CS 323, Fall 1999
Problem Set #1: Circumscription
Handout #9

Due in class on Tuesday, October 26. (Two weeks from today.)
Note: The problems in this problem set are from [**Lif93**]. You can refer to it, if you are stuck, but I suggest you try to do the problems on your own – you'll learn a lot more. There are ten problems.

1 Circumscription

Circumscription is a rule of conjecture, which is applied to a logical theory[1] *A(P)* to give another theory. This resulting theory is denoted by Circ[*A(P)*; P]. Circ[*A(P)*; P] is a logical theory, which entails the same facts as *A(P)*, except the extent of the predicate P in the theory *A(P)* is minimized. In other words, as few tuples as possible satisfy P. In terms of models, the models of the theory Circ[*A(P)*; P] are exactly the P-minimal models of *A(P)*, which are those models M for which $\neg\exists M'.M' <_P M$.
There are two ways to describe circumscription: semantically (in terms of models) and syntactically (in terms of logical formulas).

1.1 Semantic Description of Circumscription

Model-theoretically, the models of Circ[*A(P)*; P] are those P-minimal models of *A(P)*. One way to compute the circumscription is to enumerate the models of *A(P)*, order them according to $\leq_P$, and then pick the models M for which $\neg\exists M'.M' <_P M$.
Recall the definition of $\leq_P$:
Take any two models of theory *A(P)* , M_1 and M_2. Then

$$\begin{aligned} M_1 \leq_P M_2 \quad &\equiv P^{M_1} \subseteq P^{M_2} \wedge \\ &(M_1 \text{ and } M_2 \text{ agree on all interpretations of other symbols,} \\ &\text{ be they constants, functions, or predicates)} \end{aligned} \tag{1}$$

Note that $\leq_P$ is a partial order, not total – if two models disagree on some symbol, or the Ps are not subsets of each other, then we are out of luck – we cannot compare the models.
Pictures always help. Let say our $A(P) \equiv [P(a) \wedge Q(a) \wedge Q(b)]$, and say our universe consists only of three different constants, $\{a, b, c\}$. The models of this theory are:

[1] A logical theory is a set of logical sentences. When we speak of a set of sentences $\{\phi, \psi, ...\}$, they are implicitly ANDed ($\wedge$) together.

$$\begin{aligned}
M_1 &= \{P = \{a\}, Q = \{a, b\}\} \\
M_2 &= \{P = \{a, b\}, Q = \{a, b\}\} \\
M_3 &= \{P = \{a, c\}, Q = \{a, b\}\} \\
M_4 &= \{P = \{a, b, c\}, Q = \{a, b\}\} \\
M_5 &= \{P = \{a\}, Q = \{a, b, c\}\} \\
M_6 &= \{P = \{a, b\}, Q = \{a, b, c\}\} \\
M_7 &= \{P = \{a, c\}, Q = \{a, b, c\}\} \\
M_8 &= \{P = \{a, b, c\}, Q = \{a, b, c\}\}
\end{aligned} \tag{2}$$

Here is a Hasse diagram of how these models are ordered according to $\leq_P$. Note that the models M_1 and M_5 are not comparable, as the Qs are different. M_2 and M_3 are not comparable since neither P is a subset of the other. M_1 and M_5 are the P-minimal models of Circ[*A(P)*; P], as they are the only models which do not have a model below themselves.

