CS472 Foundations of Artificial Intelligence
Fall 2000
Assignment 3
SOLUTION

1. **Representing Sentences in First-order Logic**

   (a) $\exists x : [did(Fred, x) \wedge annoys(x, Wilma)]$ where $did(a, b)$ means 'a performed action b' and $annoys(a, b)$ means 'action a annoys b'.

   (b) $\forall x \forall s \exists f : ([push(x, s) \wedge hardEnough(x, f) \wedge \neg greater(s, f)] \rightarrow fall(x))$ where $push(a, b)$ means 'a is pushed with force b', $hardEnough(a, b)$ means 'b is the force required to cause a to fall over', $greater(a, b)$ means $a > b$, and $fall(x)$ means 'x falls over'.

   (c) $ugly(Me) \rightarrow \exists m : [instance(m, Monkey) \wedge uncle(You, m)]$, where $ugly(x)$ means 'x is ugly', $uncle(a, b)$ means 'a is an uncle of b', and $instance(a, b)$ has it's standard meaning.

   (d) $\exists f : instance(f, fight) \rightarrow \exists s_1 : [started(s_1, f) \wedge \exists s_2 : [\neg equal(s_1, s_2) \wedge started(s_2, f)]]$, where $started(a, b)$ means 'a started b', and $instance(a, b)$ and $equal(a, b)$ have their standard meanings.

   (e) $\forall x : [person(x) \wedge (\forall y\ vegetarian(y) \rightarrow dislikes(x, y)) \rightarrow stupid(x)]$, where $person(x)$ means 'x is a person', $vegetarian(x)$ means 'x is a vegetarian', $dislikes(x, y)$ means 'x dislikes y', and $stupid(x)$ means 'x is stupid'.

   (f) $\forall x, y : [person(x) \wedge smart(y) \wedge vegetarian(y) \rightarrow \neg smart(x)]$; the intended interpretations should be clear here, given the last problem.

   (g) $\exists x : [barber(x) \wedge \forall y man(y) \wedge \neg shaves(y, y) \rightarrow shaves(x, y)]$, where $barber(x)$ means 'x is a barber', $man(x)$ means 'x is a man', and $shaves(x, y)$ means 'x shaves y'.

   Grader comments:

   (a) Most people managed to do this problem to quite reasonable extend. There were some minor mistakes (for which no points were deducted), like (1) not following the letter-case conventions for predicate, variable and constant names and (2) failing to place parenthesis at the appropriate places.

   (b) One point was deducted for each part, in which the student failed (or missed) to explain what do their predicates mean. It was required to do so, as stated in the example from the problem set.

   (c) Generally there were no problems with (a). Many students messed up with (b). Some argued about $\exists f$ - force, but failed to state that it is the minimum force required (-1 point). Others completely avoided forces and similar entities (-2 points). Another relatively common mistake is placing $\wedge$ instead of $\rightarrow$ in some places. Finally, some students nearly got (c), but forgot to include that the "two" $\exists x \exists y$ who started the fight should be deferent, i.e. $\neg same(x, y)$ (-1 point).

   (d) It is very inappropriate to use open terms (terms with some free occurences of variables) to represent single English sentences (sentences out of context). Point deductions vary, according to severity of the problem.

2. **Unification**

   (a) In trying to unify p and q, we're forced to unify y and $f(y, A)$ which fails the "occurs check", so p and q do not unify.

   (b) Let $\sigma = \{v/j(x, y), u/j(x, y), w/j(x, y)\}$ to get $p\sigma = q\sigma = f(g(j(x, y)), h(j(x, y), j(x, y)))$.

   (c) In trying to unify p and q, we're forced to unify w and $j(x, w)$ which fails the "occurs check", so p and q do not unify.

   (d) Let $\sigma = \{x/f(B, A), u/A, y/B, z/A\}$ to get $p\sigma = q\sigma = f(f(B, A), f(A, f(B, A)))$.

3. **Clausal Form**

   (a) $\neg P(x) \vee P(x)$ (or, for you simplificationists, *true*).

   (b) $P(x) \vee \neg P(S1)$, which also simplifies to *true*. Note that the x's in the two parts of the implication are *not* the same variable.

