TARSKI'S LOGIC

Keith Simmons

1 INTRODUCTION

Alfred Tarski was born in Warsaw on January 14th 1902, under the family name of Tajtelbaum.[1] He studied at the University of Warsaw from 1918 to 1924, when he received his PhD in mathematics. His doctoral dissertation *O wyrazie pierwotnym logistyki* (*On the primitive term of logistic*) was written under the supervision of S. Lesniewski.[2] Appointed Docent in 1926 and later Adjunct Professor, Tarski taught at the University of Warsaw until 1939. During that year, Tarski visited the US for a lecture tour, but was prevented from returning to Poland by the outbreak of war. Between 1939 and 1942 Tarski was at Harvard University, the City College of New York, and the Institute for Advanced Study in Princeton. In 1942, he was appointed lecturer in mathematics at the University of California at Berkeley, where he remained for the rest of his career. He became professor of mathematics there in 1946, when his family joined him from Poland. In 1958, Tarski founded the Group in Logic and the Methodology of Science. He retired in 1968, though he taught for several more years, and continued his research throughout his retirement. Tarski died in Berkeley on October 27th 1983.

Tarski's work has had an enormous influence on the development of logic and mathematics over the last eighty years. He broke new ground with his work on metamathematics and semantics. His methods and results in those fields and many others — including algebra, geometry, and set theory — have become part of the fabric of modern logic and mathematics. The very divisions between these various fields appear somewhat artificial in the light of Tarski's work.

I have attempted to make this chapter accessible to the non-mathematician who is familiar with basic logic. I have focussed more on Tarski's work in logic, semantics and metamathematics, and less on Tarski's more purely mathematical work. Nevertheless, I hope that the reader will come away with a sense of Tarski's achievements in each of the many areas to which he contributed. My aim has been to make this chapter as self-contained as possible. For the reader who wants to follow out the formal details of a definition or a proof, I have endeavored to

[1] The family name Tarski was adopted in 1924. There is disagreement in the literature about Tarski's birthdate. I have followed Mostowski [Tarski, 1967] and the *Encyclopedia Britannica*, but other sources given the year of his birth as 1901 (see, for example, [Hodges, 1986a]).

[2] The dissertation was submitted in 1923, and the essential part of it was published as the paper [Tarski, 1923]. An English translation of the paper appears in [Tarski, 1983a].

Handbook of the History of Logic. Volume 5
Dov M. Gabbay and John Woods (Editors)

provide explanations in footnotes. And where the formal material goes beyond the scope of this chapter, I have provided further references. I quote often from Tarski, since it is very hard to improve upon Tarski's own words — no-one who is familiar with Tarski's work can fail to be struck by the clarity of his writing. Works by Tarski and others are referred to by the last two digits of the year of publication.[3] Throughout I have tried to provide some historical context for Tarski's work and its significance.

2 METAMATHEMATICS

We turn first to what Tarski initially called the *methodology of the deductive sciences* — though he later preferred the label "metalogic and metamathematics",[4] and often just used the term "metamathematics". Its objects of study are theories of a certain kind — *formalized deductive theories*. They constitute the subject matter of metamathematics just as

> "spatial entities constitute the subject matter of geometry and animals that of zoology" [Tarski, 1930d, p. 60].[5]

One task of metamathematics is to construct its objects of study. The construction of formalized deductive theories proceeds in accordance with the *deductive method*. When we study or advance any science, Tarski says,

> "a method would be ideal, if it permitted us to explain the meaning of every expression occurring in this science and to justify each of its assertions" [Tarski, 1941a, p. 117]

Tarski points out that this ideal can never be realized — since we must use expressions to explain the meaning of an expression, we have to take some terms as primitive, on pain of an infinite regress.

> "When we set out to construct a given discipline, we distinguish, first of all, a certain small group of expressions of this discipline that seems to us to be immediately understandable; the expressions of this group we call *primitive terms* or *undefined terms*." (p. 118)

All other expressions are defined via the primitive terms. And we proceed similarly with the asserted statements of the theory:

---

[3] To reduce the number of footnotes, I refer to Tarski's work parenthetically in the main text. When the work of Tarski's from which I am quoting is clear from the context, I provide only a page number. For a full bibliography of Tarski, including all his papers, abstracts, monographs, exercises and problems, contributions to discussions, reviews, publications as editor, project reports and letters, see Givant [1986].

