3.6 The Semantics of Predicate logic

The semantics of predicate logic is concerned with how the meanings of sentences, which just as in propositional logic, amount to their truth values, depend on the meanings of the parts of which they are composed. But since the parts need not themselves be sentences, or even formulas—they may also be predicate letters, constants, or variables—we will not be able to restrict ourselves to truth values in interpreting languages of predicate logic. We will need functions other than the valuations we encountered within propositional logic, and ultimately the truth values of sentences will have to reduce to the interpretations of the constants and predicate letters and everything else which appears in them. Valuations, however, retain a central role, and it is instructive to start off just with them and to build up the rest of the apparatus for the interpretation of predicate logic from there. One first attempt to do this is found in the following definition, in which valuations are extended to the languages of predicate logic. It turns out that this is in itself not enough, so remember that the definition is only preliminary.

Definition 5

A valuation for a language L of predicate logic is a function with the sentences in L as its domain and $\{0, 1\}$ as its range, and such that:

(i) $V(\neg\phi) = 1$ iff $V(\phi) = 0$;
(ii) $V(\phi \wedge \psi) = 1$ iff $V(\phi) = 1$ and $V(\psi) = 1$;
(iii) $V(\phi \vee \psi) = 1$ iff $V(\phi) = 1$ or $V(\psi) = 1$;

(iv) $V(\phi \rightarrow \psi) = 1$ iff $V(\phi) = 0$ or $V(\psi) = 1$;
(v) $V(\phi \leftrightarrow \psi) = 1$ iff $V(\phi) = V(\psi)$;
(vi) $V(\forall x\phi) = 1$ iff $V([c/x]\phi) = 1$ for all constants c in L;
(vii) $V(\exists x\phi) = 1$ iff $V([c/x]\phi) = 1$ for at least one constant c in L.

The idea is that $\forall x\phi$ is true just in case $[c/x]\phi$ is true for every c in L, and that $\exists x\phi$ is true just in case $[c/x]\phi$ is true for at least one c in L. This could be motivated with reference to (90) and (91). For (90) is true just in case every substitution of the name of an individual human being into the open space in (91) results in a true sentence. And (92) is true just in case there is at least one name the substitution of which into (91) results in a true sentence.

(90) Everyone is friendly.

(91) . . . is friendly.

(92) Someone is friendly.

One thing should be obvious right from the start: in formal semantics, as in informal semantics, it is necessary to introduce a *domain of discourse*. For (90) may very well be true if the inhabitants of the Pacific state of Hawaii are taken as the domain, but untrue if all human beings are included. So in order to judge the truth value of (90), it is necessary to know what we are talking about, i.e., what the domain of discourse is. Interpretations of a language L of predicate logic will therefore always be with reference to some domain set D. It is usual to suppose that there is always at least one thing to talk about—so by convention, the domain is not empty.

3.6.1 *Interpretation Functions*

We will also have to be more precise about the relationship between the constants in L and the domain D. For if we wish to establish the truth value of (90) in the domain consisting of all inhabitants of Hawaii, then the truth value of *Liliuokalani is friendly* is of importance, while the truth value of *Gorbachev is friendly* is of no importance at all, since Liliuokalani is the name of an inhabitant of Hawaii (in fact she is, or at least was, one of its queens), while Gorbachev, barring unlikely coincidences, is not. Now it is a general characteristic of a proper name in natural language that it refers to some fixed thing. This is not the case in formal languages, where it is necessary to stipulate what the constants refer to. So an interpretation of L will have to include a specification of what each constant in L refers to. In this manner, constants refer to entities in the domain D, and as far as predicate logic is concerned, their meanings can be restricted to the entities to which they refer. The interpretation of the constants in L will therefore be an attribution of some entity in D to each of them, that is, a function with the set of constants in L as its domain and D as its range. Such functions are called *interpretation functions*.

I(c) is called the *interpretation of a constant c*, or its *reference* or its *denotation*, and if e is the entity in D such that I(c) = e, then c is said to be one of e's *names* (e may have several different names).

Now we have a domain D and an interpretation function I, but we are not quite there yet. It could well be that

(93) Some are white.

is true for the domain consisting of all snowflakes without there really being any English sentence of the form *a is white* in which *a* is the name of a snowflake. For although snowflakes tend to be white, it could well be that none of them has an English name. It should be clear from this that definition 5 does not work as it is supposed to as soon as we admit domains with unnamed elements. So two approaches are open to us:

A. We could stick to definition 5 but make sure that all objects in our domains have names. In this case, it will sometimes be necessary to add constants to a language if it does not contain enough constants to give a unique name to everything in some domain that we are working with.

B. We replace definition 5 by a definition which will also work if some entities lack names.

We shall take both approaches. Approach B seems preferable, because of A's intuitive shortcomings: it would be strange if the truth of a sentence in predicate logic were to depend on a contingency such as whether or not all of the entities being talked about had a name. After all, the sentences in predicate logic do not seem to be saying these kinds of things about the domains in which they are interpreted. But we shall also discuss A, since this approach, where it is possible, is simpler and is equivalent to B.

3.6.2 *Interpretation by Substitution* (1)

First we shall discuss approach A, which may be referred to as *the interpretation of quantifiers by substitution*. We shall now say more precisely what we mean when we say that each element in the domain has a name in L. Given the terminology introduced in §2.4, we can be quite succinct: the interpretation function I must be a function from the constants in L *onto* D. This means that for every element d in D, there is at least one constant c in L such that I(c) = d, i.e., c is a name of d. So we will only be allowed to make use of the definition if I is a function onto D.

But even this is not wholly satisfactory. So far, the meaning of predicate letters has only been given syncategorematically. This can be seen clearly if the question is transplanted into natural language: definition 5 enables us to know the meaning of the word *friendly* only to the extent that we know which sentences of the form *a is friendly* are true. If we want to give a direct, categorematic interpretation of *friendly*, then the interpretation will have to be such that the truth values of sentences of the form *a is friendly* can be deduced from it. And that is the requirement that can be placed on it, since we have restricted the meanings of sentences to their truth values. As a result, the only thing which matters as far as sentences of the form *a is friendly* are concerned is their truth values. An interpretation which establishes which people are friendly and which are not will satisfy this requirement. For example, *Gorbachev is friendly* is true just in case Gorbachev is friendly, since *Gorbachev* is one name for the man Gorbachev. Thus we can establish which people are friendly and which are not just by taking the set of all friendly people in our domain as the interpretation of *friendly*. In general then, as the interpretation I(A) of a unary predicate letter A we take the set of all entities e in D such that for some constant a, Aa is true and I(a) = e. So I(A) = {I(a) | Aa is true} or, in other words, Aa is true just in case I(a) ∈ I(A).

