1.9: Substitutions and Substitutability

Suppose you knew that the sentence $\forall x \phi(x)$ was true in a particular structure $\mathfrak{A}$. Then, if $c$ is a constant symbol in the language, you would certainly expect $\phi(c)$ to be true in $\mathfrak{A}$ as well. What we have done is substitute the constant symbol $c$ for the variable $x$. This seems perfectly reasonable, although there are times when you do have to be careful.

Suppose that $\mathfrak{A} \models \forall x \exists y \neg(x = y)$. This sentence is, in fact, true in any structure $\mathfrak{A}$ such that $\mathfrak{A}$ has at least two elements. If we then proceed to replace the variable $x$ by the variable $u$, we get the statement $\exists y \neg(u = y)$, which will still be true in $\mathfrak{A}$, no matter what value we give to the variable $u$. If, however, we take our original formula and replace $x$ by $y$, then we find ourselves looking at $\exists y \neg(x = y)$, which will be false in any structure. So by a poor choice of substituting variable, we have changed the truth value of our formula. The rules of substitutability that we will discuss in this section are designed to help us avoid this problem, the problem of attempting to substitute a term inside a quantifier that binds a variable involved in the term.

We begin by defining exactly what we mean when we substitute a term $t$ for a variable $x$ in either a term $u$ or a formula $\phi$.

**Definition 1.8.1.** Suppose that $\mathfrak{A}$ models $\forall x \exists y \neg(x = y)$. This sentence is, in fact, true in any structure $\mathfrak{A}$ such that $\mathfrak{A}$ has at least two elements. If we then proceed to replace the variable $x$ by the variable $u$, we get the statement $\exists y \neg(u = y)$, which will still be true in $\mathfrak{A}$, no matter what value we give to the variable $u$. If, however, we take our original formula and replace $x$ by $y$, then we find ourselves looking at $\exists y \neg(x = y)$, which will be false in any structure. So by a poor choice of substituting variable, we have changed the truth value of our formula. The rules of substitutability that we will discuss in this section are designed to help us avoid this problem, the problem of attempting to substitute a term inside a quantifier that binds a variable involved in the term.

We begin by defining exactly what we mean when we substitute a term $t$ for a variable $x$ in either a term $u$ or a formula $\phi$.

**Definition 1.8.1.** Suppose that $\mathfrak{A}$ models $\forall x \exists y \neg(x = y)$. This sentence is, in fact, true in any structure $\mathfrak{A}$ such that $\mathfrak{A}$ has at least two elements. If we then proceed to replace the variable $x$ by the variable $u$, we get the statement $\exists y \neg(u = y)$, which will still be true in $\mathfrak{A}$, no matter what value we give to the variable $u$. If, however, we take our original formula and replace $x$ by $y$, then we find ourselves looking at $\exists y \neg(x = y)$, which will be false in any structure. So by a poor choice of substituting variable, we have changed the truth value of our formula. The rules of substitutability that we will discuss in this section are designed to help us avoid this problem, the problem of attempting to substitute a term inside a quantifier that binds a variable involved in the term.

We begin by defining exactly what we mean when we substitute a term $t$ for a variable $x$ in either a term $u$ or a formula $\phi$.

**Definition 1.8.1.** Suppose that $\mathfrak{A}$ models $\forall x \exists y \neg(x = y)$. This sentence is, in fact, true in any structure $\mathfrak{A}$ such that $\mathfrak{A}$ has at least two elements. If we then proceed to replace the variable $x$ by the variable $u$, we get the statement $\exists y \neg(u = y)$, which will still be true in $\mathfrak{A}$, no matter what value we give to the variable $u$. If, however, we take our original formula and replace $x$ by $y$, then we find ourselves looking at $\exists y \neg(x = y)$, which will be false in any structure. So by a poor choice of substituting variable, we have changed the truth value of our formula. The rules of substitutability that we will discuss in this section are designed to help us avoid this problem, the problem of attempting to substitute a term inside a quantifier that binds a variable involved in the term.