   (c) There are actually 2 answers to this one, because of some ambiguity in the formula. If you interpret the overall structure to be $A \rightarrow (B \wedge C)$ (as was intended), then it translates and simplifies to:

   $P(S1)$
   $\neg P(f(S1, S2)) \vee \neg Q(S1, y) \vee P(y)$

   by the following derivation:

0	$\neg\forall x : (P(x) \rightarrow \forall y : [P(x) \rightarrow P(f(x, y))] \wedge \neg\forall y : [Q(x, y) \rightarrow P(y)])$
1	$\neg\forall x : (\neg P(x) \vee \forall y : [\neg P(x) \vee P(f(x, y))] \wedge \neg\forall y : [\neg Q(x, y) \vee P(y)])$
2	$\exists x : \neg(\neg P(x) \vee \forall y : [\neg P(x) \vee P(f(x, y))] \wedge \neg\forall y : [\neg Q(x, y) \vee P(y)])$
	$\exists x : (P(x) \wedge \neg\forall y : [\neg P(x) \vee P(f(x, y))] \wedge \neg\forall y : [\neg Q(x, y) \vee P(y)])$
	$\exists x : (P(x) \wedge \neg\forall y : [\neg P(x) \vee P(f(x, y))] \vee \forall y : [\neg Q(x, y) \vee P(y)])$
	$\exists x : (P(x) \wedge \exists y : \neg[\neg P(x) \vee P(f(x, y))] \vee \forall y : [\neg Q(x, y) \vee P(y)])$
	$\exists x : (P(x) \wedge \exists y : [P(x) \wedge \neg P(f(x, y))] \vee \forall y : [\neg Q(x, y) \vee P(y)])$
3	$(P(S1) \wedge [P(S1) \wedge \neg P(f(S1, S2))] \vee \forall y : [\neg Q(S1, y) \vee P(y)])$
4	*no change*
5	$\forall y : (P(S1) \wedge [P(S1) \wedge \neg P(f(S1, S2))] \vee [\neg Q(S1, y) \vee P(y)])$
6	$\forall y : (P(S1) \wedge [P(S1) \vee \neg Q(S1, y) \vee P(y) \wedge \neg P(f(S1, S2)) \vee \neg Q(S1, y) \vee P(y))$
6.5	$\forall y : (P(S1) \wedge \neg P(f(S1, S2)) \vee \neg Q(S1, y) \vee P(y))$
7	$P(S1)$
	$\neg P(f(S1, S2)) \vee \neg Q(S1, y) \vee P(y)$

   but if you interpretted it to be $(A \rightarrow B) \wedge C)$, it becomes:

   $P(S1)$
   $\neg P(f(S1, S2)$
   $\neg Q(S1, y) \vee P(y)$

   by the following derivation:

$$
\begin{array}{ll}
0 & \neg\forall x : (P(x) \rightarrow \forall y : [P(x) \rightarrow P(f(x,y))] \wedge \neg\forall y : [Q(x,y) \rightarrow P(y)]) \\
1 & \neg\forall x : (\neg P(x) \vee \forall y : [\neg P(x) \vee P(f(x,y))] \wedge \neg\forall y : [\neg Q(x,y) \vee P(y)]) \\
2 & \exists x : \neg(\neg P(x) \vee \forall y : [\neg P(x) \vee P(f(x,y))] \wedge \neg\forall y : [\neg Q(x,y) \vee P(y)]) \\
 & \exists x : (\neg\neg P(x) \vee \forall y : [\neg P(x) \vee P(f(x,y))] \wedge \forall y : [\neg Q(x,y) \vee P(y)]) \\
 & \exists x : (P(x) \wedge \neg\forall y : [\neg P(x) \vee P(f(x,y))] \wedge \forall y : [\neg Q(x,y) \vee P(y)]) \\
 & \exists x : (P(x) \wedge \exists y : \neg[\neg P(x) \vee P(f(x,y))] \wedge \forall y : [\neg Q(x,y) \vee P(y)]) \\
 & \exists x : (P(x) \wedge \exists y : [P(x) \wedge \neg P(f(x,y))] \wedge \forall y : [\neg Q(x,y) \vee P(y)]) \\
3 & (P(S1) \wedge [P(S1) \wedge \neg P(f(S1,S2))] \wedge \forall y : [\neg Q(S1,y) \vee P(y)]) \\
4 & \textit{no change} \\
5 & \forall y : (P(S1) \wedge [P(S1) \wedge \neg P(f(S1,S2))] \wedge [\neg Q(S1,y) \vee P(y)]) \\
6.5 & \forall y : (P(S1) \wedge \neg P(f(S1,S2)) \wedge [\neg Q(S1,y) \vee P(y)]) \\
7 & P(S1) \\
 & \neg P(f(S1,S2) \\
 & \neg Q(S1,y) \vee P(y)
\end{array}
$$