Interpreting A as a set of entities is not the only approach open to us. We might also interpret A as a property and determine whether a given element of D has this property. Indeed, this seems to be the most natural interpretation. If it is a predicate letter, we would expect A to refer to a property. What we have done here is to take, not properties themselves, but the sets of all things having them, as the interpretations of unary predicate letters. This approach may be less natural, but it has the advantage of emphasizing that in predicate logic the only thing we need to know in order to determine the truth or falsity of a sentence asserting that something has some property is which of the things in the domain have that property. It does not matter, for example, how we know this or whether things could be otherwise. As far as truth values are concerned, anything else which may be said about the property is irrelevant. If the set of friendly Hawaiians were to coincide precisely with the set of bald ones, then in this approach, *friendly* and *bald* would have the same meaning, at least if we took the set of Hawaiians as our domain. We say that predicate letters are *extensional* in predicate logic. It is characteristic of modern logic that such restrictions are explored in depth and subsequently relaxed. More than extensional meaning is attributed to expressions, for example, in *intensional* logical systems, which will be studied in volume 2.

To continue with approach A, and assuming that I is a function onto D as far as the constants are concerned, we turn to the interpretations of binary predicate letters. Just as with unary predicates, the interpretation of any given binary predicate B does not have to do anything more than determine the d and e in D for which Bab is true if I(a) = d and I(b) = e. This can be done by interpreting B as a set of ordered pairs ⟨d, e⟩ in D^2 and taking Bab to be true if I(a) = d and I(b) = e. The interpretation must consist of ordered pairs, because the order of a and b matters. The interpretation of B is, in other words, a subset of D^2, and we have I(B) = {⟨I(a), I(b)⟩ | Bab is true} or equivalently, Bab is true just in case ⟨I(a), I(b)⟩ ∈ I(B). Here too it may seem more intuitive to interpret B as a relation on D and to say that Bab is true if and only if I(a) and I(b) bear this relation to each other. For reasons already mentioned, however, we prefer the extensional approach and interpret a binary predicate letter

not as a relation itself but as the set of ordered pairs of domain elements which (in the order they have in the pairs) have this relation to each other. And we thus have the principle of extensionality here too: two relations which hold for the same ordered pairs are identical. Ternary predicates and predicates of all higher arities are given an analogous treatment. If C is a ternary predicate letter, then I(C) is a subset of D^3, and if C is an *n*-ary predicate, then I(C) is a subset of D^n. We shall now summarize all of this in the following two definitions:

Definition 6

A *model* **M** for a language L of predicate logic consists of a domain D (this being a nonempty set) and an interpretation function I which is defined on the set of constants and predicate letters in the vocabulary of L and which conforms to the following requirements:

(i) if c is a constant in L, then I(c) ∈ D;
(ii) if B is an *n*-ary predicate letter in L, then $I(B) \subseteq D^n$.

Definition 7

If **M** is a model for L whose interpretation function I is a function of the constants in L onto the domain D, then V_M, *the valuation V based on* **M**, is defined as follows:

(i) If $Aa_1 \ldots a_n$ is an atomic sentence in L, then $V_M(Aa_1 \ldots a_n) = 1$ if and only if $\langle I(a_1), \ldots, I(a_n)\rangle \in I(A)$.
(ii) $V_M(\neg\phi) = 1$ iff $V_M(\phi) = 0$.
(iii) $V_M(\phi \wedge \psi) = 1$ iff $V_M(\phi) = 1$ and $V_M(\psi) = 1$.
(iv) $V_M(\phi \vee \psi) = 1$ iff $V_M(\phi) = 1$ or $V_M(\psi) = 1$.
(v) $V_M(\phi \to \psi) = 1$ iff $V_M(\phi) = 0$ or $V_M(\psi) = 1$. $\phi = \psi$
(vi) $V_M(\phi \leftrightarrow \psi) = 1$ iff $V_M(\phi) = V_M(\psi)$.
(vii) $V_M(\forall x\phi) = 1$ iff $V_M([c/x]\phi) = 1$ for all constants c in L.
(viii) $V_M(\exists x\phi) = 1$ iff $V_M([c/x]\phi) = 1$ for at least one constant c in L.

If $V_M(\phi) = 1$, then ϕ is said to be *true* in model **M**.

If the condition that I be a function onto D is not fulfilled, then approach B will still enable us to define a suitable valuation function V_M, though this function will no longer fulfill clauses (vii) and (viii) of definition 7. Before showing how this can be done, we shall first give a few examples to illustrate method A.

Example 1

We turn the key to a translation into a model.
Key: Lxy: x loves y; domain: Hawaiians.
We take H, the set of all Hawaiians, as the domain of model **M**. Besides the binary predicate L, our language must contain enough constants to give each

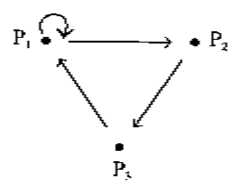
Hawaiian a name; $a_1, \ldots, a_{1,000,000}$ should be enough. Now for each *i* from 1 to 1,000,000 inclusive, a_i must be interpreted as a Hawaiian: $I(a_i) \in H$, and this in such a way that for each Hawaiian h there is some a_h which is interpreted as that Hawaiian, that is, for which $I(a_h) = h$. The interpretation of L is the following subset of H^2, i.e., the set of pairs of Hawaiians: $\{\langle d, e\rangle \mid d$ loves $e\}$. Lets us now determine the truth value of $\exists x \exists y(Lxy \wedge Lyx)$, which is the translation of *some people love each other*. Suppose that John loves Mary, that Mary's love for John is no less, that $I(a_{26})$ is Mary, and that $I(a_{27})$ is John. Then $\langle I(a_{26}), I(a_{27})\rangle \in I(L)$, and $\langle I(a_{27}), I(a_{26})\rangle \in I(L)$. According to definition 7i, we have $V_M(La_{26}a_{27}) = 1$ and $V_M(La_{27}a_{26}) = 1$, so that according to definition 7iii, we have $V_M(La_{26}a_{27} \wedge La_{27}a_{26}) = 1$. One application of definition 7viii now gives us $V_M(\exists y(La_{26}y \wedge Lya_{26})) = 1$, and a second gives us $V_M(\exists x\exists y(Lxy \wedge Lyx)) = 1$. Of course, it doesn't matter at all which constants are interpreted as which people. We could have shown that $V_M(\exists x\exists y(Lxy \wedge Lyx)) = 1$ just as well if $I(a_2)$ had been John and $I(a_9)$ had been Mary. This is a general fact: the truth of a sentence lacking constants is in any model independent of the interpretations of the constants in that model—with the proviso that everything in the domain has a name. A comment such as this should of course be proved, but we do not have the space here.