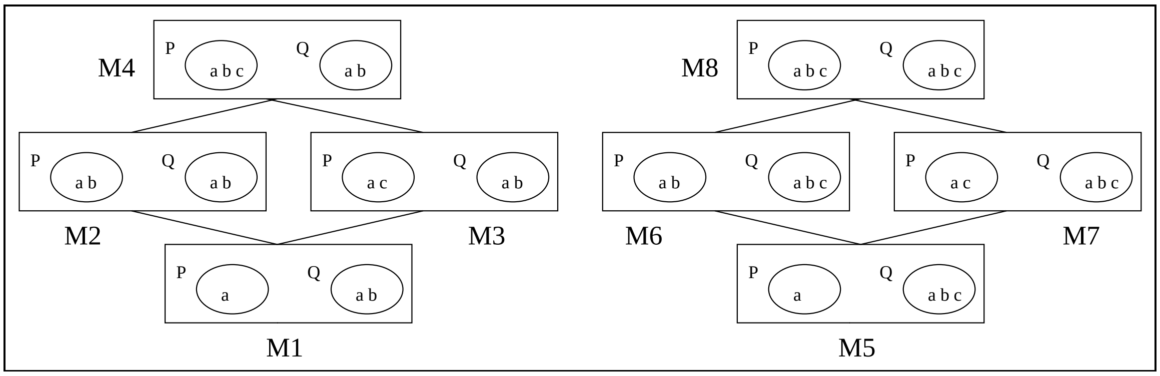

Figure 1: Models of the Theory $A(P) \equiv P(a) \wedge Q(a) \wedge Q(b)$.

1.2 Syntactic Description of Circumscription

To circumscribe a theory *A(P)* wrt to a predicate P, we append a second-order logic equation to our original theory.[2]

$$Circ[A(P); P] \equiv A(P) \;\wedge \forall\phi.[(A(\phi) \wedge \phi \leq P) \rightarrow \phi = P] \tag{3}$$

This says that the theory Circ[*A(P)*; P] is equivalent to the conjunction of two statements. The first asserts *A(P)*. The second statement says: "and if there is another predicate ϕ which looks just like P in the theory A, and it has a smaller extent, then that extent is the same as P's."

Put another (logically equivalent) way:

$$Circ[A(P); P] \equiv A(P) \;\wedge \neg\exists\phi.[A(P) \;\wedge \phi < P] \tag{4}$$

[2]Remember that $A(\phi)$ is just the theory *A(P)* where every occurrence of the predicate symbol P is replaced by ϕ. (We could probably do better by saying something like $A_P(\phi)$ to make the replacement of P with ϕ explicit.) Also recall that "$\phi \leq P$" is just shorthand for "$\forall x.\phi(x) \rightarrow P(x)$" and "$\phi = P$" for "$\forall x.\phi(x) \equiv P(x)$" You can think of "$\phi \leq P$" as saying that the tuples true of ϕ are a subset of those of P.

In plain English, the second conjunct says: "There is no predicate ϕ that can be smaller than P." This is another way of guaranteeing that P's extent will be minimized.

For you mathematicians out there, this definition is isomorphic to defining the minimum of a set X:

$$\forall x \in X.(x \leq Min(X) \rightarrow x = Min(X)). \tag{5}$$

The hard part of circumscription is finding an equivalent first order theory.[3] The form of the equivalent first order theory is usually a complete definition of P ($\forall x.P(x) \equiv ...$) conjoined with *A(P)*.

For example, the result of circumscribing *A(P)* wrt to P in the theory $A(P) \equiv [P(a) \wedge Q(a) \wedge Q(b)]$ is equivalent to: $Circ[A(P); P] \equiv (Q(a) \wedge Q(b) \wedge \forall x.[P(x) \equiv (x = a)])$. This corresponds to the models M_1 and M_5. Note that these are the only two models of *A(P)* which have the smallest Ps. All other models have two or more elements in the P relation.

Lifschitz's paper goes through many examples, two of which are repeated here. You'll never in real life have theories as simple as these, but they sure are fun to practice on.

Example 1: $A(P) \equiv P(a)$

$$\begin{aligned} Circ[A(P); P] &\equiv A(P) \wedge \forall \phi.[(A(\phi) \wedge \phi \leq P) \rightarrow \phi = P] \\ &\equiv P(a) \wedge \forall \phi.[(\phi(a) \wedge \forall x.\phi(x) \rightarrow P(x)) \rightarrow (\forall x.\phi(x) \equiv P(x))] \end{aligned} \tag{6}$$

To compute this circumscription, you must think of a definition for P that would be true for any ϕ. The best way to find this definition is to think about the problem semantically: what are the models of *A(P)* and of those, which have a P with the smallest extent? After some thought and practice (like doing this problem set), you will probably come up with a definition like $\forall x.P(x) \equiv (x = a)$. Now as a sanity check, we should ensure that these statements are equivalent:

Prove: $Circ[A(P); P] \equiv (\forall x.P(x) \equiv (x = a))$

$\rightarrow$:

$P(a)$	1. Given from Circ[*A(P)*; P]
$\forall \phi.[(\phi(a) \wedge \forall x.\phi(x) \rightarrow P(x)) \rightarrow (\forall x.\phi(x) \equiv P(x))]$	2. Given from: Circ[*A(P)*; P]
$(a = a \wedge \forall x.(x = a) \rightarrow P(x)) \rightarrow (\forall x.(x = a) \equiv P(x))$	3. Substitute $x = a$ for $\phi(x)$ (Universal instantiation in (2).)
$\forall x.(x = a) \equiv P(x)$	4. Modus Ponens (1, 3) (a = a is true, and P(a) is equivalent to $\forall x.(x = a) \rightarrow P(x)$)

(7)

We have to prove both $P(a)$ and $\forall \phi.[(\phi(a) \wedge \forall x.(\phi(x) \rightarrow P(x))) \rightarrow (\forall x.\phi(x) \equiv P(x))]$ from $\forall x.(x = a) \equiv P(x)$. The first is trivial, so let's try the second. Let's assume the antecedent, for arbitrary ϕ:

[3] Many times there is no equivalent first order theory, especially if there are an infinite number of elements in the domain.

$\leftarrow$:

$$\begin{array}{ll} \forall x.(x = a) \equiv P(x) & \text{1. Given} \\ \phi(a) \wedge \forall x.\phi(x) \rightarrow P(x) & \text{2. Assumed} \\ \phi(a) \wedge \forall x.\phi(x) \rightarrow (x = a) & \text{3. Substitute for def. of } P(x) \\ \phi(a) & \text{4. } \wedge\text{-elimination of (3)} \\ \forall x.\phi(x) \rightarrow (x = a) & \text{5. } \wedge\text{-elimination of (3)} \\ \forall x.(x = a) \rightarrow \phi(x) & \text{6. Equivalent to (4)} \\ \forall x.(x = a) \equiv \phi(x) & \text{7. Def. of } \equiv \text{ (5, 6)} \\ \forall x.\phi(x) \equiv P(x) & \text{8. Transitivity of } \equiv \text{ (1, 7)} \end{array} \tag{8}$$

Figuring out the minimal P seems like black magic, but easy if you enumerate models and think about it. The first trick of circumscription, used above, is the use of equality to completely specify the domain of a predicate. To say $P(x)$ holds only for the elements a_1, ... , a_n and no more, is to say: $P(x) \equiv (x = a_1) \vee ... \vee (x = a_n)$.

The second trick is to always start small with P(x), and then work your way up. You should always consider $P(x) \equiv \perp$ first. This is the same as saying that the extent of P is empty. This obviously wouldn't work in the first example, as we asserted $P(a)$. (P contains at least a.) This does work in the next example:

Example 2: $A(P) \ \equiv \neg P(a)$

$$Circ[A(P); P] \equiv \neg P(a) \wedge \forall \phi.[(\neg \phi(a) \wedge \forall x.[\phi(x) \rightarrow P(x)]) \rightarrow (\forall x.\phi(x) \equiv P(x))] \tag{9}$$

There's no reason for $P(x)$ to be true for any element – *A(P)* never specified it, and model-theoretically, one of the models of $\neg P(a)$ has an empty extent for P. So let's set $\forall x.P(x) \equiv \perp$, or equivalently, $\forall x.\neg P(x)$.

Prove: $Circ[A(P); P] \equiv (\forall x.P(x) \equiv \perp)$

$\rightarrow$:

$$\begin{array}{ll} \neg P(a) & \text{1. Given from Circ}[A(P)\text{; P}] \\ \forall \phi.[(\neg \phi(a) \wedge \forall x.\phi(x) \rightarrow P(x)) \rightarrow (\forall x.\phi(x) \equiv P(x))] & \text{2. Given from: Circ}[A(P)\text{; P}] \\ (\neg \perp \wedge \forall x. \perp \rightarrow P(x))) \rightarrow (\forall x. \perp \equiv P(x))) & \text{3. Substitute } \perp \text{ for } \phi(x) \\ & \text{(Universal instantiation of the empty relation.)} \\ \forall x. \perp \equiv P(x) & \text{4. Modus Ponens (3) } (\neg \perp \text{ is true,} \\ & \text{as is } \forall x. \perp \rightarrow P(x).) \end{array} \tag{10}$$

