H. Burkhardt & B. Smith (Eds.), Handbook of Metaphysics and Ontology, Vol. 1: A-K. Munich: Philosophia (Analytica), 1991, pp. 204-208.

Definition
II: Rules of Definition

In the older Aristotelian logic a definition is the delimitation of a species by stating the genus which includes it and the specific difference or distinguishing characteristic of the species. A typical example is the definition of man as a rational animal. The genus is the animal genus and the distinguishing characteristic is rationality. The most important classical texts on definition are to be found in Aristotle's *Posterior Analytics* and *Topics*.

A traditional definition *per genus et differentiam* is often called a *real* definition because it is said to characterize the essence of a species. The kinds of definition common in mathematics, that is, definitions which introduce a new symbol, are often called *verbal* or *nominal* definitions. However, it is not clear how a sharp distinction between the two kinds of definition can be made. For our purposes, it is sufficient to understand that a definition is a statement which establishes the meaning of an expression. The definition accomplishes this by relating the expression it defines to other expressions already available.

At least two questions immediately arise from this vague statement about what definitions are. What is meant by 'other expressions already available'? What restrictions, if any, are there on the logical form of sentences which may serve as definitions? The answer to the first question is that we have in mind the introduction of a definition within a specified theory, like the elementary theory of arithmetic. As understood here, a *theory* is characterized in terms of its primitive, non-logical symbols and its axioms. The second question is answered by giving rules of definition, which are illustrated later.

Although the theory of definition to be outlined here is a part of modern logic and has only been developed in this century, the texts of Aristotle mentioned earlier contain much good advice about formulating definitions in more informal contexts which do not assume an explicit theory as background.

The first definition in a theory is, then, a sentence of a certain form which establishes the meaning of a new symbol of the theory in terms of the primitive symbols of the theory. The second definition in a theory is a sentence of a certain form which establishes the meaning of a second new symbol of the theory in terms of the primitive symbols and the first defined symbol of the theory; and similarly for subsequent definitions. The point to be noted is that the definitions in a theory are introduced one at a time in some fixed sequence. Because of this fixed sequence we may speak meaningfully of *preceding* definitions in the theory. Often it is convenient to adopt the viewpoint that any defined symbol must be defined in terms only of the primitive symbols of the theory. In this case there is no need to introduce definitions in some fixed sequence. However, the common mathematical and scientific practice is to use previously defined symbols in defining new symbols; and to give an exact account of this practice, a fixed sequence of definitions is needed.

From the standpoint of the logic of inference, a definition in a theory is simply regarded as a new axiom or premiss. But it is not intended that a definition shall strengthen the theory in any substantive way. The point of introducing a new symbol is to facilitate deductive investigation of the structure of the theory, but not to add to that structure. Two criteria which make more specific these intuitive ideas about the character of definitions are that:

1. a defined symbol should always be eliminable from any formula of the theory, and
2. a new definition does not permit the proof of relationships among the old symbols which were previously unprovable, that is, it does not function as a creative axiom.

(These two criteria of eliminability and non-creativity were first introduced explicitly by the Polish logician S. Leśniewski 1929.) For instance, in arithmetic we introduce the symbols for subtraction by the equivalence:

(1) $x - y = z$ if and only if $x = y + z$.

We may use (1) to eliminate any occurrence of the subtraction symbol. Thus by virtue of (1) we eliminate '$-$' from:

If $y \neq 0$ then $x - y \neq x$,

and obtain the arithmetically equivalent statement:

If $y \neq 0$ then $x \neq y + x$.

It seems reasonable to require that any definition introducing a new symbol may be used to eliminate all subsequent meaningful occurrences of the new symbols. To be eliminable is a characteristic property of a defined symbol, as opposed to a primitive symbol. The concept of eliminability is formalized as follows, where *iff* is an abbreviation for *if and only if*:

Criterion of Eliminability. A statement S introducing a new symbol of a theory satisfies the criterion of eliminability if and only if: whenever S_1 is a statement in which the new symbol occurs, then there is a statement S_2 in which the new symbol does not occur such that *If* S *then* (S_1 *iff* S_2) is derivable from the axioms and preceding definitions of the theory.