It is perhaps worth pointing out at this stage that semantics is not really concerned with finding out which sentences are in fact true and which are false. One's ideas about this are unlikely to be influenced much by the analysis given here. Essentially, semantics is concerned with *the ways the truth values of sentences depend on the meanings of their parts and the ways the truth values of different sentences are related*. This is analogous to the analysis of the notion of grammaticality in linguistics. It is assumed that it is clear which expressions are grammatical and which are not; the problem is to conceive a systematic theory on the subject.

The following examples contain a few extremely simple mathematical structures. We shall leave off the index M in V_M if it is clear what model the valuation is based on.

Example 2

The language we will interpret contains three constants, a_1, a_2, and a_3, and the binary predicate letter R. The domain D of the model is the set of points $\{P_1, P_2, P_3\}$ represented in figure (94).

(94)


The constants are interpreted as follows: $I(a_1) = P_1$; $I(a_2) = P_2$; and $I(a_3) = P_3$. The interpretation of R is the relation holding between any two not necessarily different points with an arrow pointing from the first to the second. So the following interpretation of R can be read from figure (94): $I(R) = \{\langle P_1, P_1\rangle, \langle P_1, P_2\rangle, \langle P_2, P_3\rangle, \langle P_3, P_1\rangle\}$. Representing this by means of a key, Rxy: there is an arrow pointing from x to y. It is directly obvious that $V(Ra_1a_2) = 1$, $V(Ra_2a_3) = 1$, $V(Ra_3a_1) = 1$, and $V(Ra_1a_1) = 1$; in all other cases, $V(Rbc) = 0$, so that, for example, $V(Ra_2a_1) = 0$ and $V(Ra_3a_3) = 0$. We shall now determine the truth value of $\forall x\exists yRxy$ (which means *every point has an arrow pointing away from it*).

(a) $V(\exists yRa_1y) = 1$ follows from $V(Ra_1a_2) = 1$ with definition 7viii;
(b) $V(\exists yRa_2y) = 1$ follows from $V(Ra_2a_3) = 1$ with definition 7viii;
(c) $V(\exists yRa_3y) = 1$ follows from $V(Ra_3a_1) = 1$ with definition 7viii.

From (a), (b), and (c), we can now conclude that $V(\forall x\exists yRxy) = 1$ with definition 7vii. The truth value of $\forall x\exists yRyx$ (which means *every point has an arrow pointing to it*) can be determined in just the same way:

(d) $V(\exists yRya_1) = 1$ follows from $V(Ra_3a_1) = 1$ with definition 7viii;
(e) $V(\exists yRya_2) = 1$ follows from $V(Ra_1a_2) = 1$ with definition 7viii;
(f) $V(\exists yRya_3) = 1$ follows from $V(Ra_2a_3) = 1$ with definition 7viii.

From (d), (e), and (f), we conclude that $V(\forall x\exists yRyx) = 1$ with definition 7vii.

Finally, we shall determine the truth value of $\exists x\forall yRxy$ (which means: *there is a point from which arrows go to all other points*):

(g) $V(\forall yRa_1y) = 0$ follows from $V(Ra_1a_3) = 0$ with definition 7vii;
(h) $V(\forall yRa_2y) = 0$ follows from $V(Ra_2a_2) = 0$ with definition 7vii;
(i) $V(\forall yRa_3y) = 0$ follows from $V(Ra_3a_3) = 0$ with definition 7vii.

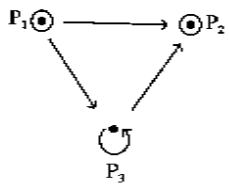
From (g), (h), and (i), we can now conclude that $V(\exists x\forall yRxy) = 0$ with definition 7viii.

Example 3

We consider a language with a unary predicate letter E, a binary predicate letter L, and constants $a_0, a_1, a_2, a_3, \ldots$. We take N, the set $\{0, 1, 2, 3, \ldots\}$ of natural numbers, as our domain. We choose $V(a_i) = i$ for every i and interpret E as the set of even numbers, so that $I(E) = \{0, 2, 4, 6, \ldots\}$. We interpret L as $<$, so that $I(L) = \{\langle m, n\rangle \mid m$ less than $n\}$. As true sentences we then have, for example, Ea_2, La_4a_5, and $\forall x\exists y(Lxy \wedge \neg Ey)$ (these mean *2 is even*, *4 is less than 5*, and *for every number there is a larger number which is odd*, respectively). We shall expand on the last of these. Consider any number m. This number must be either even or odd.

If m is even, then m + 1 is odd, so that $V(Ea_{m+1}) = 0$ and $V(\neg Ea_{m+1}) = 1$. We also have $V(La_ma_{m+1}) = 1$, since $m < m + 1$. From this we may conclude that $V(La_ma_{m+1} \wedge \neg Ea_{m+1}) = 1$, and finally that $V(\exists y(La_my \wedge \neg Ey)) = 1$.

If, on the other hand, m is odd, then m + 2 is odd too, so that $V(Ea_{m+2}) = 0$ and $V(\neg Ea_{m+2}) = 1$. We also have $V(La_ma_{m+2}) = 1$, since $m < m + 2$, and thus $V(La_m a_{m+2} \wedge \neg Ea_{m+2}) = 1$, so that we have $V(\exists y(La_m y \wedge \neg Ey)) = 1$ in this case as well. Since this line of reasoning applies to an arbitrary number m, we have for every a_m: $V(\exists y(La_m y \wedge \neg Ey)) = 1$. Now we have shown that $V(\forall x \exists y(Lxy \wedge \neg Ey)) = 1$.