(d) $P(S1,x) \vee Q(x,S1) \vee \neg P(S2,y) \vee S(S2,y)$
$P(S1,x) \vee \neg S(S1,x) \vee \neg P(S2,y) \vee S(S2,y)$
$\neg Q(x,S1) \vee \neg S(S1,x) \vee \neg P(S2,y) \vee S(S2,y)$
by the following derivation:

$$
\begin{array}{ll}
0 & \forall x \exists y : ([P(x,y) \rightarrow Q(y,x)] \wedge [Q(y,x) \rightarrow S(x,y)]) \rightarrow \exists x \forall y : [P(x,y) \rightarrow S(x,y)] \\
1 & \neg\forall x \exists y : ([\neg P(x,y) \vee Q(y,x)] \wedge [\neg Q(y,x) \vee S(x,y)]) \vee \exists x \forall y : [\neg P(x,y) \vee S(x,y)] \\
2 & \exists x \forall y : \neg([\neg P(x,y) \vee Q(y,x)] \wedge [\neg Q(y,x) \vee S(x,y)]) \vee \exists x \forall y : [\neg P(x,y) \vee S(x,y)] \\
 & \exists x \forall y : (\neg[\neg P(x,y) \vee Q(y,x)] \vee \neg[\neg Q(y,x) \vee S(x,y)]) \vee \exists x \forall y : [\neg P(x,y) \vee S(x,y)] \\
 & \exists x \forall y : ([P(x,y) \wedge \neg Q(y,x)] \vee [Q(y,x) \wedge \neg S(x,y)]) \vee \exists x \forall y : [\neg P(x,y) \vee S(x,y)] \\
3 & \forall y : ([P(S1,y) \wedge \neg Q(y,S1)] \vee [Q(y,S1) \wedge \neg S(S1,y)]) \vee \forall y : [\neg P(S2,y) \vee S(S2,y)] \\
4 & \forall x : ([P(S1,x) \wedge \neg Q(x,S1)] \vee [Q(x,S1) \wedge \neg S(S1,x)]) \vee \forall y : [\neg P(S2,y) \vee S(S2,y)] \\
5 & \forall x \forall y : [([P(S1,x) \wedge \neg Q(x,S1)] \vee [Q(x,S1) \wedge \neg S(S1,x)]) \vee [\neg P(S2,y) \vee S(S2,y)]] \\
6 & \forall x \forall y : [([P(S1,x) \vee [Q(x,S1) \wedge \neg S(S1,x)]] \\
 & \qquad \wedge[\neg Q(x,S1) \vee [Q(x,S1) \wedge \neg S(S1,x)]]) \vee [\neg P(S2,y) \vee S(S2,y)]] \\
 & \forall x \forall y : [([P(S1,x) \vee Q(x,S1)] \wedge [P(S1,x) \vee \neg S(S1,x)] \\
 & \qquad \wedge[\neg Q(x,S1) \vee Q(x,S1)] \wedge [\neg Q(x,S1) \vee \neg S(S1,x)]) \vee [\neg P(S2,y) \vee S(S2,y)]] \\
 & \forall x \forall y : ([P(S1,x) \vee Q(x,S1) \vee \neg P(S2,y) \vee S(S2,y)] \\
 & \qquad \wedge[P(S1,x) \vee \neg S(S1,x) \vee \neg P(S2,y) \vee S(S2,y)] \\
 & \qquad \wedge[\neg Q(x,S1) \vee Q(x,S1) \vee \neg P(S2,y) \vee S(S2,y)] \\
 & \qquad \wedge[\neg Q(x,S1) \vee \neg S(S1,x) \vee \neg P(S2,y) \vee S(S2,y)]) \\
6.5 & \forall x \forall y : ([P(S1,x) \vee Q(x,S1) \vee \neg P(S2,y) \vee S(S2,y)] \\
 & \qquad \wedge[P(S1,x) \vee \neg S(S1,x) \vee \neg P(S2,y) \vee S(S2,y)] \\
 & \qquad \wedge[\neg Q(x,S1) \vee \neg S(S1,x) \vee \neg P(S2,y) \vee S(S2,y)]) \\
7{+}8 & P(S1,x1) \vee Q(x1,S1) \vee \neg P(S2,y1) \vee S(S2,y1) \\
 & P(S1,x) \vee \neg S(S1,x2) \vee \neg P(S2,y2) \vee S(S2,y2) \\
 & \neg Q(x3,S1) \vee \neg S(S1,x3) \vee \neg P(S2,y3) \vee S(S2,y3)
\end{array}
$$