The notion of a definition not being creative is formalized in the following statement:

Criterion of Non-creativity. A statement S introducing a new symbol of a theory satisfies the criterion of non-creativity if and only if: there is no statement T in which the new symbol does not occur such that *If* S *then* T is derivable from the axioms and preceding definitions of the theory but T is not so derivable

In other words, we cannot permit a statement or formula S introducing a new symbol to make possible the derivation of some previously unprovable theorem stated wholly in terms of primitive and previously defined symbols. An example of a formula which does not satisfy this criterion of non-creativity is the second axiom for groups if we consider a more limited theory than that of groups. The single primitive symbol of our theory is the binary symbol 'o' and the single axiom the associative axiom:

(1) $xo(yoz) = (xoy)oz.$

As the first definition of this theory we now propose the following equation introducing the new individual constant e:

(2) $xoe = x.$

However, applying the criterion of non-creativity we reject (2) as a proposed definition in our theory, for from (2) we may derive at once:

(3) $(\exists y)(x)(xoy = x).$

We note that (3) is a formula whose only non-logical symbol is the primitive symbol of the theory, but it is trivial to find an interpretation showing that (3) cannot be derived from (1). Thus (2) is creative and must be rejected as a proper definition.

It should be noticed that a special consequence of the criterion of non-creativity is the criterion of relative consistency. If the axioms and preceding definitions are consistent and if a statement introducing a new symbol may be used to derive a contradiction, then the new statement does not satisfy the criterion of non-creativity.

In theories stated in precise language (whether the subject matter is pure mathematics or science), we ordinarily introduce three kinds of defined symbols: relation symbols, operation symbols, and individual constants. We consider here only the rules for introducing relation symbols. (A detailed but elementary treatment is to be found in Suppes 1957, Chapter 8.)

In dealing with definitions, which are ordinarily equivalences or identities, it is customary to introduce the new symbol on the left side and to call this side the *definiendum* (thing to be defined). The right side is called the *definiens* (thing defining).

Rule for Defining Relation Symbols. An equivalence D introducing a new n-place relation symbol R is a proper definition in a theory if and only if D is of the form $R(v_1, \ldots, v_n)$ iff S, and the following restrictions are satisfied: (i) $v_1, \ldots, v_n$ are distinct variables; (ii) S has no free variables other than $v_1, \ldots, v_n$; and (iii) S is a statement in which the only non-logical constants are primitive symbols and previously defined symbols of the theory.

The *definiendum* $R(v_1, \ldots, v_n)$ is an atomic formula, which form is needed to guarantee elimination of the defined relation symbol from every possible context. Some examples will help clarify the three restrictions on the rule. The requirement that the variables $v_1, \ldots, v_n$ be distinct prevents definitions like:

(4) $R(x,x)$ if and only if $x \div x \leqslant 1.$

Formula (1) does not really define the binary relation symbol R, since only one variable occurs in the *definiendum*. With (1) at hand, we would not know how to eliminate R from the statement 'if $R(x,y)$ then $x \leqslant y$'. The *definiens* of (1) must be regarded as defining a unary relation symbol, which is a trivial universal property possessed by every number $x \neq 0$.

The second restriction prevents definitions like:

(5) $R(x)$ if and only if $x + y = 0.$

When (5) is added to the axioms of arithmetic we may derive a contradiction. The source of the trouble is the appearance of the variable y in the *definiens* but not in the *definiendum*, for (5) is logically equivalent to the pair of statements:

(6) If $x + y = 0$ then $R(x)$,
(7) If $R(x)$ then $x + y = 0.$

But from the logic of quantifiers we know that (6) is equivalent to:

(8) If there is a y such that $x + y = 0$ then $R(x)$,

and (7) is equivalent to:

(9) If $R(x)$ then for every y, $x + y = 0.$

From (8) and (9) we immediately infer the false statement:

(10) If there is a y such that $x + y = 0$ then for every y, $x + y = 0$.

(Note that the variable x is left free in this discussion, since it appears in a proper manner in both the *definiendum* and *definiens* of (5).)

The third restriction simply prohibits two kinds of circularity of definition. We could not admit as a proper definition:

(11) $R(x)$ if and only if $R(x)$;

a logical truth such as (11) would not be creative. Its defect is that it does not satisfy the criterion of eliminability. Of a similar sort is the pair of equivalences:

(12) $R(x)$ if and only if it is not the case $P(x)$,
(13) $P(x)$ if and only if it is not the case $R(x)$.

If we define the relation symbol R in terms of the new relation symbol P, and vice versa, then we are not able to eliminate either in favour of the primitive notation.