(95) 

Exercise 7

Model **M** is given in figure (95). The language has three constants a_1, a_2, and a_3 interpreted as the points P_1, P_2, and P_3, a unary predicate letter A interpreted as the predicate that applies to a point if it has a circle around it, and a binary predicate letter R to be interpreted as in example 2.

(a) Describe exactly the interpretation function I of the model **M**.

(b) Determine on the basis of their meaning the truth or falsity of the following sentences on model **M** and then justify this in detail, using definition 7:

(i) $\exists x \exists y \exists z(Rxy \wedge Ay \wedge Rxz \wedge \neg Az)$.
(ii) $\forall x Rxx$.
(iii) $\forall x(Rxx \leftrightarrow \neg Ax)$.
(iv) $\exists x \exists y(Rxy \wedge \neg Ax \wedge \neg Ay)$.
(v) $\forall x(Rxx \rightarrow \exists y(Rxy \wedge Ay))$.
(vi) $\forall x(Ax \rightarrow \exists y\, Rxy)$.
(vii) $\exists x \exists y(Rxy \wedge \neg Ryx \wedge \exists z(Rxz \wedge Rzy))$.

3.6.3 *Interpretation by means of assignments*

We have now come to the explication of **approach B.** To recapitulate: we have a language L, a domain D, and an interpretation function I which maps all of L's constants into D but which is not necessarily a function *onto* D. That is, we have no guarantee that everything in the domain has some constant as its name. This means that the truth of sentences $\exists x\phi$ and $\forall x\phi$ can no longer be reduced to that of sentences of the form $[c/x]\phi$. Actually, this reduction is not that attractive anyway, if we wish to take the principle of compositionality strictly. This principle requires that the meaning (i.e., the truth value) of an expression be reducible to that of its composite parts. But sentences $\exists x\phi$ and $\forall x\phi$ do not have sentences of the form $[c/x]\phi$ as their component parts, because they are obtained by placing a quantifier in front of a formula ϕ, which normally has a free variable x and therefore is not even another sentence. What this means is that we will have to find some way to attach meanings to formulas in general; we can no longer restrict ourselves to the special case of sentences.

We have reserved the name *propositional function* for formulas with free variables, in part because sentences can be obtained by replacing the free variables with constants, and in part because a formula with free variables does not seem to express a proposition but rather a property or a relation. But we could also take a different view and say that formulas with free variables express propositions just as much as sentences do, only these propositions are about unspecified entities. This would be why they are suited to express properties and relations.

In order to see how a meaning can be attached to these kinds of formulas, let us return again to (96) (=(93)):

(96) Some are white.

This was to be interpreted in the domain consisting of all snowflakes. What we want to do is determine the truth value of (96) with reference to the meaning of *x is white* interpreted in the domain consisting of all snowflakes. Now x, as we have emphasized, has no meaning of its own, so it must not refer to some fixed entity in the domain as if it were a constant. This may be compared with the way pronouns refer in sentences like *he is white* and *she is black*. But precisely for this reason, it may make sense to consider x as the *temporary* name of some entity. The idea is to consider model **M** together with an extra attribution of denotations to x and all the other variables; x will receive a temporary interpretation as an element in D. It is then quite easy to determine the truth value of (96): (96) is true if and only if there is some attribution of a denotation in the domain of all snowflakes to x, such that *x is white* becomes a true sentence. In other words, (96) is true just in case there is some snowflake which, if it is given the name x, will turn *x is white* into a true sentence—and that is exactly what we need.

The meaning of

(97) They are all black.

in the domain consisting of all snowflakes can be handled in much the same way: (97) is true if and only if *every* attribution of a denotation to x in this domain turns *x is black* into a true sentence. Analyzing this idea brings up more technical problems than most things we have encountered so far.

In order to determine the truth value of a sentence like $\exists x \exists y(Hxy \wedge Hyx)$, it is necessary to work back (in two steps) to the meaning of its subformula $Hxy \wedge Hyx$, which has two free variables. Obviously since no limitation is placed on the length of formulas, such subformulas can contain any number of free variables. This means that we must deal with the meanings of formulas with any number of free variables in order to determine the truth values of sentences. What matters is the truth value of a formula once all of its free variables have been given a temporary denotation, but it turns out that it is easiest to give all free variables a denotation at the same time. It is unnecessarily difficult to keep track of what free variables each formula has and to assign denotations to them. What we do is use certain functions called *assignments* which have the set of all variables in the language as their domain, and D, the domain of the model, as their range.

We will now describe the truth values a model **M** gives to the formulas of L under an assignment g by means of a valuation function $V_{M,g}$. This function will be defined by modifying conditions (i)–(viii) of definition 7 above.

The complications begin with clause (i). There is no problem as long as we deal with an atomic formula containing only variables and no constants: we are then dealing with $V_{M,g}(Ax_1 \ldots x_n)$, and it is clear that we wish to have $V_{M,g}(Ax_1 \ldots x_n) = 1$ if and only if $\langle g(x_1), \ldots, g(x_n)\rangle \in I(A)$, since the only difference from the earlier situation is that we have an assignment g attributing denotations to variables instead of an interpretation I attributing denotations to constants. But it becomes more difficult to write things up properly for formulas of the form $At_1 \ldots t_n$, in which $t_1, \ldots, t_n$ may be either constants or variables. What we do is introduce *term* as the collective name for the constants and variables of L. We first define what we mean by $[\![t]\!]_{M,g}$, the interpretation of a term t in a model **M** under an assignment g.

Definition 8

$[\![t]\!]_{M,g} = I(t)$ if t is a constant in L, and
$[\![t]\!]_{M,g} = g(t)$ if t is a variable.

Now we can generalize (i) in definition 7 to:

$$V_{M,g}(At_1 \ldots t_n) = 1 \text{ iff } \langle [\![t_1]\!]_{M,g}, \ldots, [\![t_n]\!]_{M,g}\rangle \in I(A).$$

It is clear that the value of $V_{M,g}(At_1 \ldots t_n)$ does not depend on the value of g(y) if y does not appear among the terms $t_1, \ldots, t_n$.

