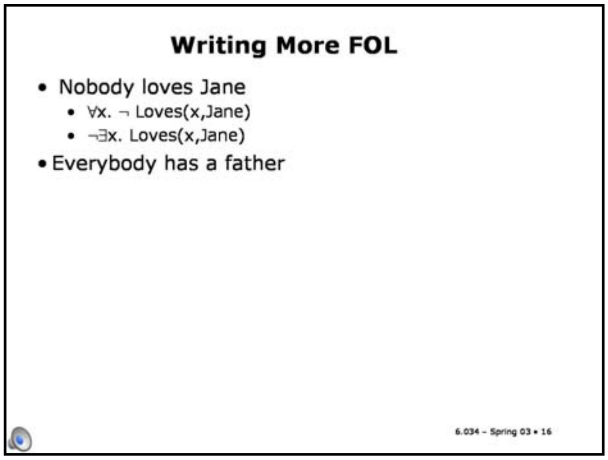




Slide 9.6.14
For all x, not Loves(x, Jane). So, for everybody, every single person, that person doesn't love Jane.

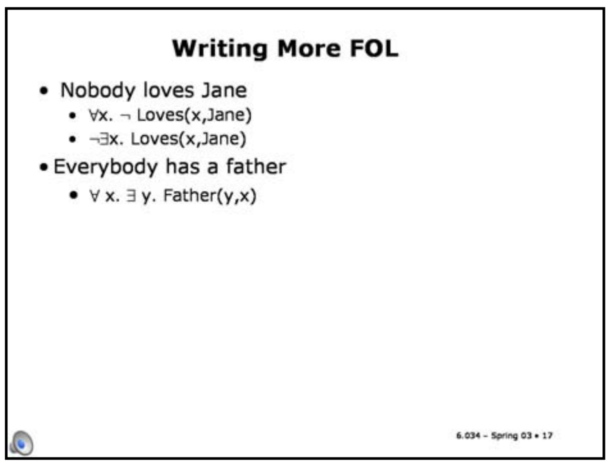
Slide 9.6.15
An equivalent thing to write is there does not exist an **x** such that **Loves(x, Jane)**. This is a general transformation, if you have for all x not something, then it's the same as having not there exists an x something. It's like saying I can't find a single x such that x Loves Jane.

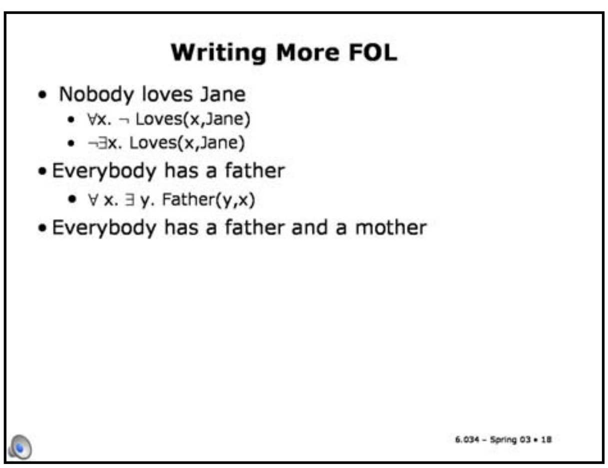




Slide 9.6.16
Everybody has a father.

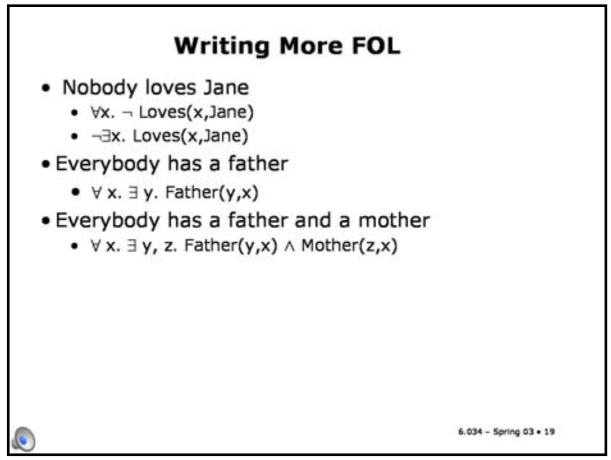
Slide 9.6.17
For all x, exists y such that Father(y,x)

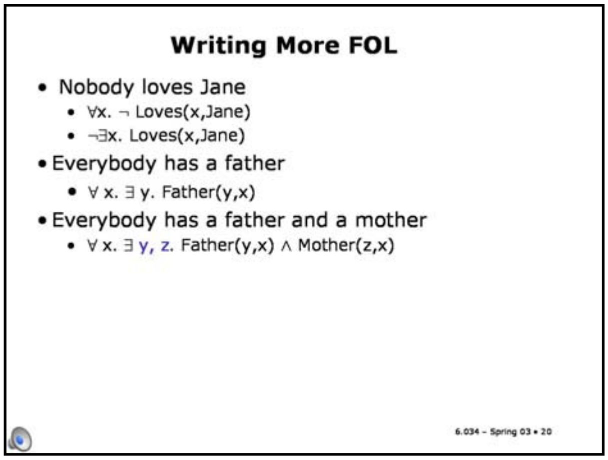




Slide 9.6.18
Everybody has a father and a mother.

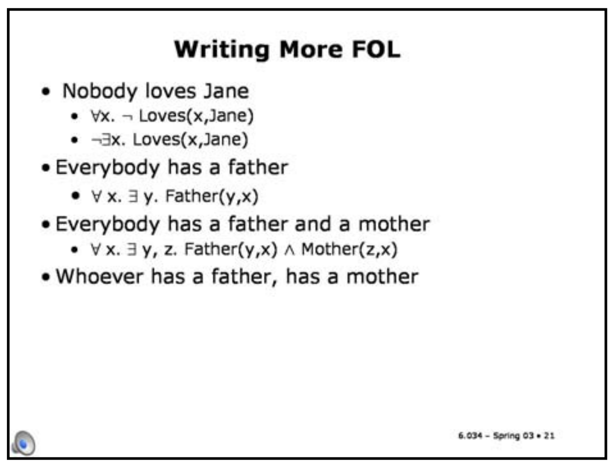
Slide 9.6.19
For all x, exists y, z such that Father(y,x) and Mother(z,x)

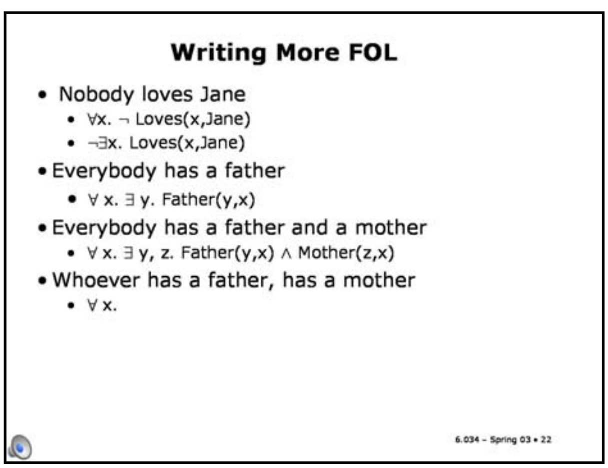




Slide 9.6.20
Now, you might ask whether **y** and **z** are necessarily different. The answer is, no, that's not enforced by the logic. For that matter, they could be the same as **x**. Now, if you use the typical definitions of father and mother, they won't be the same, but that's up to the interpretation.