[4] See [Tarski, 1941a] *Introduction to Logic*, p. 140.

[5] Compare this remark in [Tarski, 1941a]: "The methodology of the deductive sciences became a general science of deductive sciences in an analogous sense as arithmetic is the science of numbers and geometry is the science of geometrical configurations." (p. 138)

> "Some of these statements which to us have the appearance of evidence are chosen as the so-called *primitive statements* or *axioms*..." (p. 118)

And further:

> "we agree to accept any other statement as true only if we have succeeded in establishing its validity, and to use, while doing so, nothing but axioms, definitions and such statements of the discipline the validity of which has been established previously." (p. 118)

The method of constructing a theory in accordance with these principles is the *deductive method*,[6] and the theory so constructed is a *formalized deductive theory*.[7]

So one task of metamathematics is the construction of formalized deductive theories via the deductive method. Now, as a result of the application of this method "deductive theories acquire certain interesting and important features" [Tarski, 1941a, p. 120]. It is a further task of metamathematics to uncover these features.

This further task can be carried out at two distinct levels. One can investigate specific deductive theories. In the 1920s, Tarski carried out metamathematical investigations into a number of deductive theories - for example, the sentential calculus (see [Tarski, 1930c]), the algebra of logic, the arithmetic of real numbers, the geometry of straight lines, the theory of order, and the theory of groups.[8] Tarski writes in [Tarski, 1930d]:

> "Strictly speaking metamathematics is not to be regarded as a single theory. For the purpose of investigating each deductive discipline a special metadiscipline should be constructed." (p. 60)

---

[6] Tarski points out that logic occupies a special position in regard to the deductive method. In using the method to construct any other discipline or theory, we cannot but presuppose logic. In Tarski's terminology, given a target discipline, logic will be a discipline *preceding the given discipline*. And it may be convenient to presuppose other disciplines too. Tarski writes:

> "Thus logic itself does not presuppose any preceding discipline; in the construction of arithmetic as a special mathematical discipline logic is presupposed as the only preceding discipline; on the other hand, in the case of geometry it is expedient — though not unavoidable — to presuppose not only logic but also arithmetic." [Tarski, 1941a, p. 119]

So, whenever we apply the deductive method, we must be sure to enumerate the preceding disciplines.

[7] The phrase 'formalized deductive discipline' appears in the early papers [Tarski, 1930b] and [Tarski, 1930d]. In [Tarski, 1930d], Tarski writes:

> "Naturally not all deductive systems are presented in a form suitable for objects of scientific investigation. Those, for example, are not suitable which do not rest on a definite logical basis, have no precise rules of inference, and the theorems of which are formulated in the usually ambiguous and inexact terms of colloquial language — in a word those which are not formalized. (p.60)

[8] See [Tarski, 1983a, p. 205, fn. 2], where Tarski mentions that he investigated these last five theories during 1926–8.

But one can also proceed at a more abstract level, and explore concepts that are common to *all* these special metadisciplines. In [30d], Tarski proceeds at this higher level of generality, and we will start with Tarski's study of general deductive systems.

2.1 General deductive systems

In his paper "Fundamental Concepts of the Methodology of the Deductive Sciences" [Tarski, 1930d], Tarski's aim is

> "*to make precise the meaning of a series of important metamathematical concepts* which are common to the special metadisciplines, *and to establish the fundamental properties of these concepts.*" (p. 60, original emphases)

Tarski starts out with the fundamental concept of *consequence*, and provides the first rigorous axiomatic characterization of this notion. Tarski goes on to define a number of further concepts, including *logical equivalence*, *axiomatizability*, *independence*, *consistency*, and *completeness*. Tarski thus provided the first precise, systematic treatment of these basic metamathematical concepts.