We begin by defining exactly what we mean when we substitute a term $t$ for a variable $x$ in either a term $u$ or a formula $\phi$.

**Definition 1.8.1.** Suppose that $\mathfrak{A}$ models $\forall x \exists y \neg(x = y)$. This sentence is, in fact, true in any structure $\mathfrak{A}$ such that $\mathfrak{A}$ has at least two elements. If we then proceed to replace the variable $x$ by the variable $u$, we get the statement $\exists y \neg(u = y)$, which will still be true in $\mathfrak{A}$, no matter what value we give to the variable $u$. If, however, we take our original formula and replace $x$ by $y$, then we find ourselves looking at $\exists y \neg(x = y)$, which will be false in any structure. So by a poor choice of substituting variable, we have changed the truth value of our formula. The rules of substitutability that we will discuss in this section are designed to help us avoid this problem, the problem of attempting to substitute a term inside a quantifier that binds a variable involved in the term.
In the fourth clause of the definition above and in the first two clauses of the next definition, the parentheses are not really there. However, we believe that no one can look at \( (u_1 \lor t^x) \) and figure out what it is supposed to mean. So the parentheses have been added in the interest of readability.

For example, if we let \( (t) \) be \( (g \left( c \right) \) and we let \( (u) \) be \( (f \left( x, y \right) + h \left( z, x, g \left( x \right) \right) \) then \( (u_t^x) \) is

\[
\left[ f \left( g \left( c \right), y \right) + h \left( z, g \left( c \right) , g \left( g \left( c \right) \right) \right) \right].
\]

The definition of substitution into a formula is also by recursion:

**Definition 1.8.2.** Suppose that \( (\phi) \) is an \( \langle \text{mathcal{L}} \rangle \)-formula, \( (t) \) is a term, and \( (x) \) is a variable. We define the formula \( (\phi_t^x) \) (read "\( (\phi) \) with \( (x) \) replaced by \( (t) \)"") as follows:

1. If \( (\phi) \equiv = u_1 u_2 \), then \( (\phi_t^x) \) is \( = \left( u_1 \right)_t^x \left( u_2 \right)_t^x \).
2. If \( (\phi) \equiv R u_1 u_2 \), then \( (\phi_t^x) \) is \( R \left( u_1 \right)_t^x \left( u_2 \right)_t^x \).
3. If \( (\phi) \equiv \neg \left( \alpha \right) \), then \( (\phi_t^x) \) is \( \neg \left( \alpha_t^x \right) \).
4. If \( (\phi) \equiv \left( \alpha \lor \beta \right) \), then \( (\phi_t^x) \) is \( \left( \alpha_t^x \lor \beta_t^x \right) \).
5. If \( (\phi) \equiv \left( \forall y \right) \left( \alpha \right) \), then \( (\phi_t^x) \) is \( \begin{cases} \phi \text{ if } x = y \\ \left( \forall y \right) \left( \alpha_t^x \right) \text{ otherwise} \end{cases} \).

As an example, suppose that \( (\phi) \) is the formula

\[
P \left( x, y \right) \rightarrow \left[ \left( \forall x \right) Q \left( g \left( x \right), z \right) \right] \lor \left( \forall y \right) \left( R \left( x, h \left( x \right) \right) \right].
\]

Then, if \( (t) \) is the term \( (g \left( c \right) \), we get

\[
\left[ P \left( x, y \right) \rightarrow \left[ \left( \forall x \right) Q \left( g \left( x \right), z \right) \right] \lor \left( \forall y \right) \left( R \left( x, h \left( x \right) \right) \right].\right.
\]

Having defined what we mean when we substitute a term for a variable, we will now define what it means for a term to be substitutable for a variable in a formula. The idea is that if \( (t) \) is substitutable for \( (x) \) in \( (\phi) \), we will not run into the problems discussed at the beginning of this section - we will not substitute a term in such a way that a variable contained in that term is inadvertently bound by a quantifier.