Grader Comments

(a) The following codes represent the bulk of the mistakes that students in the class make.

(b) *&&1*: Most answers are not entirely correct, but marks are awarded for partially correct solutions. However about 1/5 of the class gave only the final solutions instead of the step by step derivations. 1 or 2 marks are usually awarded for those scripts with only partially correct final answers as there are no means to give credits for the derivation.

(c) *&&2*: Skolemization is the process of removing existential quantifiers by elimination. In the case when an existential quantifier, $\exists y$:], is nested inside another universal (not existential) quantifier, $\forall x$:], S(x) can replace all instances of y in the first-order logic. However, y must be nested within the scope of x for this to apply. In part (c) and part (d), the skolemization are not nested within any universal quantifiers if the steps are done properly, and hence simple replacement by constants suffice. Note that variable y is not nested within the universal quantifier z in this example:
$\forall x \exists y \forall z : P(x) \vee P(y) \vee P(z)$

(d) *&&3*: About 1/4 of the class makes this mistake when removing the implication in part (c) and part (d). Removing implication from
$\forall x \exists y : ([P(x,y) \rightarrow Q(y,x)] \wedge [Q(y,x) \rightarrow S(x,y)]) \rightarrow \exists x \forall y : [P(x,y) \rightarrow S(x,y)]$
should yield
$\neg \forall x \exists y : ([\neg P(x,y) \vee Q(y,x)] \wedge [\neg Q(y,x) \vee S(x,y)]) \vee \exists x \forall y : [\neg P(x,y) \vee S(x,y)]$
and not
$\forall x \exists y : (\neg[\neg P(x,y) \vee Q(y,x)] \wedge [\neg Q(y,x) \vee S(x,y)]) \vee \exists x \forall y : [\neg P(x,y) \vee S(x,y)]$
The negation affects the quantifiers too. This mistake often leads to a wrong skolemization step in part (d) because the quantifier for variable x in the wrong case will be an universal quantifier instead of an existential quantifier.

(e) The sequences of steps taken to convert to normal form is not arbitrary. You can either follow the sequences shown in the lecture notes or the sequence in the textbook. For example, you must remove all implications before the quantifiers are moved to the left. Also if you are using the textbook model where quantifiers are moved left before the skolemization process, make sure that the original nested order of the quantifiers are preserved, otherwise &&2 may be repeated. The lecture notes sequence perform skolemization before moving the quantifiers, and hence does not have this particular problem.

4. **Modes of Inference (X pts.)**

(a) Everyday examples

i. Examples of Deductive Inference

- *If someone gets caught in a rainstorm, he/she is wet.* Given "David is caught in a rainstorm", conclude "David is wet".
- *If too many people are logged into a computer, then the computer is slow.* Given "Too many people are logged into the computer", conclude "The computer is slow".
- *If less than 8 players on a team show up for a softball game, then that team loses the game.* Given "Only 7 members of the computer science softball team showed up for their game", conclude "The computer science softball team lost their game".

ii. Examples of Abductive Inference

- *If someone gets caught in a rainstorm, he/she is wet.* Given "David is wet", conclude "David got caught in a rainstorm".
- *If someone is drunk, then he/she will not be able to walk a straight line.* Given "Kevin cannot walk a straight line", conclude "Kevin is drunk".
- *If someone has the flu, then he/she feels terrible.* Given "Jane feels terrible", conclude "Jane has the flu".

iii. Examples of Plausible Inference in Social Interactions
   - Given "A wants object B. C has B and A is a mean, sly person", conclude "A will do something devious to get B from C.
   - Given "A is an honest person and A finds one million dollars on the highway", conclude "A will turn in the money to the police".
   - Given "Women W_1 loves a man M, but M loves another woman W_2", conclude "W_1 will be jealous of W_2.