Clauses (ii) to (vi) in definition 7 can be transferred to the definition of $V_{M,g}$ without modification. The second clause we have to adapt is (viii), the clause for $V_{M,g}(\exists y\phi)$. Note that ϕ may have free variables other than y. Let us return to the model given in example 1, only this time for a language lacking constants. We take Lxy as our ϕ. Now how is $V_{M,g}(\exists yLxy)$ to be defined? Under an assignment g, x is treated as if it denotes g(x), so $\exists yLxy$ means that g(x) loves someone. So the definition must result in $V_{M,g}(\exists yLxy) = 1$ if and only if there is a $d \in H$ such that $\langle g(x), d\rangle \in I(L)$. The idea was to reduce the meaning of $\exists yLxy$ to the meaning of Lxy. But we cannot take $V_{M,g}(\exists yLxy) = 1$ if and only if $V_{M,g}(Lxy) = 1$, since $V_{M,g}(Lxy) = 1$ if and only if $\langle g(x), g(y)\rangle \in I(L)$, that is, if and only if g(x) loves g(y). For it may well be that g(x) loves someone without this someone being g(y). The existential quantifier forces us to consider assignments other than g which only differ from g in the value which they assign to y, since the denotation of x may clearly not be changed.

On the one hand, if there is an assignment g′ which differs from g only in the value it assigns to y and such that $V_{M,g'}(Lxy) = 1$, then $\langle g'(x), g'(y)\rangle \in I(L)$, and thus, because $g(x) = g'(x)$, $\langle g(x), g'(y)\rangle \in I(L)$. So for some $d \in H$, $\langle g(x), d\rangle \in I(L)$. On the other hand, if there is some $d \in H$ such that $\langle g(x), d\rangle \in I(L)$, then it can easily be seen that there is always an assignment g′ such that $V_{M,g'}(Lxy) = 1$. Choose g′, for example, the assignment obtained by taking g and then just changing the value assigned to y to d. Then $\langle g'(x), g'(y)\rangle \in I(L)$, and so $V_{M,g'}(Lxy) = 1$. This argument can be repeated for any given formula, so now we can give a first version of the new clause for existential formulas. It is this: $V_{M,g}(\exists yLxy) = 1$ if and only if there is a g′ which differs from g only in its value for y and for which $V_{M,g'}(Lxy) = 1$. So g′ is uniquely determined by g, and the value g′ is assigned to the variable y. This means that we can adopt the following notation: we write g[y/d] for g′ if this assignment assigns d to y and assigns the same values as g to all the other variables. (Note that c in the notation $[c/x]\phi$ refers to a constant in L, whereas the d in g[y/d] refers to an entity in the domain; the first expression refers to the result of a syntactic operation, and the second does not.) The assignments g[y/d] and g tend to differ. But that is not necessarily the case, since they are identical if $g(y) = d$. So now we can give the final version of the new clause for existential formulas. It is this:

$$V_{M,g}(\exists y\phi) = 1 \text{ iff there is a } d \in D \text{ such that } V_{M,g[y/d]}(\phi) = 1.$$

A similar development can be given for the new clause for the universal quantifier. So now we can complete this discussion of the B approach by giving the following definition. It is well known as Tarski's truth definition, in honor of the mathematician A. Tarski who initiated it; it is a generalization of definition 7. Although clauses (ii)–(vi) are not essentially changed, we give the definition in full for ease of reference.

Definition 9

If **M** is a model, D is its domain, I is its interpretation function, and g is an assignment into D, then

(i) $V_{M,g}(At_1 \ldots t_n) = 1$ iff $\langle [\![t_1]\!]_{M,g}, \ldots, [\![t_n]\!]_{M,g}\rangle \in I(A)$;
(ii) $V_{M,g}(\neg\phi) = 1$ iff $V_{M,g}(\phi) = 0$;
(iii) $V_{M,g}(\phi \wedge \psi) = 1$ iff $V_{M,g}(\phi) = 1$ and $V_{M,g}(\psi) = 1$;
(iv) $V_{M,g}(\phi \vee \psi) = 1$ iff $V_{M,g}(\phi) = 1$ or $V_{M,g}(\psi) = 1$;
(v) $V_{M,g}(\phi \to \psi) = 1$ iff $V_{M,g}(\phi) = 0$ or $V_{M,g}(\psi) = 1$;
(vi) $V_{M,g}(\phi \leftrightarrow \psi) = 1$ iff $V_{M,g}(\phi) = V_{M,g}(\psi)$;
(vii) $V_{M,g}(\forall x\phi) = 1$ iff for all $d \in D$, $V_{M,g[x/d]}(\phi) = 1$;
(viii) $V_{M,g}(\exists x\phi) = 1$ iff there is at least one $d \in D$ such that $V_{M,g[x/d]}(\phi) = 1$.

We now state a few facts about this definition which we shall not prove. First, the only values of g which $V_{M,g}(\phi)$ is dependent on are the values which g assigns to variables which occur as free variables in ϕ; so ϕ has the same value for every g in the extreme case in which ϕ is a sentence. This means that for sentences ϕ we can just write $V_M(\phi)$. Consequently, it holds for sentences ϕ that if ϕ is true with respect to some g, then it is true with respect to all g. If all elements of the domain of **M** have names, then for any sentence ϕ, approach A and approach B give the same values for $V_M(\phi)$. In such cases then, either can be taken. We shall now return to the examples given in connection with approach A, and reconsider them with B.

Example 1

There is just a single binary predicate letter L in the language; the domain is H, the set of all Hawaiians; $I(L) = \{\langle d, e\rangle \in H^2 | d \text{ loves } e\}$, and John and Mary are two members of the domain who love one another. We now define g(x) = John and g(y) = Mary; we complete g by assigning the other variables at random. Then $V_{M,g}(Lxy) = 1$, since $\langle [\![x]\!]_{M,g}, [\![y]\!]_{M,g}\rangle = \langle g(x), g(y)\rangle = \langle \text{John}, \text{Mary}\rangle \in I(L)$. Analogously, $V_{M,g}(Lyx) = 1$, so that we also have $V_{M,g}(Lxy \wedge Lyx) = 1$. This means that $V_{M,g}(\exists y(Lxy \wedge Lyx)) = 1$, since g = g[y/Mary], and that $V_{M,g}(\exists x \exists y(Lxy \wedge Lyx)) = 1$ too, since g = g[x/John].