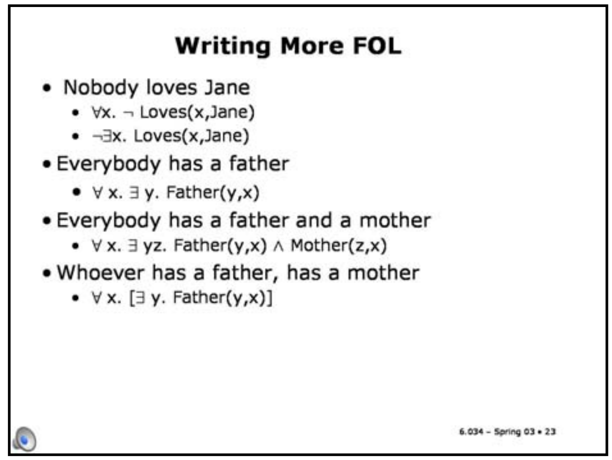
Slide 9.6.21
Whoever has a father has a mother.

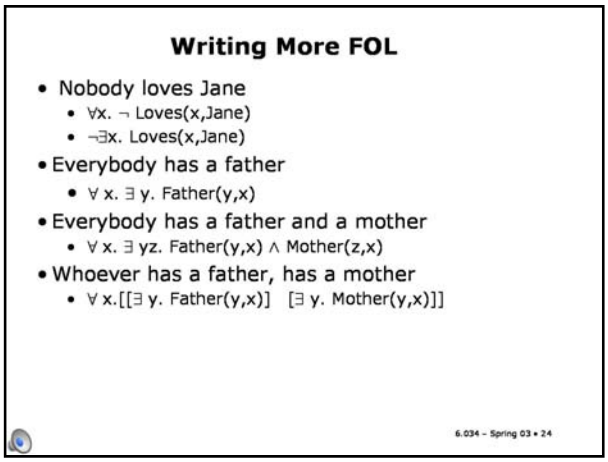




Slide 9.6.22
This is a general statement about objects of the kind, everything that has one property has another property. All right? So we'll talk about everything by starting with **forall x**.

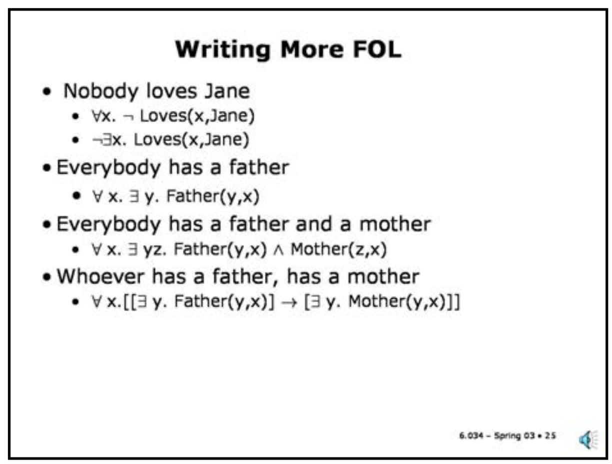
Slide 9.6.23
Now, how do we describe x's that have a father? **Exists y such that Father(y,x)**.

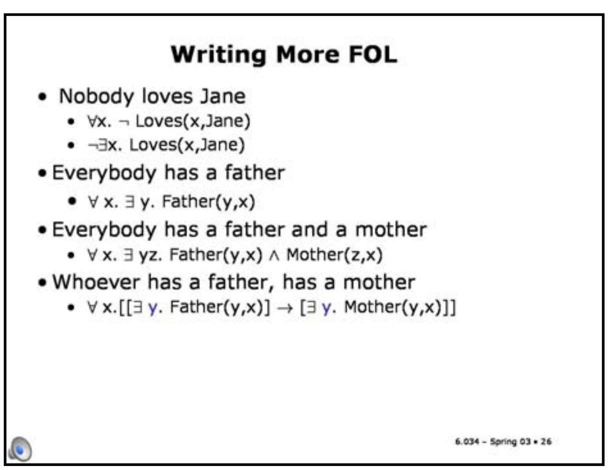




Slide 9.6.24
And we can describe x's that have a mother by **exists y such that Mother (y,x)**.

Slide 9.6.25
Finally, we put these together using implication, just as we did with the "all cats are mammals" example. We want to say objects with a Father are a subset of the set of objects with a Mother (in this case, it will turn out that the sets are equal). So, we end up with "for all x, if there exists a y such that y is the father of x, then there exists a y such that y is the mother of x".

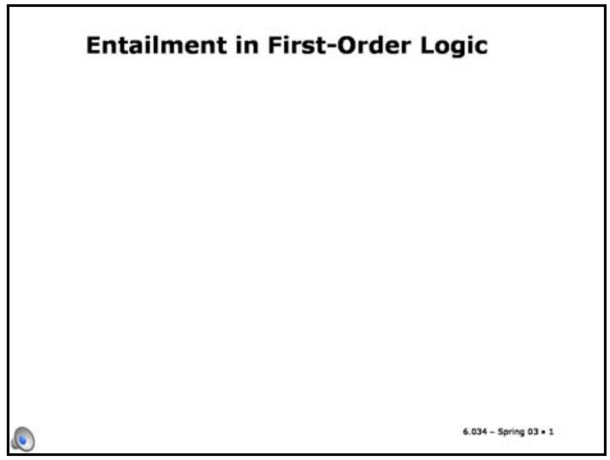


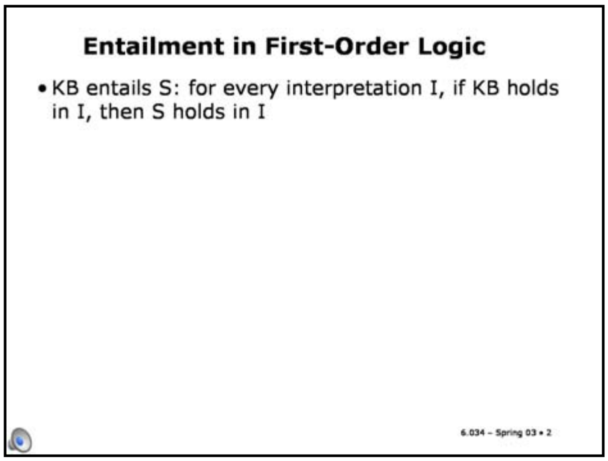


Slide 9.6.26
Note that the two variables named y have separate scopes, and are entirely unrelated to one another. You could rename either or both of them and the semantics of the sentence would remain the same. It's technically legal to have nested quantifiers over the same variable, and there are rules for figuring out what it means, but it's very confusing, so it's just better not to do it.

6.034 Notes: Section 9.7

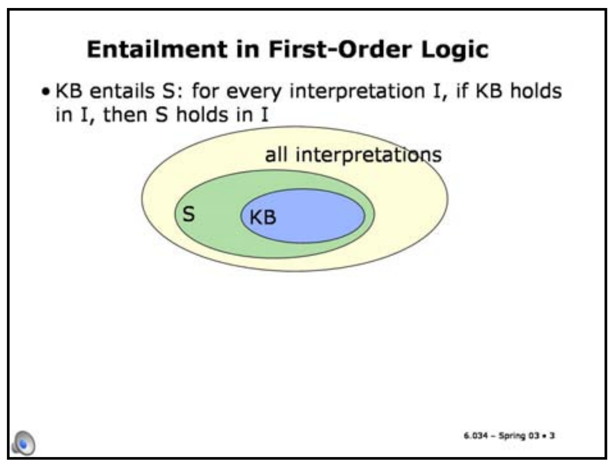
Slide 9.7.1
Now that we understand something about first-order logic as a language, we'll talk about how we can use it to do things. As in propositional logic, the thing that we'll most often want to do with logical statements is to figure out what conclusions we can draw from a set of assumptions. In propositional logic, we had the notion of entailment: a **KB** entails a sentence if and only if the sentence is true in every interpretation that makes **KB** true.

