# First-order logic:
# Free and bound variables. Scope of a quantifier. Substitution of terms for variables. Capture. Variable renaming.

Valentin Goranko

School of Mathematics, University of the Witwatersrand
Johannesburg, South Africa

February 2008

# Free and bound variables

Two essentially different ways in which we use individual variables in first-order formulae:

1. Free variables: used to denote *unknown or unspecified objects*, as in $(\mathbf{5} < x) \vee (x^2 + x - \mathbf{2} = \mathbf{0})$.
2. Bound variables: used *to quantify*, as in
   $\exists x((\mathbf{5} < x) \vee (x^2 + x - \mathbf{2} = \mathbf{0}))$
   and $\forall x((\mathbf{5} < x) \vee (x^2 + x - \mathbf{2} = \mathbf{0}))$.

Note that the same variable can be *both free and bound in a formula*, e.g. $x$ in the formula $x > \mathbf{0} \wedge \exists x(\mathbf{5} < x)$.

A formula with no bound variables is an open formula.
A formula with no free variables is a closed formula, or a sentence.

# Scope of a quantifier

Scope of (an occurrence of a) quantifier in a given formula $A$: the *unique* subformula $QxB$ begun by that occurrence of the quantifier. E.g.:

$$\underline{\forall x}((x > \mathbf{5}) \to \forall y(y < \mathbf{5} \to (y < x \land \exists x(x < \mathbf{3})))).$$

$$\forall x((x > \mathbf{5}) \to \underline{\forall y}(y < \mathbf{5} \to (y < x \land \exists x(x < \mathbf{3})))).$$

$$\forall x((x > \mathbf{5}) \to \forall y(y < \mathbf{5} \to (y < x \land \underline{\exists x}(x < \mathbf{3})))).$$

A bound occurrence of a variable $x$ is bound by the *innermost* occurrence of a quantifier $Qx$ in the scope of which it occurs. E.g.:

$$\forall x((x > \mathbf{5}) \to \forall y(y < \mathbf{5} \to (y < x \land \exists x(x < \mathbf{3})))),$$

while

$$\forall x((x > \mathbf{5}) \to \forall y(y < \mathbf{5} \to (y < x \land \exists x(x < \mathbf{3})))).$$

# Using bound and free variables in a formula

Free variables have their own values in a given formula (determined by a variable assignment), while bound variables only play a dummy role and can be replaced (with care!) by one another.

For instance, the sentence

$$\exists x(\mathbf{5} < x \wedge x^2 + x - \mathbf{2} = \mathbf{0})$$

means exactly the same as

$$\exists y(\mathbf{5} < y \wedge y^2 + y - \mathbf{2} = \mathbf{0})$$

Likewise, $\forall x(\mathbf{5} < x \vee x^2 + x - \mathbf{2} = \mathbf{0})$
means the same as $\forall y(\mathbf{5} < y \vee y^2 + y - \mathbf{2} = \mathbf{0})$.

On the other hand, the meaning of

$$\mathbf{5} < x \wedge x^2 + x - \mathbf{2} = \mathbf{0}$$

is essentially different from the meaning of

$$\mathbf{5} < y \wedge y^2 + y - \mathbf{2} = \mathbf{0}.$$

# Reusing variables as free and bound in a formula

The same variable can occur both free and bound in a formula:

$$x > \mathbf{5} \rightarrow \forall x(\mathbf{2}x > x).$$

However, the free occurrence of $x$ has nothing to do with the bound occurrences of $x$:

$$x > \mathbf{5} \rightarrow \forall x(\mathbf{2}x > x).$$

Thus, the formula above has the same meaning as

$$x > \mathbf{5} \rightarrow \forall y(\mathbf{2}y > y),$$

but not the same meaning as

$$y > \mathbf{5} \rightarrow \forall x(\mathbf{2}x > x).$$

# Binding a variable by different quantifiers in a formula

Different occurrences of the same variable can be bound by different quantifiers:

$$\exists x(x > \mathbf{5}) \vee \forall x(\mathbf{2}x > x).$$

Again, the occurrences of $x$, bound by the first quantifier, have nothing to do with those bound by the second one.

For instance, the two $x$'s claimed to exist in the formula

$$\exists x(x > \mathbf{5}) \wedge \exists x(x < \mathbf{3}).$$

need not (and, in fact, cannot) be the same.

Thus, the formula above has the same meaning as each of

$$\exists y(y > \mathbf{5}) \wedge \exists x(x < \mathbf{3}),$$

$$\exists x(x > \mathbf{5}) \wedge \exists z(z < \mathbf{3}),$$

$$\exists y(y > \mathbf{5}) \wedge \exists z(z < \mathbf{3}).$$

# Nested bindings of a variable in a formula

Different bindings of the same variable can be nested, e.g.:

$$\forall x(x > \mathbf{5} \to \exists x(x < \mathbf{3})).$$

Again, the occurrences of $x$ in the subformula $\exists x(x < \mathbf{3})$ are bound by $\exists$ and not related to the first two occurrences of $x$, bound by $\forall$:

$$\forall x(x > \mathbf{5} \to \exists x(x < \mathbf{3})).$$

Thus, the formula above has the same meaning as each of

$$\forall x(x > \mathbf{5} \to \exists y(y < \mathbf{3})),$$

$$\forall z(z > \mathbf{5} \to \exists x(x < \mathbf{3})),$$

$$\forall z(z > \mathbf{5} \to \exists y(y < \mathbf{3})).$$

# Renaming of a bound variable in a formula

Using the same variable for different purposes in a formula can be confusing, and is often unwanted, so we may want to eliminate it.