Example 2

There is just a single binary predicate letter R in the language; the domain is $\{P_1, P_2, P_3\}$; $I(R) = \{\langle P_1, P_2\rangle, \langle P_1, P_1\rangle, \langle P_2, P_3\rangle, \langle P_3, P_1\rangle\}$. Now for an arbitrary g we have:

if $g(x) = P_1$, then $V_{M,g[y/P_2]}(Rxy) = 1$, since $\langle P_1, P_2\rangle \in I(R)$;
if $g(x) = P_2$, then $V_{M,g[y/P_3]}(Rxy) = 1$, since $\langle P_2, P_3\rangle \in I(R)$;
if $g(x) = P_3$, then $V_{M,g[y/P_1]}(Rxy) = 1$, since $\langle P_3, P_1\rangle \in I(R)$.

This means that for every g there is a $d \in \{P_1, P_2, P_3\}$ such that $V_{M,g[y/d]}(Rxy) = 1$. This means that $V_{M,g}(\exists yRxy) = 1$. Since this holds for an arbitrary g, we may conclude that $V_{M,g[x/d]}(\exists yRxy) = 1$ for every $d \in D$. We have now shown that $V_{M,g}(\forall x \exists yRxy) = 1$. That $V_{M,g}(\forall x \exists yRyx) = 1$ can be shown in the same way.

Now for the truth value of $\exists x \forall yRxy$. For arbitrary g, we have:

if $g(x) = P_1$, then $V_{M,g[y/P_3]}(Rxy) = 0$, since $\langle P_1, P_3\rangle \notin I(R)$;
if $g(x) = P_2$, then $V_{M,g[y/P_2]}(Rxy) = 0$, since $\langle P_2, P_2\rangle \notin I(R)$;
if $g(x) = P_3$, then $V_{M,g[y/P_3]}(Rxy) = 0$, since $\langle P_3, P_3\rangle \notin I(R)$.

This means that for every g there is a $d \in \{P_1, P_2, P_3\}$ such that $V_{M,g[y/d]}(Rxy) = 0$. From this it is clear that for every g we have $V_{M,g}(\forall yRxy) = 0$, and thus that for every $d \in D$, $V_{M,g[y/d]}(\forall yRxy) = 0$; and this gives $V_{M,g}(\exists x \forall yRxy) = 0$.

Example 3

The language contains a unary predicate letter E and a binary predicate letter L. The domain of our model **M** is the set N, $I(E) = \{0, 2, 4, 6, \ldots\}$, and I(L) $= \{\langle m, n\rangle | m < n\}$. Now let g be chosen at random. Then there are two possibilities:

(a) g(x) is an even number. In that case g(x) + 1 is odd, so that $g(x) + 1 \notin I(E)$, from which it follows that $V_{M,g[y/g(x)+1]}(Ey) = 0$ and that $V_{M,g[y/g(x)+1]}(\neg Ey) = 1$. Furhermore, $\langle g(x), g(x) + 1\rangle \in I(L)$, and therefore $V_{M,g[y/g(x)+1]}(Lxy) = 1$, so that we have $V_{M,g[y/g(x)+1]}(Lxy \wedge \neg Ey) = 1$.

(b) g(x) is an odd number. In that case, g(x) + 2 is an odd number too. From this it follows, as in (a), that $V_{M,g[y/g(x)+2]}(Lxy \wedge \neg Ey) = 1$. In both cases, then, there is an $n \in N$ such that $V_{M,g[y/n]}(Lxy \wedge \neg Ey) = 1$. This means that for every g, $V_{M,g}(\exists y(Lxy \wedge \neg Ey)) = 1$, from which it is clear that $V_{M,g}(\forall x \exists y(Lxy \wedge \neg Ey)) = 1$.

Exercise 8

Work out exercise 7bi, iii, and v again, now according to approach B (definition 9).

3.6.4 *Universal Validity*

In predicate logic as in propositional logic, we speak of *contradictions*, these being sentences ϕ such that $V_M(\phi) = 0$ for all models **M** in the language from which ϕ is taken. Here are some examples of contradictions: $\forall x(Ax \wedge \neg Ax)$, $\forall xAx \wedge \exists y \neg Ay$, $\exists x \forall y(Ryx \leftrightarrow \neg Ryy)$ (the last one is a formalization of Russell's paradox).

Formulas ϕ such that $V_M(\phi) = 1$ for all models **M** for the language from which ϕ is taken are called *universally valid* formulas (they are not normally called tautologies). That ϕ is universally valid is written as $\models \phi$. Here are some examples of universally valid formulas (more will follow later): $\forall x(Ax \vee \neg Ax)$, $\forall x(Ax \wedge Bx) \rightarrow \forall xAx$, $(\forall x(Ax \vee Bx) \wedge \exists x \neg Ax) \rightarrow \exists xBx$.

And in predicate logic as in propositional logic, sentences ϕ and ψ are said to be *equivalent* if they always have the same truth values, that is, if for every model **M** for the language from which ϕ and ψ are taken, $V_M(\phi) = V_M(\psi)$. On approach B, this can be generalized to: two formulas ϕ and ψ are equivalent if for every model **M** for the language from which they are taken and every assignment g into **M**, $V_{M,g}(\phi) = V_{M,g}(\psi)$. As an example of a pair of equivalent sentences, we have $\forall xAx$, $\forall yAy$, as can easily be checked. More generally, are $\forall x\phi$ and $\forall y([y/x]\phi)$ always equivalent? Not when y occurs free in ϕ; obviously $\exists xLxy$ is not equivalent to $\exists yLyy$: somebody may love y without anybody loving him- or herself.

It might be thought though, that $\forall x\phi$ and $\forall y([y/x]\phi)$ are equivalent for any ϕ in which y does not occur free. This is, however, not the case, as can be seen from the fact that $\forall x \exists yAxy$ and $\forall y \exists yAyy$ are not equivalent. In $\forall y \exists yAyy$, the quantifier $\forall y$ does not bind any variable y, and therefore $\forall y \exists yAyy$ is equivalent to $\exists yAyy$. But clearly $\forall x \exists yAxy$ can be true without $\exists yAyy$ being true. Everyone has a mother, for example, but there is no one who is his or her own mother. The problem, of course, is that y has been substituted for a free variable x within the range of the quantifier $\forall y$. If we want to turn the above into a theorem, then we need at least one restriction saying that this may not occur. The following definition enables us to formulate such restrictions more easily:

Definition 10

y is free (for substitution) for x in ϕ if x does not occur as a free variable within the scope of any quantifier $\forall y$ or $\exists y$ in ϕ.