Padoa's Principle and Independence of Primitive Symbols. When the primitive symbols of a theory are given, it is natural to ask if it would be possible to define one of them in terms of the others. The Italian logician Alessandro Padoa (1868–1937) formulated (1902, 1903) a principle applying the method of interpretation which may be used to show that the primitive symbols are independent, that is, that one may not be defined in terms of the other. The principle is simple. To prove that a given primitive symbol is independent of the remaining primitives, find two interpretations of the axioms of the theory such that the given primitive has two different interpretations and the remaining primitive symbols have the same interpretation. (Theoretical justification of this principle was first given in Tarski 1935.) For instance, consider the theory of preference based on the primitive relation symbols P (for strict preference) and I (for indifference). The axioms of the theory are:

A1. *If* xPy and yPz, *then* xPz.
A2. *If* xIy and yIz, *then* xIz.
A3 *Exactly one of the following*: xPy, yPx, xIy.

We want to show that P is independent of I, that is, cannot be defined in terms of I. Let the domain of interpretation for both interpretations be the set $\{1,2\}$. Let I be interpreted as identity in both cases. In one case let P be interpreted as $<$ and in the other case as $>$. In the first interpretation we have: $1 \; P \; 2$ since $1 < 2$, and consequently by Axiom A3 not $2 \; P \; 1$.

But in the second interpretation, we have: $2 \; P \; 1$ since $2 > 1$.

Now if P were definable in terms of I then P would have to be the same in both interpretations, since I is. However, P is not the same, and we conclude that P cannot be defined in terms of I. E. W. Beth (1953) proved the converse of Padoa's principle for theories formulated in first-order logic, namely, if two models of the sort described above for proving a concept independent do not exist, then the concept is explicitly definable.

Conditional Definitions. In many situations the rules of definition, illustrated above for the case of relation symbols, are too severe. The reason is that many significant defined concepts have an intended natural meaning only when some hypothesis is satisfied. A familiar example is the definition of division in arithmetic.

If $y \neq 0$, *then* $x/y = z$ *if and only if* $x = y \cdot z$.

The main disadvantage of such conditional definitions is that they do not fully satisfy the criterion of eliminability, for instance, in the case of division when $y = 0$. But, with the obvious modifications in the rules of definition, such definitions do satisfy eliminability whenever the hypothesis of the conditional definition is satisfied.

A historically important application in philosophy of such conditional definitions was given in Rudolf Carnap's concept of a *reduction sentence* (1936). Such sentences were introduced to provide a method of relating dispositional predicates to directly observable predicates. For example, *if x is placed in water, then x is soluble in water if and only if x dissolves.*

In general, the kinds of questions about axioms of a theory that are standard, e.g., whether the axioms are mutually independent, can be reflected in corresponding questions about the primitive concepts of a theory. To give just one more example, the axioms of a theory are *categorical* if any two models of the theory are isomorphic. Tarski (1935) defined the primitive concepts of a theory as *complete* if there is not another set of axioms that:

1. use additional primitive concepts,
2. are categorical, and
3. characterize an essentially richer theory.

Both Euclidean geometry and the elementary algebra of the real numbers are categorical theories, but only the second is complete in its primitive concepts.

FURTHER READING

Beth, E W., 1953, "On Padoa's method in the theory of definition", *Indagationes Mathematicae*, 15, 330–9
Carnap, R., 1936–7, "Testability and meaning", *Philosophy of Science*, 3, 419–71; 4, 1–40.
Leśniewski, S., 1929, "Grundzuge eines neuen Systems der Grundlagen der Mathematik, §§1–11", *Fundamenta Mathematicae*, 14, 1–81
Padoa, A , 1902, "Un nouveau système irréductible de postulats pour l'algèbre", *Comptes rendus du deuxième congrès international des mathématiciens (Paris, 1900)*, Paris, Gauthier-Villars, 249–56.
— 1903, "Le problème no. 2 de M. David Hilbert", *L'Enseignement Mathématique*, 5, 85–91
Suppes, P , 1957, *Introduction to Logic*, New York: Van Nostrand Reinhold Company
Tarski, A , 1935, "Einige methodologische Untersuchungen uber die Definierbarkeit der Begriffe", *Erkenntnis*, 5, 80–100 (Eng trans "Some methodological investigations on the definability of concepts", in A. Tarski, *Logic, Semantics, Metamathematics*, 2nd ed , revised, Indianapolis, Ind · Hackett, 1983, 296–319)

PATRICK SUPPES