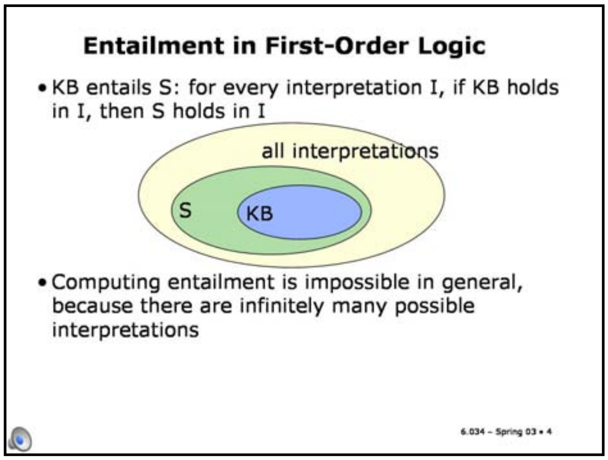




Slide 9.7.2
In first-order logic, the notion of entailment is the same. A knowledge base entails a sentence if and only if the sentence holds in every interpretation in which the knowledge base holds.

Slide 9.7.3
It's important that entailment is a relationship between a set of sentences, **KB**, and another sentence, **S**. It doesn't directly involve a particular intended interpretation that we might have in mind. It has to do with the subsets of all possible interpretations in which **KB** and **S** hold; entailment requires that the set of interpretations in which **KB** holds be a subset of those in which **S** holds. This is sort of a hard thing to understand at first, since the number (and potential weirdness) of all possible interpretations in first-order logic is just huge.



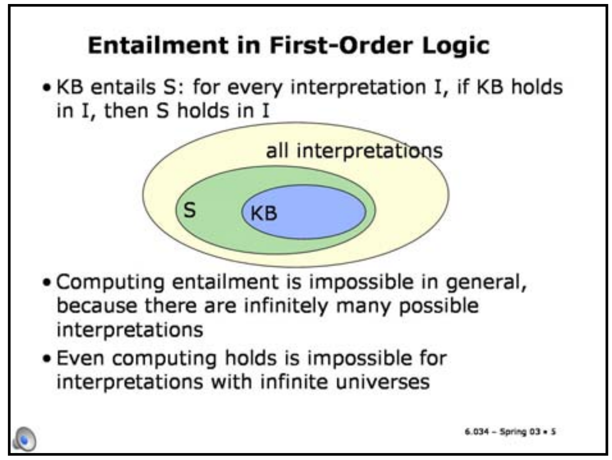


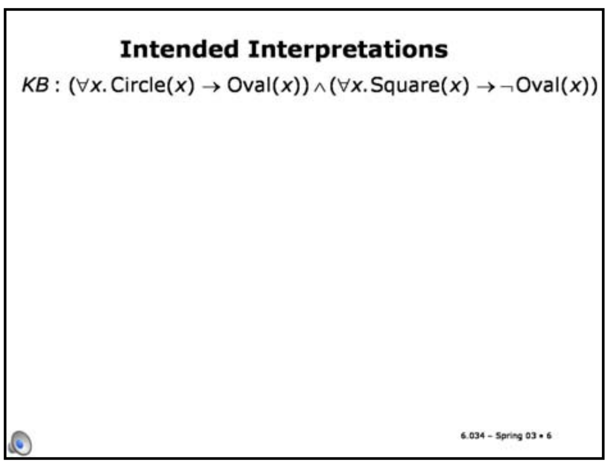
Slide 9.7.4
In propositional logic, we were able to think about computing entailment by doing a brute-force enumeration of all intepretations, then, in each interpretation, checking to see whether the sentence and/or the knowledge base were true in that interpretation.

This will be impossible in first-order logic for two reasons.

First, it's completely out of the question to enumerate all possible interpretations. How many universes are there? More than *I* can count...

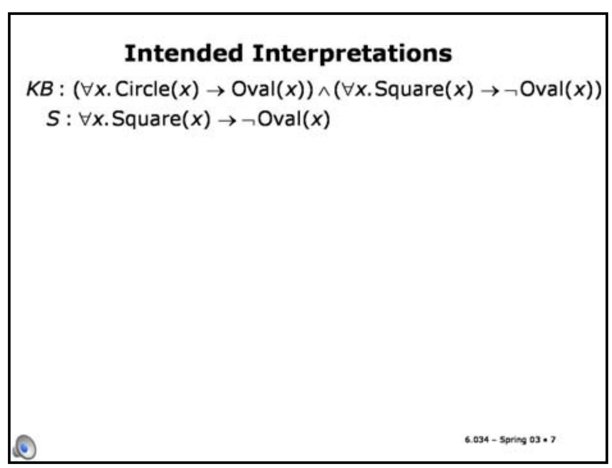
Slide 9.7.5
Second, even for a single interpretation, it's not necessarily possible to compute whether a sentence holds in that interpretation. Why? Because if it has a universal or existential quantifier, it will require testing whether a sentence holds for every substitution of an element in **U** for the quantified variable. And if the universe is infinite, you just can't do that.

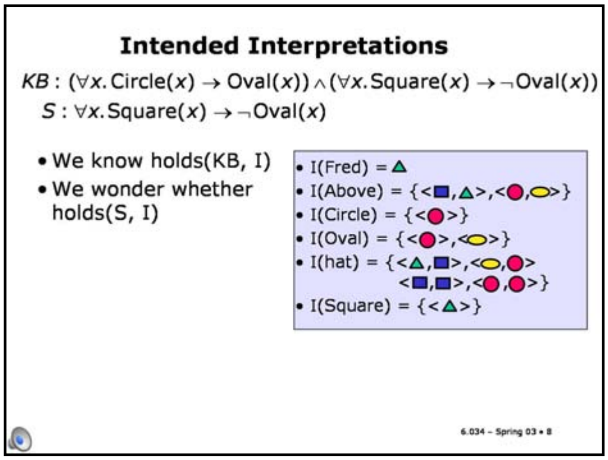




Slide 9.7.6
Let's look at a particular situation in which we might want to do logical inference. Consider our shapes example from before. Let's say that we know, as our knowledge base, that all circles are ovals, and that no squares are ovals. We can write this as **for all x, Circle(x) implies Oval(x)**. And **for all x, Square (x) implies not Oval(x)**.

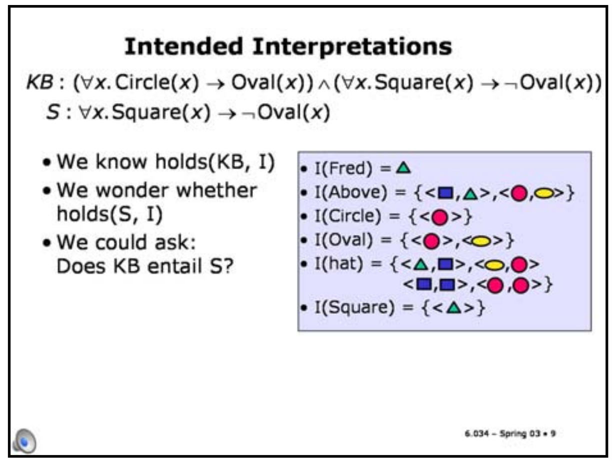
Slide 9.7.7
Now, let's say we're wondering whether it's also true that no squares are circles. We'll call that sentence **S**, and write it **for all x, Square(x) implies not Circle(x)**.