For example, y will clearly be free for x in ϕ if y doesn't appear in ϕ. In general, it is not difficult to prove (by induction on the complexity of ϕ) that for ϕ in which y does not occur free, ϕ and $\forall y([y/x]\phi)$ are indeed equivalent if y is free for x in ϕ.

In predicate logic as in propositional logic, substituting equivalent subformulas for each other does not affect equivalence. We will discuss this in §4.2, but we use it in the following list of pairs of equivalent formulas:

(a) $\forall x\neg\phi$ is equivalent to $\neg\exists x\phi$. This is apparent from the fact that $V_{M,g}(\forall x\neg\phi) = 1$ iff for every $d \in D_M$, $V_{M,g[x/d]}(\neg\phi) = 1$; iff for every $d \in D_M$, $V_{M,g[x/d]}(\phi) = 0$; iff it is not the case that there is a $d \in D_M$ such that $V_{M,g[x/d]}(\phi) = 1$; iff it is not the case that $V_{M,g}(\exists x\phi) = 1$; iff $V_{M,g}(\exists x\phi) = 0$; iff $V_{M,g}(\neg\exists x\phi) = 1$.

(b) $\forall x\phi$ is equivalent to $\neg\exists x\neg\phi$, since $\forall x\phi$ is equivalent to $\forall x\neg\neg\phi$, and thus, according to (a), to $\neg\exists x\neg\phi$ too.

(c) $\neg\forall x\phi$ is equivalent to $\exists x\neg\phi$, since $\exists x\neg\phi$ is equivalent to $\neg\neg\exists x\neg\phi$, and thus, according to (b), to $\neg\forall x\phi$ too.

(d) $\neg\forall x\neg\phi$ is equivalent to $\exists x\phi$. According to (c), $\neg\forall x\neg\phi$ is equivalent to $\exists x\neg\neg\phi$, and thus to $\exists x\phi$.

(e) $\forall x(Ax \wedge Bx)$ is equivalent to $\forall xAx \wedge \forall xBx$, since $V_{M,g}(\forall x(Ax \wedge Bx)) = 1$ iff for every $d \in D_M$: $V_{M,g[x/d]}(Ax \wedge Bx) = 1$; iff for every $d \in D_M$: $V_{M,g[x/d]}(Ax) = 1$ and $V_{M,g[x/d]}(Bx) = 1$; iff for every $d \in D_M$: $V_{M,g[x/d]}(Ax) = 1$, while for every $d \in D_M$: $V_{M,g[x/d]}(Bx) = 1$; iff $V_{M,g}(\forall xAx) = 1$ and $V_{M,g}(\forall xBx) = 1$; iff $V_{M,g}(\forall xAx \wedge \forall xBx)) = 1$.

(f) $\forall x(\phi \wedge \psi)$ is equivalent to $\forall x\phi \wedge \forall x\psi$. This is a generalization of (e), and its proof is the same.

(g) $\exists x(\phi \vee \psi)$ is equivalent to $\exists x\phi \vee \exists x\psi$, since $\exists x(\phi \vee \psi)$ is equivalent to $\neg\forall x\neg(\phi \vee \psi)$, and thus to $\neg\forall x(\neg\phi \wedge \neg\psi)$ (de Morgan) and thus, according to (f), to $\neg(\forall x\neg\phi \wedge \forall x\neg\psi)$, and thus to $\neg\forall x\neg\phi \vee \neg\forall x\neg\psi$ (de Morgan), and thus, according to (d), to $\exists x\phi \vee \exists x\psi$.

N.B. $\forall x(\phi \vee \psi)$ is not necessarily equivalent to $\forall x\phi \vee \forall x\psi$. For example, each is male or female in the domain of human beings, but it is not the case that either all are male or all are female. $\exists x(\phi \wedge \psi)$ and $\exists x\phi \wedge \exists x\psi$ are not necessarily equivalent either. What we do have, and can easily prove, is:

(h) $\forall x(\phi \vee \psi)$ is equivalent to $\phi \vee \forall x\psi$ if x is not free in ϕ, and to $\forall x\phi \vee \psi$ if x is not free in ψ. Similarly:

(k) $\exists x(\phi \wedge \psi)$ is equivalent to $\exists x\phi \wedge \psi$ if x is not free in ψ, and to $\phi \wedge \exists x\phi$ if x is not free in ϕ.

(l) $\forall x(\phi \rightarrow \psi)$ is equivalent to $\phi \rightarrow \forall x\psi$ if x is not free in ϕ, since $\forall x(\phi \rightarrow \psi)$ is equivalent to $\forall x(\neg\phi \vee \psi)$ and thus, according to (h), to $\neg\phi \vee \forall x\psi$, and thus to $\phi \rightarrow \forall x\psi$. An example: *For everyone it holds that if the weather is fine, then he or she is in a good mood* means the same as *If the weather is fine, then everyone is in a good mood.*

(m) $\forall x(\phi \rightarrow \psi)$ is equivalent to $\exists x\phi \rightarrow \psi$ if x is not free in ψ, since $\forall x(\phi \rightarrow \psi)$ is equivalent to $\forall x(\neg\phi \vee \psi)$ and thus, according to (h), to $\forall x\neg\phi \vee \psi$, and thus, according to (a), to $\neg\exists x\phi \vee \psi$, and thus to $\exists x\phi \rightarrow \psi$. An example: *For everyone it holds that if he or she puts a penny in the slot, then a package of chewing gum drops out* means the same as *If someone puts a penny in the machine, then a package of chewing gum rolls out.*

(n) $\exists x\exists y(Ax \wedge By)$ is equivalent to $\exists xAx \wedge \exists yBy$, since $\exists x\exists y(Ax \wedge By)$ is equivalent to $\exists x(Ax \wedge \exists yBy)$, given (k), and with another application of (k), to $\exists xAx \wedge \exists yBy$.