The axiomatic system (or theory) that Tarski presents in [Tarski, 1930d] contains just two primitive concepts, *sentence* and *consequence*. Sentences are certain inscriptions of a well-defined form. And a deductive discipline is a set of (meaningful) sentences. Let S be the set of all meaningful sentences of a given language, and let A be an arbitrary set of sentences from S that compose a particular deductive discipline. '$Cn(A)$' denotes the set of consequences of set A, those sentences derived from A via rules of inference. The *schema* of a definition of consequence can be given as follows: the set of all consequences of the set A is the intersection of all sets which contain the set A and are closed under the given rules of inference. But we are working at too high a level of generality to give an exact definition of consequence; it is the task of a specific metadiscipline to establish rules of inference, and thereby an exact definition of consequence. Instead, we are to regard the notion of consequence as primitive, and characterized only through the axioms. From just these two primitive concepts, sentence and consequence,

> "almost all basic concepts of metamathematics can be defined; on the basis of the given axiom system various fundamental properties of these concepts can be established." (p. 69)

Tarski presents an axiom system composed of four axioms which express basic properties of the primitive concepts and "are satisfied in all known formalized disciplines" (p. 63). The axioms are as follows

Axiom 1. $\bar{\bar{S}} \leq \aleph_0$ (where "$\overline{\overline{S}}$" is the cardinality of the set S)

Axiom 2. If $A \subseteq S$, then $A \subseteq Cn(A)S$.

Axiom 3. If $A \subseteq S$, then $Cn(Cn(A)) = Cn(A)$.

Axiom 4. If $A \subseteq S$, then $Cn(A) = \cap\{Cn(X) : X$ is a finite subset of $A\}$.

Axiom 1, Tarski says, "scarcely requires comment" (p. 63), though he does acknowledge the challenge that S cannot be denumerably infinite if sentences are regarded as physical inscriptions.[9]

Axiom 2 tells us that consequences of A are sentences, and Axiom 3 tells us that consequences of consequences of A are consequences of A. According to Axiom 4, any consequence of A is a consequence of a finite subset of A — and this respects the idea that "in concrete disciplines" (p. 64) rules of inference operate on a finite number of sentences.

Tarski goes on to single out "an especially important category of sets of sentences ... namely, the *deductive systems*." A deductive system, or "*closed system*" or "simply a '*system*'" (p. 70), is a set of sentences that is closed under consequence — that is, it contains all of its consequences. Tarski writes:

> "Deductive systems are, so to speak, organic units which form the subject matter of metamathematical investigations. Various important notions, like consistency, completeness, and axiomatizability, which we shall encounter in the sequel, are theoretically applicable to any sets of sentences, but in practice are applied chiefly to systems." (p. 70)

The bulk of [Tarski, 1930d] is a detailed examination of the properties of (deductive) systems. Tarski defines a number of fundamental metamathematical concepts. Two sets of sentences are *logically equivalent* if they have all their consequences in common (p. 72). An *axiom system* of a set of sentences is a finite set which is equivalent to that set, and a set of sentences which has at least one axiom system is called *axiomatizable* (p. 72). A set of sentences is *independent* if is is not equivalent to any of its proper subsets (p. 83). A *basis of the set A* is an independent set of sentences which is equivalent to the set A (p. 88). A set of sentences is *consistent* if it is not equivalent to the set of all meaningful sentences (p. 90).[10] The *decision domain* of a set A of sentences is the set of all sentences that are either consequences of A or which, when added to A, yield an inconsistent

[9] Of the assumption that S is denumerably infinite, Tarski writes:

> "... it may be noted here that I personally regard such an assumption as quite sensible, and that it appears to me even to be useful from a metamathematical standpoint to replace the inequality sign by the equality sign in Ax. 1." [Tarski, 1930d, p. 64]

Throughout his career, Tarski never showed any finitistic qualms.