**Definition 1.8.3.** Suppose that \( (\phi) \) is an \( \langle \text{mathcal{L}} \rangle \)-formula, \( (t) \) is a term, and \( (x) \) is a variable. We way that \( (t) \) is substitutable for \( (x) \) in \( (\phi) \) if

1. \( (\phi) \) is atomic, or
2. \( (\phi) \equiv \neg \left( \alpha \right) \) and \( (t) \) is substitutable for \( (x) \) in \( (\alpha) \), or
3. \( (\phi) \equiv \left( \alpha \lor \beta \right) \) and \( (t) \) is substitutable for \( (x) \) in both \( (\alpha) \) and \( (\beta) \), or
4. \( \phi : \equiv \left( \forall y \right) \left( \alpha \right) \) and either
   
   (a) \( \left( x \right) \) is not free in \( \left( \phi \right) \), or
   (b) \( \left( y \right) \) does not occur in \( \left( t \right) \) and \( \left( t \right) \) is substitutable for \( \left( x \right) \) in \( \left( \alpha \right) \)

Notice that \( \phi_t^x \) is defined whether or not \( \left( t \right) \) is substitutable for \( \left( x \right) \) in \( \left( \phi \right) \). Usually we will not want to do a substitution unless we check for substitutability, but we have the ability to substitute whether or not it is a good idea. In the next chapter, however, you will often see that certain operations are allowed only if \( \phi_t^x \) is substitutable for \( \left( x \right) \) in \( \left( \phi \right) \). That restriction is there for good reason, as we will be concerned with preserving the truth of formulas after performing substitutions.

---

**Exercises**

1. For each of the following, write out \( u_t^x \):
   
   (a) \( u : \equiv \cos \left( x \right) \), \( t \) is \( \sin \left( y \right) \).
   (b) \( u : \equiv \left( y \right) \), \( t \) is \( \left( S \left( y \right) \right) \).
   (c) \( u : \equiv \sharp \left( x, y, z \right) \), \( t \) is \( 423 - w \).

2. For each of the following, first write out \( \phi_t^x \), then decide if \( \left( t \right) \) is substitutable for \( \left( x \right) \) in \( \left( \phi \right) \), and then (if you haven’t already) use the definition of substitutability to justify your conclusions.
   
   (a) \( \phi : \equiv \left( \forall x \right) \left( x = y \rightarrow Sx = Sy \right) \), \( t \) is \( S0 \).
   (b) \( \phi : \equiv \left( \forall y \right) \left( x = y \rightarrow Sx = Sy \right) \), \( t \) is \( \left( S \left( y \right) \right) \).
   (c) \( \phi : \equiv x = y \rightarrow \left( \forall x \right) \left( Sx = Sy \right) \), \( t \) is \( \left( S \left( y \right) \right) \).

3. Show that if \( \left( t \right) \) is variable-free, then \( \left( t \right) \) is always substitutable for \( \left( x \right) \) in \( \left( \phi \right) \).

4. Show that \( \left( x \right) \) is always substitutable for \( \left( x \right) \) in \( \left( \phi \right) \).

5. Prove that if \( \left( x \right) \) is not free in \( \left( \psi \right) \), then \( \psi_t^x \) is \( \left( \psi \right) \).

6. You might think that \( \left( \phi_t^x \right) \) is \( \phi \), but a moment’s thought will give you an example to show that this doesn’t always work. (What if \( \left( y \right) \) is free in \( \left( \phi \right) \)?) Find an example that shows that even if \( \left( y \right) \) is not free in \( \left( \phi \right) \), we can still have \( \phi_t^x \) different from \( \phi \). Under what conditions do we know that \( \phi_t^x \) is \( \phi \)?

7. Write a computer program (in your favorite language, or in pseudo-code) that accepts as input a formula \( \left( \phi \right) \), a variable \( \left( x \right) \), and a term \( \left( t \right) \) and outputs “yes” or “no” depending on whether or not \( \left( t \right) \) is substitutable for \( \left( x \right) \) in \( \left( \phi \right) \).