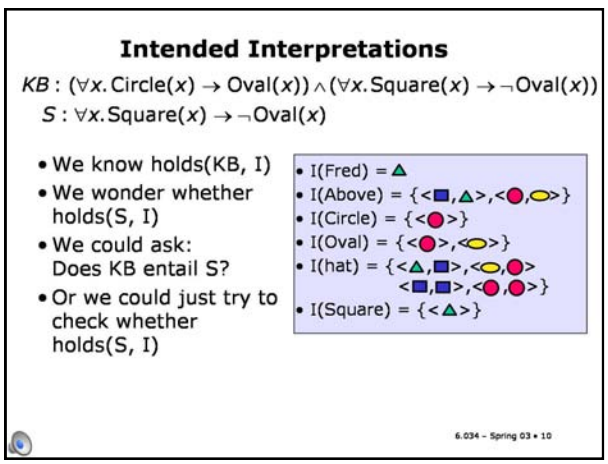




Slide 9.7.8
We know **KB** holds in interpretation **I**, and we wonder whether **S** holds in **I**.

Slide 9.7.9
We could answer this by asking the question: Does **KB** entail **S**? Does our desired conclusion follow from our assumptions.



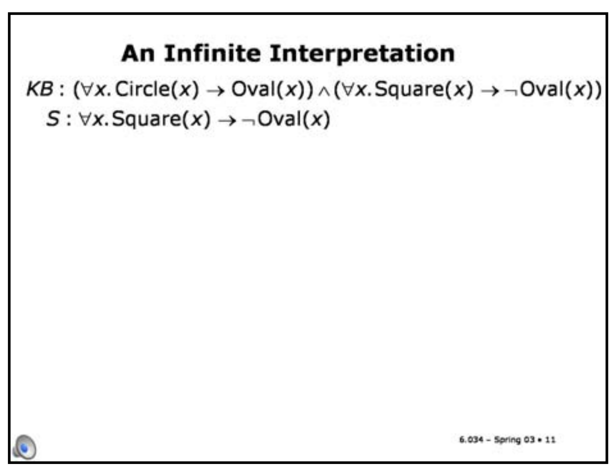


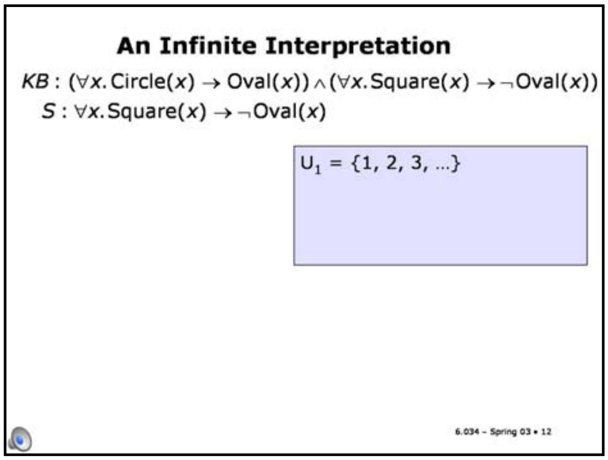
Slide 9.7.10
You might say that entailment is too big a hammer. I don't actually care whether **S** is true in all possible interpretations that satisfy **KB**. Why? because I have a particular interpretation in mind (namely, our little world of geometric shapes, embodied in interpretation **I**). And I know that **KB** holds in **I**. So what I really want to know is whether **S** holds in **I**.

Unfortunately, the computer does not know what interpretation I have in mind. We want the computer to be able to reach valid conclusions about my intended interpretation without my having to enumerate it (because it may be infinite).

For this particular example of **I**, it's not too hard to check whether **S** holds (because the universe is finite and small). But, as we said before, in general, we won't be able even to test whether a sentence holds in a particular interpetation.

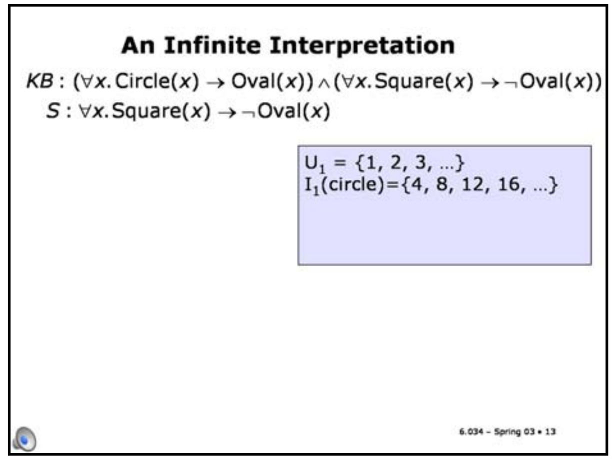
Slide 9.7.11
Let's look at our **KB** for a minute. When we wrote it down, we had a particular interpretation in mind, as evidenced by the names of the propositions. But now, here's another interpretation, $\mathbf{I_1}$

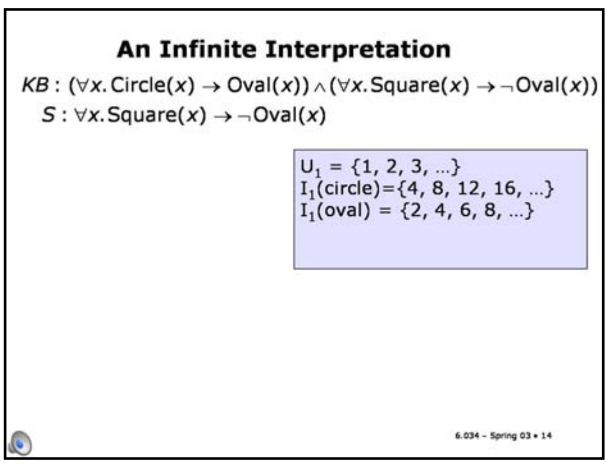




Slide 9.7.12
The universe is the positive integers (numbers 1, 2, etc.). This universe is clearly infinite.

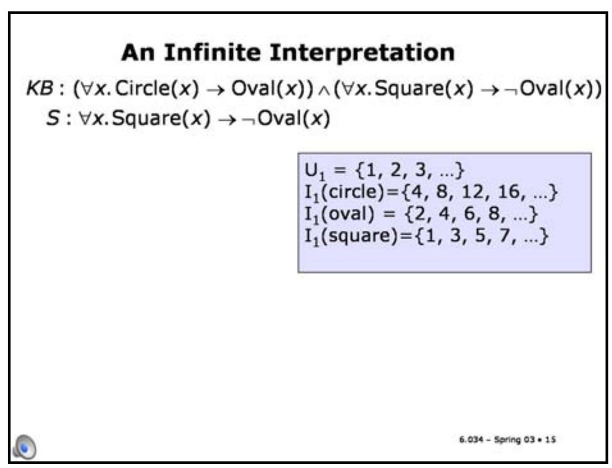
Slide 9.7.13
Let's let the circle relation stand for positive integers evenly divisible by 4. So it's the infinite set {4, 8, 12, 16, ...}.

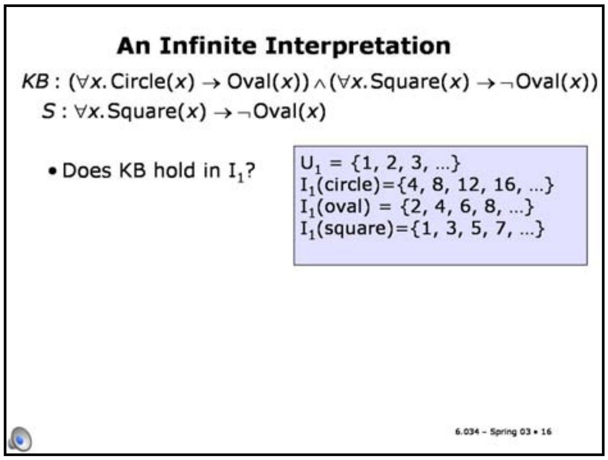




Slide 9.7.14
We'll let **Oval** stand for the even positive integers, {2, 4, 6, 8, ...}.