[10] Tarski points out that his definition of consistency departs from the usual one, according to which a set of sentences is consistent if there is no sentence which together with its negation belongs to the consequences of this set. Tarski notes that the two definitions are equivalent for all disciplines based on the system of sentential calculus — indeed, this is the content of Theorem 9* of Tarski's paper [Tarski, 1930b]. Tarski also notes that his definition is much more general, since it applies to theories without negation, and to theories in which negation does not have the usual properties. Tarski cites Post [Tarski, 1921] as the source of this definition.

set of sentences; and A is said to be *complete* if its decision domain contains all meaningful sentences (p. 93).[11] The *cardinal degree of completeness* of the set A of sentences is the cardinality of the total number of systems which include the set A (pp. 100–101). And the *ordinal degree of completeness* is the smallest ordinal number π for which there is no strictly increasing sequence of type π of consistent systems which include the set A (p. 101).[12]

[Tarski, 1930d] contains proofs of a large number of seminal theorems on axiomatizability, independence, and consistency and completeness. For example, Corollary 37 (p. 87) tells us that if a system includes an infinite independent set of sentences, then it has uncountably many subsystems, of which only countably many are finitely axiomatizable. Tarski writes:

> "It is to be noted that within almost all deductive disciplines, and in particular within the simplest of them — the sentential calculus — it has been found possible to construct a set of sentences which is both infinite and independent, and thus to realize the hypothesis of the last corollary [Corollary 37]. Hence it turns out that in all these disciplines there are more unaxiomatizable than axiomatizable systems; the deductive systems are, so to speak, as a rule unaxiomatizable, although in practice we deal almost exclusively with axiomatizable systems. This paradoxical circumstance was first noticed by Lindenbaum in application to the sentential calculus." (p. 88)

(Tarski's use of the term 'paradoxical' indicates that unaxiomatizable theories came as a surprise.) For another example, Tarski establishes (by Corollary 44) that "every axiomatizable set of sentences possesses at least one basis" (p. 90). This result does not extend to unaxiomatizable sets of sentences on the basis of Axioms 1–4 (though as we will see below, it does extend to deductive disciplines that presuppose the sentential calculus). Another result, Theorem 56 (p. 98), tells us that "every consistent set of sentences can be extended to a consistent and complete system" (*ibid.*), a generalization of Lindenbaum's lemma for the sentential calculus.

2.2 Classical deductive systems

While in [Tarski, 1930d] Tarski investigated the broadest category of deductive systems, in his paper "On some fundamental concepts of metamathematics" [Tarski, 1930b], Tarski investigated a certain subcategory — the classical systems. In

[11] As with the definition of consistency, this definition of completeness departs from the usual one, and does not rest on the concept of negation. Tarski again cites Post [1921] as the source of this definition of completeness. Theorem 10* of Tarski's [1930b] states the equivalence of the two definitions for systems based on the sentential calculus.

[12] That is, there is no increasing sequence of type π of the form $\langle X, \ldots, X_\nu, \ldots, X_\eta, \ldots, \rangle$, where $X_\nu \subseteq X_\eta \subseteq S$ and and $Cn(X_\nu)?Cn(X_\eta)$ for $\nu < \eta < \pi$. Or, in the now-standard terminology, there exists no chain of consistent theories of type π beginning with X.

[Tarski, 1930b], Tarski lays out two groups of axioms. The first contains Axioms 1–4, together with a fifth:

Axiom 5. There exists a sentence $x \in S$ such that $Cn(\{x\}) = S$.

The axioms in the second group are "of a more special nature":

> "In contrast to the first group of axioms those of the second group apply, not to all deductive disciplines, but only to those which presuppose the sentential calculus, in the sense that in considerations relating to these disciplines we may use as premises all true sentences of the sentential calculus." (p. 31)

These additional axioms contain two new primitive concepts: the conditional and negation. *Axiom 6* tells us that if x and y are meaningful sentences, so are $x \to y$ and $\neg x$. The remaining four axioms are as follows (where '∅' below denotes the empty set):

Axiom 7. If $X \subseteq S, y \in S, z \in S$ and $y \to z \in Cn(X)$, then $z \in Cn(X \cup \{y\})$.

Axiom 8. If $X \subseteq S, y \in S, z \in S$ and $z \in Cn(X \cup \{y\})$, then $y \to z \in Cn(X)$.

Axiom 9. If $x \in S$, then $Cn(\{x, \neg x\}) = S$.

Axiom 10. If $x \in S$, then $Cn(\{x\}) \cap Cn(\{\neg x\}) = Cn(\varnothing)$.