(o) $\exists x\phi$ is equivalent to $\exists y([y/x]\phi)$ if y does not occur free in ϕ and y is free for x in ϕ, since $\exists x\phi$ is equivalent to $\neg\forall x\neg\phi$, according to (d). This in turn is equivalent to $\neg\forall y([y/x]\neg\phi)$, for y is free for x in ϕ if y is free for x in $\neg\phi$. And $\neg\forall y(\exists y/x]\neg\phi)$, finally, is equivalent to $\exists y([y/x]\phi)$ by (d), since $\neg([y/x]\phi)$ and $[y/x]\neg\phi$ are one and the same formula.

(p) $\forall x\forall y\phi$ is equivalent to $\forall y\forall x\phi$, as can easily be proved.

(q) $\exists x\exists y\phi$ is equivalent to $\exists y\exists x\phi$, on the basis of (d) and (p).

(r) $\exists x\exists yAxy$ is equivalent to $\exists x\exists yAyx$. According to (o), $\exists x\exists yAxy$ is equivalent to $\exists x\exists zAxz$, with another application of (o), to $\exists w\exists zAwz$, with (q), to $\exists z\exists wAwz$, and applying (o) another two times, to $\exists x\exists yAyx$.

In predicate logic too, for sentences ϕ and ψ, $\vDash\phi \leftrightarrow \psi$ iff ϕ and ψ are equivalent. And if $\vDash\phi \leftrightarrow \psi$, then both $\vDash\phi \rightarrow \psi$ and $\vDash\psi \rightarrow \phi$. But it is quite possible that $\vDash\phi \rightarrow \psi$ without ϕ and ψ being fully equivalent.

Here are some examples of universally valid formulas (proofs are omitted):

(i)	$\forall x\phi \rightarrow \exists x\phi$	(vi)	$\exists x\forall y\phi \rightarrow \exists y\forall x\phi$
(ii)	$\forall x\phi \rightarrow [t/x]\phi$	(vii)	$\forall xAxx \rightarrow \forall x\exists yAxy$
(iii)	$[t/x]\phi \rightarrow \exists x\phi$	(viii)	$\exists x\forall yAxy \rightarrow \exists xAxx$
(iv)	$(\forall x\phi \wedge \forall x\psi) \rightarrow \forall x(\phi \wedge \psi)$	(ix)	$\forall x(\phi \rightarrow \psi) \rightarrow (\forall x\phi \rightarrow \forall x\psi)$
(v)	$\exists x(\phi \wedge \psi) \rightarrow (\exists x\phi \wedge \exists x\psi)$	(x)	$\forall x(\phi \rightarrow \psi) \rightarrow (\exists x\phi \rightarrow \exists x\psi)$

Exercise 9

Prove of (i), (ii), (v) and (vii) of the above formulas that they are universally valid: prove (i) and (v) using approach A, assuming that all elements of a model have a name; prove (ii) and (vii) using approach B.

Exercise 10 ◇

Find as many implications and nonimplications as you can in the set of all possible formulas of the form Rxy prefixed by two quantifiers Q_1x, Q_2y (not necessarily in that order).

3.6.5 *Rules*

In order to discover universally valid formulas we may use certain *rules*. First, there is *modus ponens:*

(i) If $\models\phi$ and $\models\phi \rightarrow \psi$, then $\models\psi$.

It is not difficult to see that this rule is correct. For suppose that $\models\phi$ and $\models\phi \rightarrow \psi$, but that $\not\models\psi$. It follows from $\not\models\psi$ that there is some model **M** with $V_M(\psi) = 0$, and it follows from $\models\phi$ that $V_M(\phi) = 1$, and thus that $V_M(\phi \rightarrow \psi) = 0$, which contradicts $\models\phi \rightarrow \psi$. Here are some more rules:

(ii) If $\models\phi$ and $\models\psi$, then $\models\phi \wedge \psi$.
(iii) If $\models\phi \wedge \psi$, then $\models\phi$.
(iv) If $\models\phi$, then $\models\phi \vee \psi$.
(v) If $\models\phi \rightarrow \psi$, then $\models\neg\psi \rightarrow \neg\phi$.
(vi) $\models\neg\neg\phi$ iff $\models\phi$.

Such rules can be reduced to modus ponens. Take (v), for example, and suppose $\models\phi \rightarrow \psi$. It is clear that $\models(\phi \rightarrow \psi) \rightarrow (\neg\psi \rightarrow \neg\phi)$, since this formula has the form of a propositional tautology (theorem 13 in §4.2.2 shows that substitutions into tautologies like this are universally valid). Then with modus ponens it follows that $\models\neg\psi \rightarrow \neg\phi$. Here is a different kind of rule:

(vii) $\models\phi$ iff $\models\forall x([x/c]\phi)$, if x is free for c in ϕ.

Intuitively this is clear enough: if ϕ is universally valid and c is a constant appearing in ϕ, then apparently the truth of ϕ is independent of the interpretation given to c (ϕ holds for an 'arbitrary' c), so we might as well have a universal quantification instead of c.

Proof of (vii):

$\Leftarrow$: Suppose $\models\forall x([x/c]\phi)$. From example (ii) at the end of §3.6.4, we may conclude that $\models\forall x([x/c]\phi) \rightarrow [c/x][x/c]\phi$, and $[c/x][x/c]\phi$ is the same formula as ϕ (since x is free for c in ϕ). Now $\models\phi$ follows with modus ponens.

$\Rightarrow$: Suppose $\models\phi$, while $\not\models\forall x([x/c]\phi)$. Then apparently there is a model **M** with $V_M([x/c]\phi) = 0$. This means that there is an assignment g into **M** such that $V_{M,g}([x/c]\phi) = 0$. If we now define **M′** such that **M′** is the same as **M** (the same domain, the same interpretations), except that $I_{M'}(c) = g(x)$, then it is clear that

$V_{M'}(\phi) = 0$, since x appears as a free variable in $[x/c]\phi$ at precisely the same points at which c appears in ϕ, because x is free for c in ϕ. This, however, cannot be the case, since ϕ is universally valid, so $\not\models \forall x([x/c]\phi)$ cannot be the case either. □

Rule (vii) now opens all kinds of possibilities. From $\models (Ac \wedge Bc) \rightarrow Ac$ (by substitution into a taulogy), it now follows that $\models \forall x((Ax \wedge Bx) \rightarrow Ax)$. And applying (ix) in §3.6.4 and modus ponens to this result, we obtain $\models \forall x(Ax \wedge Bx) \rightarrow \forall xAx$.