For the other direction, we have to prove both $\neg P(a)$ and $\forall \phi.[(\neg \phi(a) \wedge \forall x.(\phi(x) \rightarrow P(x))) \rightarrow (\forall x.\phi(x) \equiv P(x))]$ from $\forall x. \perp \equiv P(x)$ The first is trivial, so let's try the second. Let's assume the antecedent, for arbitrary ϕ:

$\leftarrow$:

$$\begin{array}{ll} \forall x.P(x) \equiv \perp & \text{1. Given} \\ \neg \phi(a) \wedge \forall x.\phi(x) \rightarrow P(x) & \text{2. Assumed} \\ \neg \phi(a) \wedge \forall x.\phi(x) \rightarrow \perp & \text{3. Substitute for def. of } P(x) \\ \neg \phi(a) & \text{4. } \wedge\text{-elimination of (3)} \\ \forall x.\phi(x) \rightarrow \perp & \text{5. } \wedge\text{-elimination of (3)} \\ \forall x. \perp \rightarrow \phi(x) & \text{6. Tautology} \\ \forall x.\phi(x) \equiv \perp & \text{7. Definition (5, 6)} \\ \forall x.\phi(x) \equiv P(x) & \text{8. Transitivity of } \equiv \text{ (1, 7)} \end{array} \tag{11}$$

You can think of Circ[$A(P)$; P]as this funny operator which looks at the models of P, and snaps the domain of P like a rubber band to one that is as small as possible, while still obeying whatever $A(P)$ says.

1.3 Problems

Given the tricks above, these formulas should be a piece of cake. Just return (don't prove) the first order equivalent of Circ[$A(P)$; P]. (Hint: they are all of the form $\forall x.[P(x) \equiv ...].$) If you are having trouble, try enumerating the models of $A(P)$, and looking at the P-minimal ones.

1. $A(P) \equiv [P(a) \wedge P(b)]$
2. $A(P) \equiv [P(a) \rightarrow P(b)]$
3. $A(P) \equiv [\forall x.P(x)]$
4. $A(P) \equiv [\forall x.\neg P(x)]$
5. $A(P) \equiv [\forall x.Q(x) \rightarrow P(x)]$
6. $A(P) \equiv [(\forall x.Q(x) \rightarrow P(x)) \wedge P(a)]$
7. $A(P) \equiv [P(a) \vee P(b)]$ (This one is tricky! The form won't be exactly like the others.)
8. Pick one of the above and prove that the circumscription is equivalent to your answer, as we did above.

2 Circumscription with Varying Symbols

We can extend the operation of circumscription to let other symbols vary while trying to minimize P, which gives us extra flexibility. Here's a common sense example to give you some intuition. Say you are trying to maximize the amount of time you spend on problem sets, and you come up with the following equation:

$$[\text{time spent on problem sets}] = 24 - [\text{time spent in class}] - [\text{time spent sleeping}]. \quad (12)$$

First say you fix [time spent sleeping] to 8, and maximize problem set time by only letting [time spent in class] vary. If you skipped all your classes, you could devote 16 hours to problem sets. This is ok. But say you let your sleep time vary as well. Now you have added flexibility, and can spend the whole 24 hours doing homework, while skipping every class and not getting too much sleep!

2.1 Semantic Description

As before there are semantic and syntactic ways to describe circumscription with variation. The following example first shows the result of circumscription without variation.

Example 3: Let $A(P) \equiv [P(a) \vee Q(a)]$.

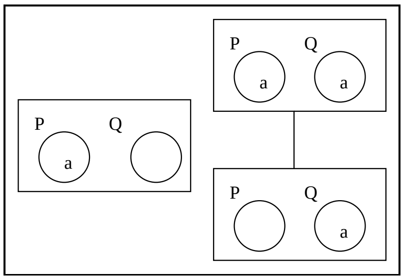

Figure 2: Models of the Theory *A(P)*$\equiv [P(a) \vee Q(a)]$.