Axiom 7 is the rule of detachment, or modus ponens. Tarski takes Axiom 8 to be one formulation of the so-called deduction theorem, a discovery of Tarski's. Tarski writes:

> "This theorem, in its application to the formalism of *Principia Mathematica*, was first established by the author as far back as 1921 .... Subsequently the deduction theorem was often applied in metamathematical discussion."[13]

Axiom 9 is the classical law of non-contradiction, in the form "Everything follows from a contradiction". Axiom 10 is a formulation of the classical law of excluded middle.

So the axiomatic system of [Tarski, 1930b] applies to a *subcategory* of deductive systems, the *classical* deductive systems, encompassing the sentential calculus and all classical systems that extend it, including the first-order predicate calculus.

[13] Tarski [1983a, p. 32, fn. †]]. Tarski reports that he established the theorem in connection with a discussion in the monograph of Ajdukiewicz [Tarski, 1921], and discussed the result in a lecture to the Warsaw Philosophical Institute, listed by title in *Ruch Filozoficzny*, vol. 6 (1921–2). As to applications of the deduction theorem, Tarski mentions that it was essential to the proofs of theorems in Lindenbaum and Tarski's note [Tarski, 1927b] and Tarski's note [Tarski, 1929b]. In 1933, Tarski outlined a proof of the deduction theorem for a particular formalized theory (see [Tarski, 1933b, Theorem 2(a), p. 286]). The deduction theorem is often attributed to Herbrand, who published the result in [Herbrand, 1928].

Tarski defines for this subcategory nearly all of the metamathematical notions that he defined in [Tarski, 1930d]] for deductive systems generally, and in much the same way.[14] He also states without proof a dozen or so theorems. Among these are the first published statement of Lindenbaum's lemma (Theorem 12, p. 34), and the result that every (countable) set of sentences possesses a base (Theorem 17, p.35) — a result that holds for the present subcategory but not generally, as we noted above.

2.3 The sentential calculus

Another subcategory of deductive systems — the sentential calculus — is investigated in Lukasiewicz and Tarski's "Investigations into the sentential calculus" [Tarski, 1930c]. The authors remark that

> "as the simplest deductive discipline, the sentential calculus is particularly suitable for metamathematical investigations. It can be regarded as a laboratory in which metamathematical methods can be discovered and mathematical concepts constructed which can then be carried over to more complicated mathematical systems." (p. 59)

[Tarski, 1930c] is "a compilation of theorems and concepts belonging to five different persons" (p. 38, fn ‡), namely, Lukasiewicz, Tarski, Lindenbaum, Sobocinski, and Wajsberg — all members of the Warsaw School of Logic.

At the outset of [Tarski, 1930c], Lukasiewicz and Tarski refer the reader to the conceptual apparatus and notation developed in [Tarski, 1930b]. The concepts of *sentential variable*, *conditional* and *negation* are taken as primitive. The metamathematical expressions '$c(x, y)$' and '$n(x)$' denote respectively the conditional with antecedent x and consequent y, and the negation of x. The set S of all sentences is defined as follows:

> "The *set S of all sentences* is the intersection of all those sets which contain all sentential variables (elementary sentences) and are closed under the operations of forming implications [conditionals] and negations." (p. 39)

Lukasiewicz and Tarski go on to define the notion of consequence in terms of *substitution* and *detachment* (or modus ponens).

> "The *set of consequences $Cn(X)$ of the set X of sentences* is the intersection of all those sets which include the set $X \subseteq S$ and are closed under the operations of substitution and detachment." (p. 40)

From this definition it follows that S and $Cn(X)$ satisfy the axioms 1–5 of [Tarski, 1930b] (see above). Now consider X such that $X \subseteq S$ and X is closed under consequence (that is, $Cn(X) = X$). Then by Tarski's characterization of a deductive

[14] The only exceptions are the notions of *decision domain* and *cardinal degree of completeness*.

system, X is a deductive system. It is these deductive systems that Lukasiewicz and Tarski investigate in [Tarski, 1930c].