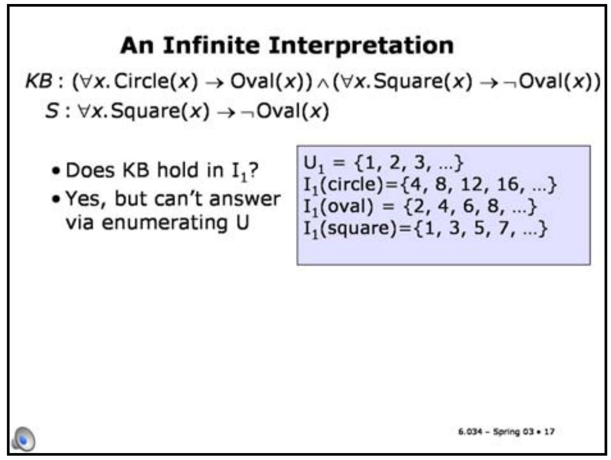
Slide 9.7.15
And we'll let **Square** stand for the odd positive integers, {1, 3, 5, 7, ...}.

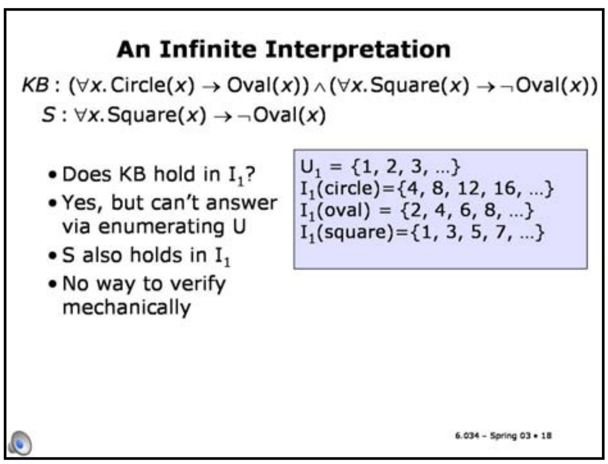




Slide 9.7.16
Now, does **KB** hold in $\mathbf{I_1}$?

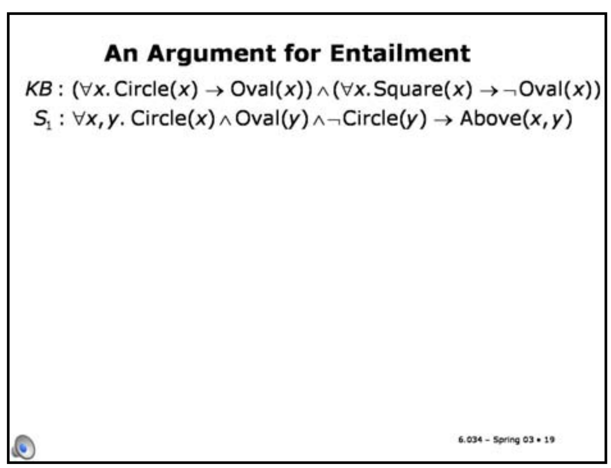
Slide 9.7.17
We can't verify that by enumerating **U** and checking the sentences inside the universal quantifier. However, we all know, due to basic math knowledge, that it does.

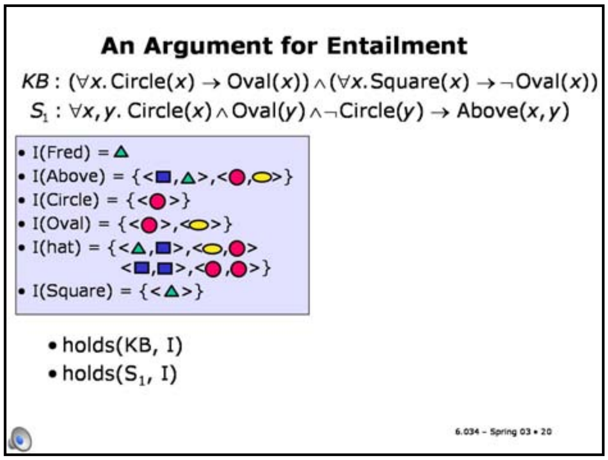




Slide 9.7.18
Similarly, *we* can see that **S** holds in $\mathbf{I_1}$, as well. Unfortunately, we can't rely on our computers to be as smart as we are (yet!). So, if we want a computer to arrive at the conclusion that **S** follows from **KB**, it will have to do it more mechanically.

Slide 9.7.19
Let's think about a different **S**, which we'll call $\mathbf{S_1}$: For every circle and every oval that is not a circle, the circle is above the oval.

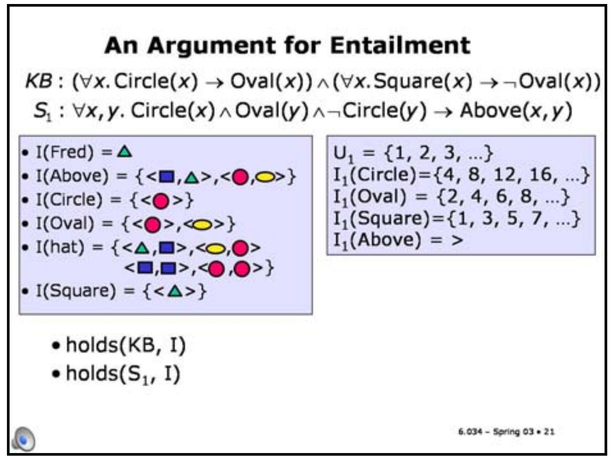


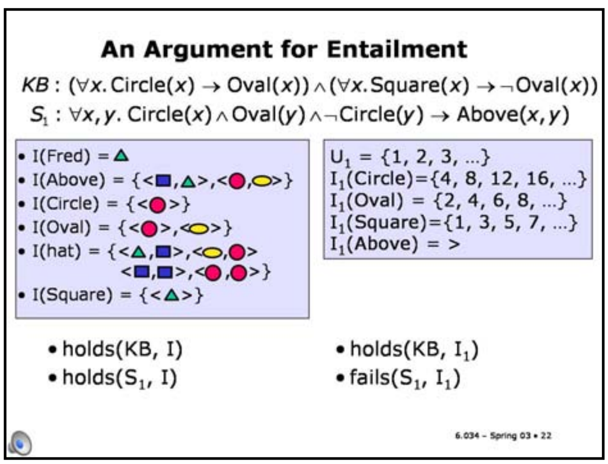


Slide 9.7.20
Back in I, our original geometric interpretation, this sentence holds, right?

But does it "follow from" **KB**? Is it entailed by **KB**?

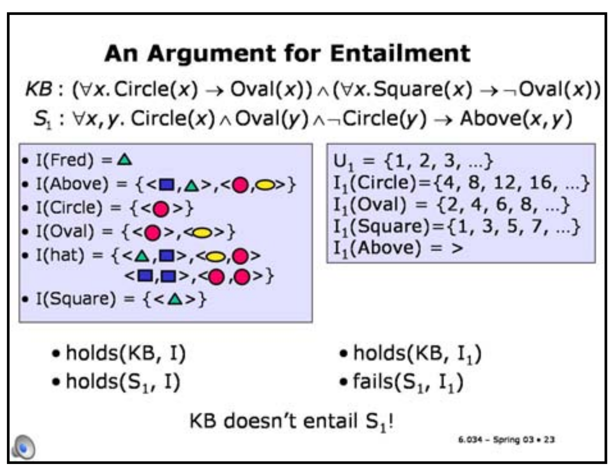
Slide 9.7.21
No. We can see this by going back to interpretation $\mathbf{I_1}$, and letting the interpretation of the "above" relation be greater-than on integers.