(b) Credibility

i. Credibility of Deductive Inference

Deductive inference is guaranteed by the inference method itself (*modus ponens*) since deduction is truth-preserving. Given the truth of the premise and the implication, deductive inference doesn't allow incorrect conclusions.

ii. Credibility of Abductive Inference

Abductive inference does not guarantee credibility. Some credible examples are given above. Here's a not-so-credible example:

> Given "If it is 6 a.m. then I am asleep" and "I am asleep", conclude "It is 6 a.m.".

Here's an even less-credible example:

> Given "If green martians live in my bedroom, then I will eat my lunch" and "I am eating my lunch", conclude "Green martians live in my bedroom".

Here the A part of the implication $A \rightarrow B$ is always false, so the implication is always true.

Some possible factors determining credibility:

- *Causality*: if the implication represents a causal relationship, then the inference is more believable. Being drunk *causes* a person to be unable to walk in a straight line, while the fact that it is 6 a.m. does not *cause* a person to be asleep.
- Credibility increases as the quality of the explanation increases. If a precedent A is an especially good *explanation* for the antecedent B, based on the information available, then abductive inference may be very credible.
- The probability of the precedent actually occurring is a third factor. If if the precedent is extremely unlikely or impossible, then the implication doesn't provide much information, as in the silly "martians" example above.

Here's one more credible example:

> Given "If there's a terrible earthquake, then some large buildings will collapse" and "Some large buildings have collapsed", conclude "There's been a terrible earthquake".

Even though the probability of a terrible earthquake occurring is fairly small, an earthquake is a good explanation for collapsed buildings, and the implication certainly represents a causal relationship.

Clearly, we have to be more careful when making abductive inferences.

iii. Credibility of Plausible Inference in Social Interactions

Plausible inference in social interactions does not guarantee credibility. Some credible examples are given above. Here's a not-so-credible example:

> Given "A wants object B. C has B and A is a mean, sly person", conclude "A will wait for C to give object B to him/her".

Some possible factors determining credibility:

- The degree to which the people involved have the qualities (emotions, traits, etc.) specified in the statements determines the credibility. If A wants object B *very* much, then A is more likely to do something devious to get it. Moreover, is A is truly despicable, he/she might be more likely to do something mean or devious to get B without even wanting it that much.
- The more consistent the inference is with general world knowledge concerning the social phenomenon, the more believable the inference.
- The relevance of and belief in the world knowledge used to derive the inference from the viewpoint of the person evaluating the credibility of the inference is also a factor. A Marxist, for example, may find an inference about human nature to be credible, while a staunch capitalist could find the same inference to be ridiculous.
- The culture in which the inference is made. For example, the following inference:

  > Given "A is unmarried and is very attracted to B", conclude "A will ask B on a date".

  may be applicable in Western culture, but not applicable in cultures where casual dating does not occur for, say, religious reasons.

Note that it is incorrect to dismiss this kind of reasoning as not credible. Clearly, we need to make such inferences all the time, or we wouldn't have much luck finding friends, jobs, dates, etc. Some of our inferences are better than others; we were hoping for a discussion of the subtleties that arise.

Grader comments:

i. Most people received full marks (15) for part (a). Some people only included one example for each case, instead of three, and I gave them two marks for each case instead of 5. A number of students were very unclear about the direction of the implication for the abductive case. The question asked for an example of an *inference*, not an *implication*, so each example should have read something like "If A, then B. From B, conclude A." Many students only included the $A \rightarrow B$ part, which didn't always make clear which way the inference went.

ii. Grading for part (b) was fairly subjective. I gave 5 points for the discussion of deductive reasoning, which most people got. (Occasionally students got the implication and the inference methods mixed up; clearly, a false implication can lead to a false conclusion, but this doesn't mean that the inference method itself is flawed.) For the discussion of abductive reasoning, worth 10 points, many students didn't discuss the subtleties included in the solution above. It's not true, for example, that abductive reasoning is only credible if the implication represents an "if and only if" relationship: there are degrees of credibility. I took off several marks if students didn't discuss these subtleties, or failed to include examples of non-so-credible examples. Grading for the last part, the discussion of the credibility of plausible inference in social interactions, was similar.