Lukasiewicz and Tarski identify two methods of constructing such a deductive system. One is the familiar *axiomatic method*, according to which

> "an arbitrary, usually finite, set X of sentences — an *axiom system* — is given, and the set $Cn(X)$, i.e. the smallest deductive system over X, is formed." (p. 40)

The second method is the *matrix method*. The following definition of a matrix is due to Tarski:

> "A *(logical) matrix* is an ordered quadruple $M = [A, B, f, g]$ which consists of two disjoint sets (with elements of any kind whatever) A and B, a function f of two variables and a function g of one variable, where the two functions are defined for all elements of the set $A \cup B$ and take as values elements of $A \cup B$ exclusively." (p. 41)[15]

The elements of the set B are called the designated elements (following Bernays). For an example of a matrix, set A, B, f, g as follows: $A = \{0\}, B = \{1\}, f(0,0) = f(0,1) = f(1,1) = 1, f(1,0) = 0, g(0) = 1, g(1) = 0$. This is the matrix associated with the classical sentential calculus (and the designated value is 1). The idea is that the functions f and g correspond to the syntactic operations of forming conditionals and negations.[16] This correspondence is made precise by Tarski's recursive definition of a value function:

> "The function h is called a *value function* of the matrix $M = [A, B, f, g]$ if it satisfies the following conditions:
>
> 1. the function h is defined for every $x \in S$;
> 2. if x is a sentential variable, then $h(x) \in A \cup B$;
> 3. if $x \in S$ and $y \in S$, then $h(c(x, y) = f(h(x), h(y))$;
> 4. if $x \in S$ then $h(n(x) = g(h(x))$." (p. 41)

Given the notion of a value function, Tarski defines the notion of satisfaction of a sentence by a matrix:

[15] I have replaced Tarski's symbol '+' for set-theoretical union by the more familiar '∪'.

[16] The functions f and g may be captured by the these tables for '*to*' and '¬':

x	y	$x \to y$
1	1	1
1	0	0
0	1	1
0	0	1

x	$\neg x$
1	0
0	1

Replacing '1' and '0' by 'T' and 'F' yields the familiar truth tables (where T is the designated value).

"The sentence x is *satisfied (or verified) by the matrix* $M = [A, B, f, g]$ ... if the formula $h(x) \in B$ holds for every value function h of this matrix." (p. 41)

In order to construct a deductive system of the sentential calculus we proceed as follows: we set up a matrix and consider all those sentences satisfied by it.[17]

In [Tarski, 1930c], the matrix method is used to construct not only the classical (two-valued) system of the sentential calculus, but also an entire class of many-valued systems. The definition of the classical system L of the sentential calculus has already been anticipated:

"The *ordinary [classical] system* L *of the sentential calculus* is the set of all sentences which are satisfied by the matrix $M = [A, B, f, g]$ where $A = \{0\}, B = \{1\}$ and the functions f and g are defined by the formulas: $f(0,0) = f(0,1) = f(1,1) = 1, f(1,0) = 0, g(0) = 1, g(1) = 0$." (p. 42)

Lukasiewicz and Tarski observe that the consistency and completeness of L follows easily from this definition. They also define L by the axiomatic method, presenting an axiom system due to Lukasiewicz, namely the set $\{(p \rightarrow q) \rightarrow ((q \rightarrow r) \rightarrow (p \rightarrow r)), (\neg p \rightarrow p) \rightarrow p, p \rightarrow (\neg p \rightarrow q)\}$.

Tarski proved the following noteworthy theorem about L:

The system L, as well as every axiomatizable system of the sentential calculus which contains the sentences '$p \rightarrow (q \rightarrow p)$' and '$p \rightarrow (q \rightarrow ((p \rightarrow (q \rightarrow r)) \rightarrow r))$', possesses a basis consisting of a single sentence. (Theorem 8, p. 44.)

A generalization of this theorem is also stated:

The system L, as well as every axiomatizable system of the sentential calculus which contains the sentences '$p \rightarrow (q \rightarrow p)$' and '$p \rightarrow (q \rightarrow ((p \rightarrow (q \rightarrow r)) \rightarrow r))$', possesses for every natural number m a basis containing exactly m elements. (Theorem 10, p. 45.)[18]

---

[17]This procedure rests on the following theorem:
If M is a normal matrix, then the set of all sentences satisfied by M is a deductive system, where a matrix is normal if the formulas $x \in B$ and $y \in A$ always imply $f(x, y) \in A$.