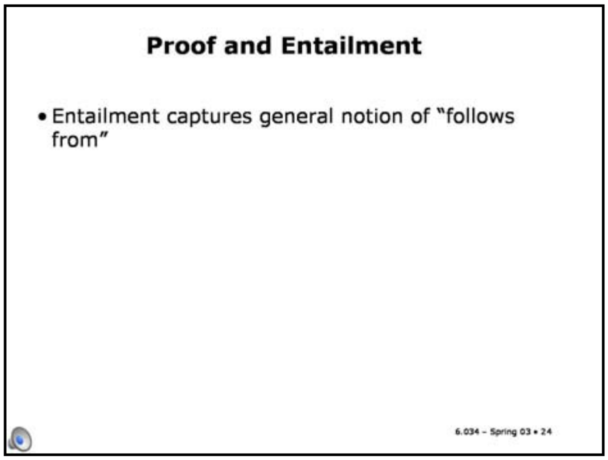




Slide 9.7.22
Then **S** holds in $\mathbf{I_1}$ if all integers divisible by 4 are greater than all integers divisible by 2 but not by 4, which is clearly false.

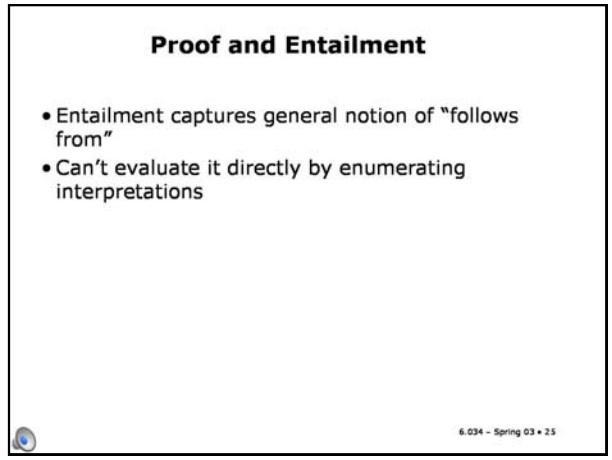
Slide 9.7.23
So, although **KB** and **S** both hold in our original intended interpretation I, **KB** does not entail **S**, because we can find an interpretation in which **KB** holds but **S** does not.

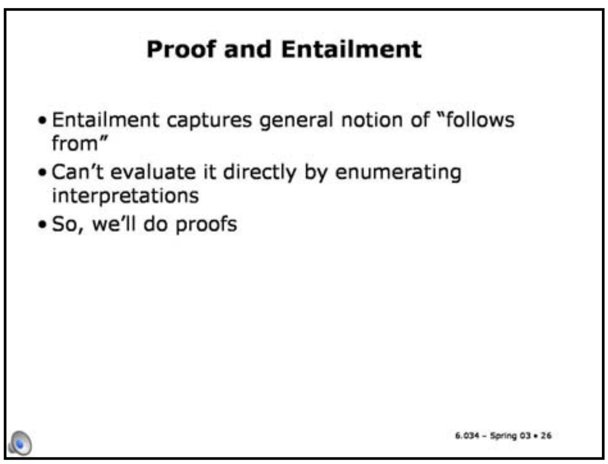




Slide 9.7.24
We can see from this example that entailment captures this general notion of a sentence following from a set of assumptions; of being able to justify the truth of **S** based only on the truth of **KB**.

Slide 9.7.25
So, we like the notion of entailment, but we can't compute it directly.

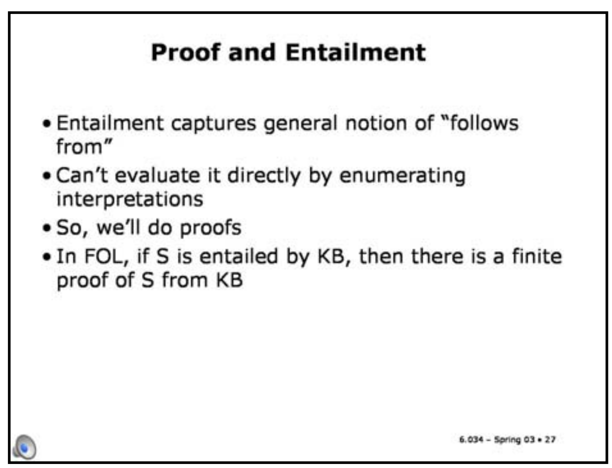


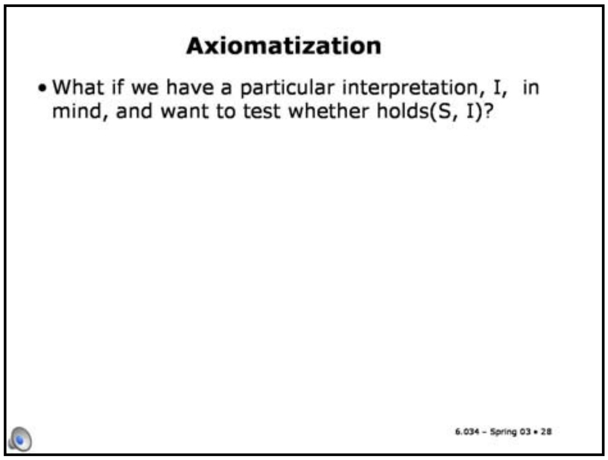


Slide 9.7.26
So what do we do? As we did in propositional logic, we will stay in the domain of syntax, and do proofs to figure out whether **S** is entailed by **KB**.

Slide 9.7.27
There are proof rules that are sound and complete, in the sense that if **S** is entailed by **KB**, there is a finite proof of that. So, it's easier, in general, though not for every particular case, to do a proof of general entailment than to test whether a sentence holds in a given interpretation.

The next few segements of this material will show how to extend the notion of resolution refutation from propositional logic to first-order logic.

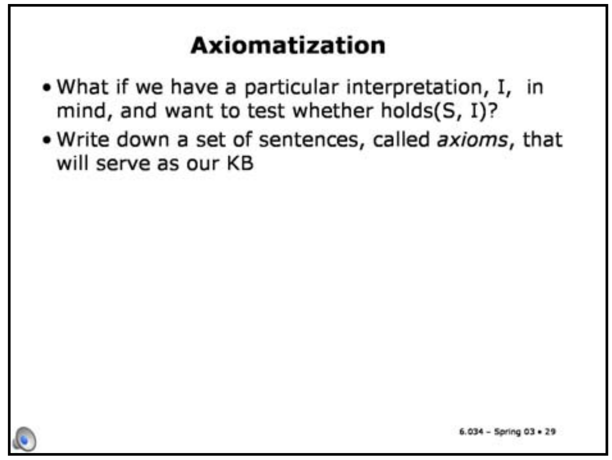


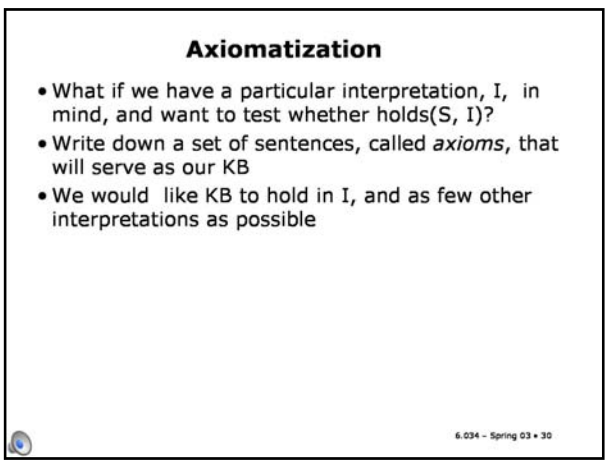


Slide 9.7.28
We just argued that entailment is the right notion when we want to ask the question whether a sentence **S** follows from a **KB**. And that we're going to show entailment via proof.

But what if we have a particular interpretation in mind? We've seen that we can't in general, test whether a sentence holds in that interpretation. How can we use the ability to use proof to show entailment, in order to test whether a sentence holds in an interpretation?