What happens when we circumscribe *A(P)* wrt P? Let's look at the models, here drawn as a Hasse diagram over the relation $\leq_P$:
Somewhat unintuitively, we get two P-minimal models: $\{Q = \{\}, P = \{a\}\}$, and $\{Q = \{a\}, P = \{\}\}$.
We would really like to have the only P-minimal model be the second model, but the existence of the relation Q prevents the models from being comparable.
We can get around this by letting other symbols vary while doing the minimization. We write this as Circ[A(P, Z); P; Z],[4] where P is the predicate to be minimized, and Z is the symbol that is varied. Model theoretically, this just means we have a different definition of $\leq_{P;Z}$:

$$M_1 \leq_{P;Z} M_2 \equiv P^{M_1} \subseteq P^{M_2} \wedge \text{(M_1 and M_2 agree on all interpretations of symbols, } \textbf{except Z}) \quad (13)$$

When we do the comparisons, we can ignore Z. So back to Example 3, if we instead ordered the models of $P(a) \vee Q(a)$ with $\leq_{P;Q}$, we would get:

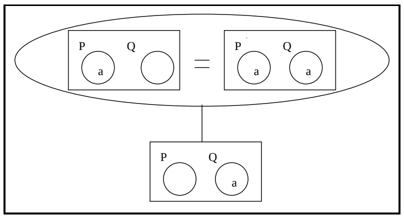

Figure 3: Models of the Theory *A(P)*$\equiv P(a) \vee Q(a)$, ordered by $\leq_{P;Q}$.

Under the $\leq_{P;Q}$ relation, that pesky first model $\{P = \{a\}, Q = \{\}\}$ looks like the model $\{P = Q = \{a\}\}$, and we get only one minimal model, the one we wanted! The $\leq_{P;Z}$ relation is so much more powerful, by letting the symbol Z be ignored while doing the minimization.

2.2 Syntactic Description

You can also see this flexibility in the syntactic definition:

[4] $A(P, Z)$ is just the theory A with P replaced for P, and Z replaced for Z. To get $A(\phi, \zeta)$, we would replace P with ϕ and Z with ζ.

$$
\begin{aligned}
Circ[A(P,Z);P;Z] &\equiv A(P) \ \wedge \forall\phi,\zeta.[(A(\phi,\zeta) \wedge \phi < P) \to \phi = P] \\
&\equiv A(P) \ \wedge \forall\phi.[(\exists\zeta.A(\phi,\zeta) \wedge \phi < P) \to \phi = P]
\end{aligned} \tag{14}
$$

Remember the trick to circumscription is to find a definition for P which satisfies the implication for any ϕ. Replacing Z with an existentially quantified variable makes it easier for ϕ to satisfy the antecedent of the implication. (ζ can take on whatever value it needs to.) This in turn means that P must work harder to be the minimal P, regardless of Z.

Example 4: Let $A(P) \ \equiv [\forall x.Q(x) \to P(x)]$

$$
\begin{aligned}
Circ[A(P,Q);P;Q] \equiv \ & [\forall x.Q(x) \to P(x) \wedge \\
& \forall\phi.[(\exists\zeta.\forall x.\zeta(x) \to \phi(x) \wedge \forall x.\phi(x) \to P(x)) \to (\forall x.\phi(x) \equiv P(x))]
\end{aligned} \tag{15}
$$

This looks intimidating, but consider the models of *A(P)*. All that is asserted is that the tuples of Q are also in P, so really we have this general picture of the models of *A(P)*:

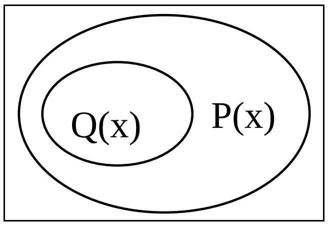

Figure 4: Models of the Theory *A(P)*$\equiv \forall x.Q(x) \to P(x)$.