Renaming of the variable $x$ in a formula $A$ is the substitution of *all occurrences of* $x$ *bound by the same occurrence of a quantifier* in $A$ with another variable, not occurring in $A$.

Example: the formula $(x > \mathbf{5}) \wedge \forall x((x > \mathbf{5}) \rightarrow \neg\exists x_1(x_1 < y))$ is a renaming of the formula $(x > \mathbf{5}) \wedge \forall x(x > \mathbf{5} \rightarrow \neg\exists x(x < y))$, while neither of the following formulae is a correct renaming:

$$(x_1 > \mathbf{5}) \wedge \forall x((x > \mathbf{5}) \rightarrow \neg\exists x(x < y)),$$

$$(x > \mathbf{5}) \wedge \forall x_1((x_1 > \mathbf{5}) \rightarrow \neg\exists x_1(x_1 < y)),$$

$$(x > \mathbf{5}) \wedge \forall x(x > \mathbf{5} \rightarrow \neg\exists y(y < y)).$$

Proposition: The result of renaming a variable in a formula is logically equivalent to that formula.

# Clean formulae

A formula $A$ is clean if no variable occurs both free and bound in $A$ and every two occurrences of quantifiers bind different variables.

Thus, $\exists x(x > \mathbf{5}) \wedge \exists y(y < z)$ is clean,

while $\exists x(x > \mathbf{5}) \wedge \exists y(y < x)$

and $\exists x(x > \mathbf{5}) \wedge \exists x(y < x)$ are not.

PROPOSITION: Every formula can be transformed into a clean formula by means of several consecutive renamings of variables.

E.g.,

$$(x > \mathbf{5}) \wedge \forall x((x > \mathbf{5}) \rightarrow \neg \exists x(x < y))$$

can be transformed into a clean formula as follows:

$$(x > \mathbf{5}) \wedge \forall x_1((x_1 > \mathbf{5}) \rightarrow \neg \exists x(x < y)),$$

$$(x > \mathbf{5}) \wedge \forall x_1((x_1 > \mathbf{5}) \rightarrow \neg \exists x_2(x_2 < y)).$$

# Substitution of a term for a variable in a formula

Unform substitution of a term $t$ for a variable $x$ in a formula $A$ means that all *free* occurrences $x$ in $A$ are simultaneously replaced by $t$. The result of the substitution is denoted $A[t/x]$.

Example: given the formula

$$A = \forall x (P(x, y) \to (\neg Q(y) \lor \exists y P(x, y)))$$

we have

$$A[f(y, z)/y] = \forall x (P(x, f(y, z)) \to (\neg Q(f(y, z)) \lor \exists y P(x, y))),$$

while

$$A[f(y, z)/x] = A$$

because $x$ does not occur free in $A$.

Intuitively, $A[t/x]$ is supposed to say about the individual denoted by $t$ the same as what $A$ says about the individual denoted by $x$.

Question: is that always the case?

Is a substitution of a term for a formula always 'safe'?

# Capture of a variable in substitution

The formula $A = \exists y(x < y)$ is true in $\mathcal{N}$ *for any value of* $x$.

However, $A[(y+\mathbf{1})/x] = \exists y(y+\mathbf{1} < y)$, which is false in $\mathcal{N}$.

Therefore, the formula $A[(y+\mathbf{1})/x]$ does not say about the term $y+\mathbf{1}$ the same as what $A$ says about $x$.

What went wrong?

The occurrence of $y$ in the term $y+\mathbf{1}$ got *captured* by the quantifier $\exists y$, because we mixed the free and the bound uses of $y$.

Capture: new occurrences of a variable $x$ in the scope of a quantifier $Qx$ introduced as a result of substitution of a term $t$ containing $x$ for another variable $y$ in a formula.

# Terms free for substitution for a variable in a formula

A term $t$ is free for (substitution for) a variable $x$ in a formula $A$, if no variable in $t$ is captured by a quantifier when $t$ is substituted for $x$ in $A$. Examples:

The term $f(x, y)$ is free for substitution for $y$ in the formula
$A = \forall x(P(x, z) \wedge \exists y Q(y)) \rightarrow P(y, z)$,

resulting in $A[f(x, y)/y] = \forall x(P(x, z) \wedge \exists y Q(y)) \rightarrow P(f(x, y), z)$,

but it is not free for substitution for $z$ in the formula $A$,
resulting in
$A[f(x, y)/z] = \forall x(P(x, f(x, y)) \wedge \exists y Q(y)) \rightarrow P(y, f(x, y))$,
because a capture occurs:
$A[f(x, y)/z] = \forall x(P(x, f(x, y)) \wedge \exists y Q(y)) \rightarrow P(y, f(x, y))$.

Note that every ground term (not containing variables), in particular every constant symbol, is always free for substitution.

# Renamings and substitutions in a formula

NB: renaming and substitution are different operations: renaming always acts on bound variables, while substitution always acts on free variables.

Also, renamings preserve the formula up to logical equivalence, while substitutions do not.

On the other hand, a suitable renaming of a formula can prepare it for a substitution, by rendering the term to be substituted free for such substitution.

For instance, the term $f(x, y)$ is not free for substitution for $y$ in

$$A = \forall x(P(x, y) \wedge \exists y Q(y)),$$

but it becomes free for such substitution after renaming of $A$ to

$$A' = \forall x'(P(x', y) \wedge \exists y Q(y)).$$