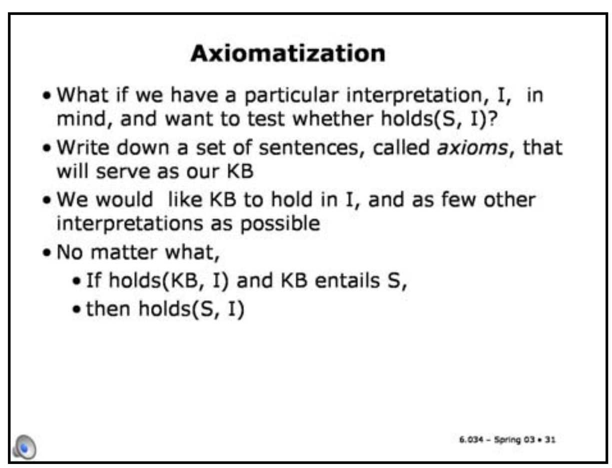
Slide 9.7.29
The answer is that we have to *axiomatize* our domain. That is, we have to write down a set of sentences, or axioms, which will serve as our **KB**.

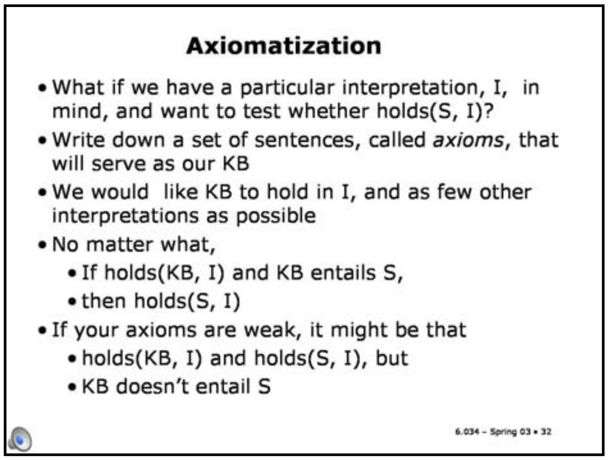




Slide 9.7.30
Ideally the axioms would be so specific that that there was a single interpretation, our intended interpretation, in which they held. In general, though, this will be impossible. You might be able to constrain your axioms to describe domains that contain exactly 4 objects, but you'll never be able to say exactly which 4. You can often give axioms that put stringent enough requirements on the relationships between those objects that all of the interpretations in which the axioms hold are essentially the same as (isomorphic to) your intended interpretation.

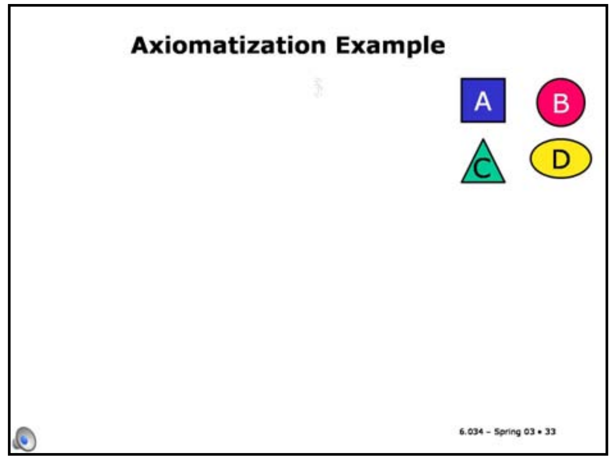
Slide 9.7.31
No matter how constraining your axioms are, you can rely on the fact that if your **KB** holds in your intended interpretation and **KB** entails **S**, then **S** holds in the intended interpretation.

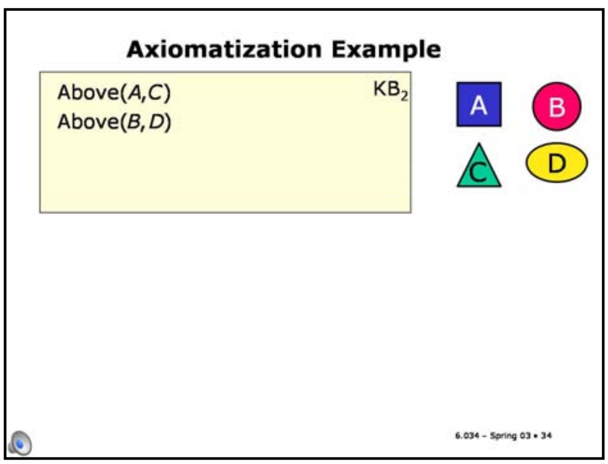




Slide 9.7.32
But that's only half of what we need. There might be some fact, **S**, about your intended interpretation that you would like to be able to derive from your axioms. But, if your axioms are not specific enough, then they might admit some interpretations in which **S** does not hold, and in that case, the axioms will not entail **S**, even though it might hold in the intended interpretation.

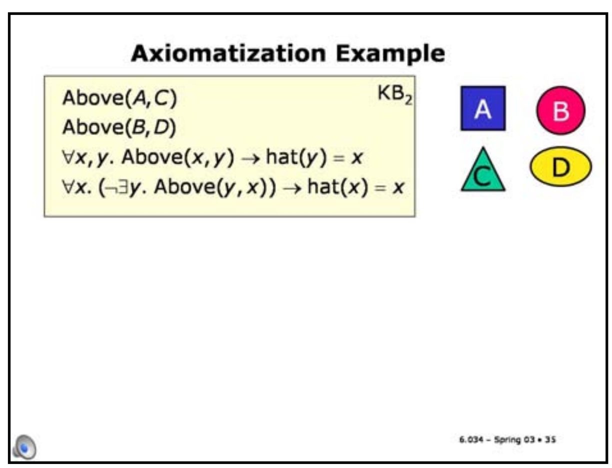
Slide 9.7.33
Let's work through an example of axiomatizing a domain. We'll think about our good old geometric domain, but, to simplify matters a bit, let's assume that we have the constant symbols **A**, **B**, **C**, and **D**. And let our interpretation specify that **A** is the square, **B** is the circle, **C** is the triangle, and **D** is the oval.

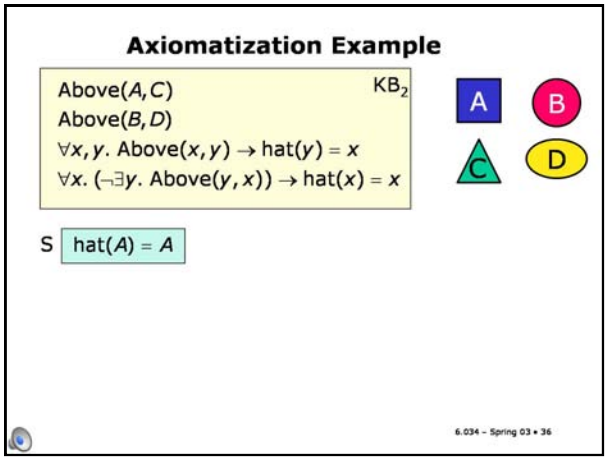




Slide 9.7.34
We propose to axiomatize this domain by specifying the above relation on these constants: **Above(A, C) and Above (B, D)**.

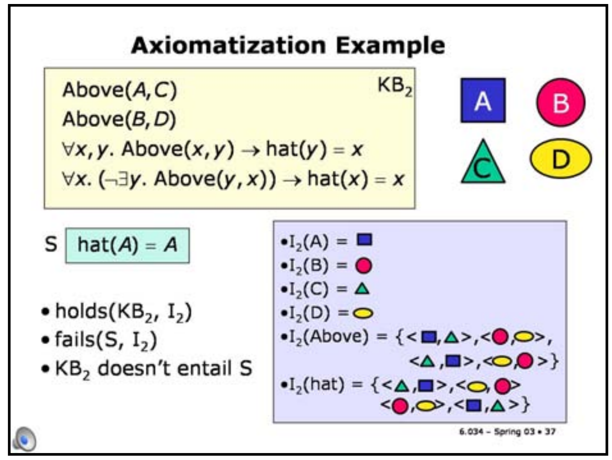
Slide 9.7.35
And we'll give some axioms that say how the hat function can be derived from **Above**: "for all x and y, if x is above y, then hat of y equals x; and for all x, if there is no y such that y is above x, then hat of x equals x".