But we want P to be as small as possible, empty in fact if we can. If we set $P(x)$ to be empty (think of the rubber band) then Q will also have to be empty. Thus the equivalent first order theory is $\forall x.\neg P(x) \wedge \forall x.\neg Q(x)$. We can let $Q(x)$ be whatever we like, since we said we are letting it vary, just as long as the definition we chose satisfies *A(P)*. You can check the proof of equivalence yourself. You should also compare the result to when Q isn't allowed to vary. (In problem 5 above.)

Example 5: Let $A(P) \ \equiv P(a) \wedge P(b)$.

$$
\begin{aligned}
Circ[A(P,b);P;b] \equiv \ & P(a) \wedge P(b) \wedge \\
& \forall\phi.[(\exists\zeta.(\phi(a) \wedge \phi(\zeta)) \wedge \forall x.[\phi(x) \to P(x)]) \to \forall x.(\phi(x) \equiv P(x))]
\end{aligned} \tag{16}
$$

Here is an example where we let the constant b vary. Now, remember we want to make ϕ as small as possible. We could set $P(x) \equiv (x = a) \vee (x = b)$, but we can actually find a smaller extent, if we let ζ above be set to a. Then $\forall x.P(x) \equiv (x = a)$, which works, so that $Circ[A(P,b);P;b] \equiv \forall x.(x = a) \equiv P(x)$, which implicitly means that $a = b$ – we've varied b to be equal to a.

2.3 Problems

You should now have more than enough ammunition to do these problems:

1. Prove: $Circ[A(P,Z);P;Z] \equiv A(P,Z) \wedge Circ[\exists\zeta.A(P,\zeta);P]$ (We can always reduce circumscription with variation to simple circumscription.)

2. Consider the following theory $A(ab)$ about birds:

$$\begin{array}{l} \forall x.(\neg ab(x) \wedge bird(x)) \rightarrow flies(x) \\ bird(Tweety) \end{array} \tag{17}$$

   (a) Assume that Tweety is the only constant in the universe. Draw all the models of this theory, and order them according to $\leq_{ab}$. Can we conclude anything about Tweety flying from the theory $Circ[A(ab);ab]$?

   (b) Draw the models again, this time with ordering $\leq_{ab;flies}$. What can you conclude from the theory $Circ[A(ab);ab;flies]$?

   (c) Find and prove the first order equivalent of $\texttt{Circ}[A(ab);ab;flies]$

   (d) DLS and SCAN are two algorithms which reduce second-order logical formulas to first order theories. (Sometimes.)

   SCAN can be found at `http://www.mpi-sb.mpg.de/õhlbach/scan/`. It works by "resolving" second-order predicates against each other.

   DLS is at `http://www.ida.liu.se/labs/kplab/projects/dls/`. DLS works using Ackerman's lemma.[5]

   Since SCAN's webpage form is broken, we are going to use DLS's. Go to `http://www.ida.liu.se/labs/kplab/projects/dls/circsimple.html`, and modify the theory to match $A(ab)$. Set "Predicates to be minimized" to Ab, and nothing for "Predicates to be varied." (Erase the entire field or it will complain.) Submit the form, and return the output as your answer.

   (e) The output is only a consequence of the input – it isn't complete. If you combine the output with the original theory above, what are the only two possibilities for Tweety in terms of $bird$, $flies$, and ab? (This should be the same as your answer to 2a.)

   (f) Now go back and set "Predicates to be varied" to Flies. What's the answer now? What can we conclude about Tweety?

Feel free to use this tool to check your answers for the other problems!

[5] Ackerman's lemma is $\exists\Phi.\forall x.[\Phi(x) \vee A(x,z)] \wedge B[\neg\Phi] \equiv B[A(x,z)]$, where Φ is a predicate symbol, A is a theory with variables x and z with no mention of Φ, and B is a theory which is positive with respect to its argument. (So in $B(\neg\Phi)$, Φ appears negatively.) You can roughly see why this lemma holds. $\Phi(x) \vee A(x,z)$ is equivalent to $\neg\Phi(x) \rightarrow A(x,z)$, so we can "substitute" A for $\neg\Phi$ in B. To go the in the other direction, just note that we can take Φ to be $\neg A$.