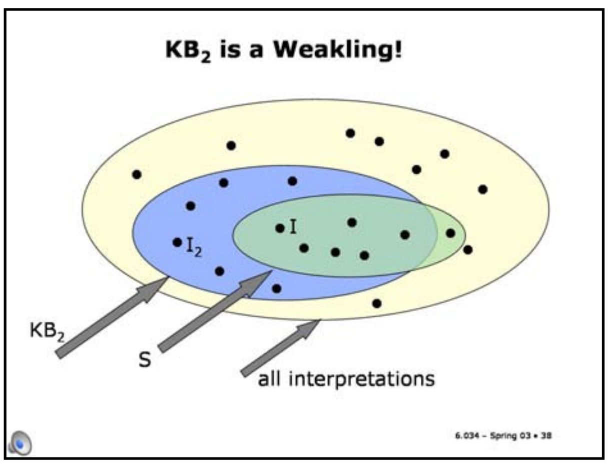




Slide 9.7.36
These four axioms will constitute our **KB**. Now, we're curious to know whether it's okay to conclude that the hat of **A** is **A**. It's true in our intended interpretation, and we'd like it to be a consequence of our axioms.

Slide 9.7.37
So, does our **KB** entail **S**? Unfortunately not. Consider the interpretation $\mathbf{I_2}$. It has two extra pairs in the interpretation of Above. Our axioms definitely hold in this interpretation, but **S** does not. In fact, in this interpretation, the sentence **hat(A) = C** will hold.

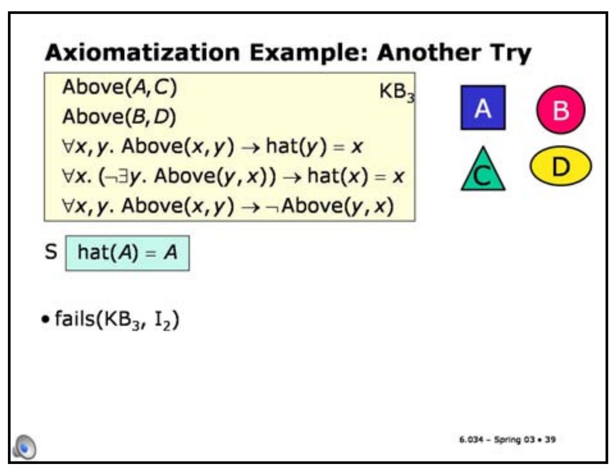


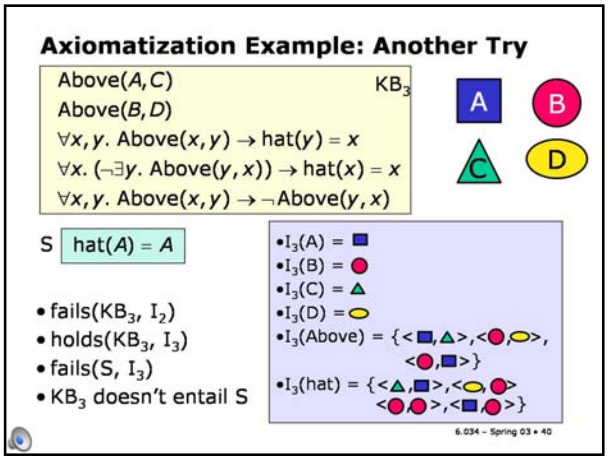


Slide 9.7.38
Just so we can see what's going on, let's go back to our Venn diagram for entailment. In this case, the blue set of interpretations in which the **KB** holds is not a subset of the green set of interpretations in which **S** holds. So, it is possible to have an interpretation, $\mathbf{I_2}$, in which **KB** holds but not **S**. **KB** does not entail **S** (for it to do so, the blue area would have to be a subset of the green), and so we are not licensed to conclude **S** from **KB**.

How can we fix this problem? We need to add more axioms, in order to rule out $\mathbf{I_2}$ as a possible interpretation. (Our goal is to make the blue area smaller, until it becomes a subset of the green area).

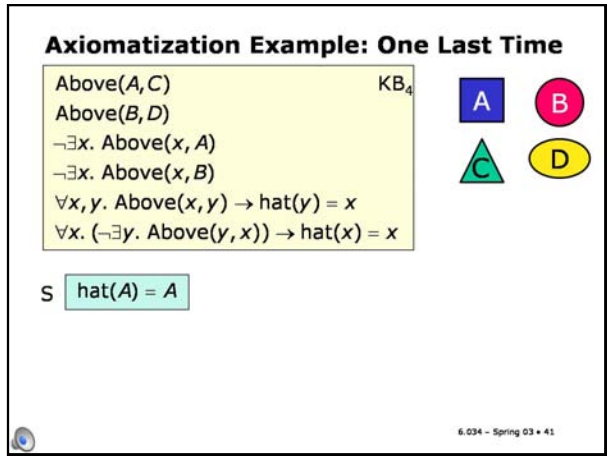
Slide 9.7.39
Here's a reasonable axiom to add: "for all x and y, if x is above y then y is not above x". It says that above is asymmetric. With this axiom added to our **KB**, **KB** no longer holds in $\mathbf{I_2}$, and so our immediate problem is solved.

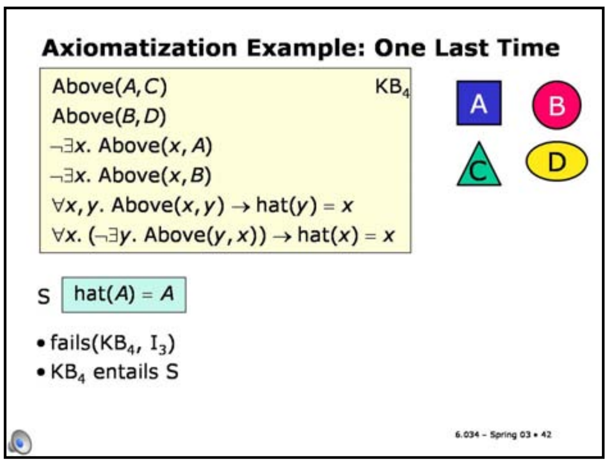




Slide 9.7.40
But we're not out of the woods yet. Now consider interpretation $\mathbf{I_3}$, in which the circle is above the square. **KB** holds in $\mathbf{I_3}$, but **S** does not. So **S** is still not entailed by $\mathbf{I_3}$.

Slide 9.7.41
Clearly, we're missing some important information about our domain. Let's add the following important piece of information to our set of axioms: there is nothing above the square or the circle.





Slide 9.7.42
If we let our new **KB** have these axioms as well, then it fails in $\mathbf{I_3}$, and does, in fact, entail **S**. Whew.

Slide 9.7.43
So, when you are axiomatizing a domain, it's important to be as specific as you can. You need to find a way to say everything that's crucial about your domain. You will never be able to draw false conclusions, but if you are too vague, you may not be able to draw some of the conclusions that you desire.

It turns out, in fact, that there is no way to axiomatize the natural numbers without including some weird unintended interpretations that have multiple copies of the natural numbers.

Still this shouldn't deter us from the enterprise of using logic to formalize reasoning inside computers. We don't have any substantially better alternatives, and, with care, we can make logic serve a useful purpose.