The generality of the matrix method is established by a converse of this theorem, due to Lindenbaum. Lindenbaum's theorem runs as follows:

For every deductive system X there exists a normal matrix $M = [A, B, f, g]$, with an at most denumerable set $A \cup B$, which satisfies exactly the sentences of X. (See [Tarski, 1930c, pp. 41–42].)

[18]The proof of the theorem for L is credited to Sobocinski, and the generalizations to other systems to Tarski.

Tarski also proved the following theorem, which provides a contrast between L and other systems of the sentential calculus: "Theorem 11. For every natural number, systems of the sentential calculus exist every basis of which contains exactly m elements." [Tarski, 1930c, p. 45]

Lukasiewicz and Tarski also use the matrix method to define a class of many-valued systems founded by Lukasiewicz — the so-called "n-valued systems":

> "The *n-valued system L_n of the sentential calculus* (where n is a natural number or $n = \aleph_0$) is the set of all sentences which are satisfied by the matrix $M = [A, B, f, g]$ where, in the case $n = 1$ the set A is null, in the case $1 < n < \aleph_0 A$ consists of all fractions of the form $k/n - 1$ for $0 \leq k < n - 1$, and in the case $n = \aleph_0$it consists of all fractions $k/1$ for $0 \leq k < 1$; further the set B is equal to $\{1\}$ and the functions f and g are defined by the formulas: $f(x, y) = \min(1, 1 - x + y), g(x) = 1 - x$." (pp. 47–8)[19]

Lukasiewicz and Tarski report a number of significant results about these systems. They include the following:

(i) For every $n, 1 \leq n < \aleph_0, L_n$ is axiomatizable. (Theorem 22, p. 49).[20]

(ii) Let $M = [A, B, f, g]$ be a normal matrix [see fn. 18] in which the set $A \cup B$ is finite. If the sentences '$(p \to q) \to ((q \to r) \to (p \to r))$', $(q \to r) \to ((p \to q) \to (p \to r)), (p \to q) \to (\neg q \to \neg p), \neg q \to ((p \to q) \to \neg p)$ are satisfied by M, then the set of all sentences satisfied by M is axiomatizable. (Theorem 24, p. 50)

(iii) Every system L_n, where $2 \leq n < \aleph_0$, possesses, for every natural number m (and in particular for $m = 1$), a basis which has exactly m elements. (Theorem 26, p. 50.)

In comparing the axiomatic method and the matrix method, Tarski writes:

---

[19]For example, consider the case $n = 3$. Then $A = \{0, \frac{1}{2}\}$ and $B = \{1\}$. The functions f and g may be captured by the following tables for $\to$ and $\neg$ respectively:

p	q	$p \to q$
1	1	1
1	$\frac{1}{2}$	$\frac{1}{2}$
1	0	0
$\frac{1}{2}$	1	1
$\frac{1}{2}$	$\frac{1}{2}$	1
$\frac{1}{2}$	0	$\frac{1}{2}$
0	1	1
0	$\frac{1}{2}$	1
0	0	1

p	$\neg p$
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

Given values m for p and n for q, the value of $p \to q$ is $\min(1, 1 - m + n)$, and the value of $\neg p$ is $1 - m$.

In the given definition, A is the set of all proper fracctions. Lindenbaum showd that any other denumerably infinite subset of the interval $\langle 0, 1\rangle$ will serve as well.

[20]This theorem was proved by Wajsberg for $n = 3$ and for all n for which n-1 is a prime number, and later extended to all natural numbers by Lindenbaum.

> "Each of the two methods has its advantages and disadvantages. Systems constructed by means of the axiomatic method are easier to investigate regarding their axiomatizability, but systems generated by matrices are easier to test for completeness and consistency." (p. 42)

Tarski was keenly aware of the value of the matrix method as an alternative to the axiomatic method. Blok and Pigozzi point out that although Tarski was not the first to use matrices to construct deductive systems (or 'logics'), he was responsible for recognizing the importance of matrices for general metamathematics. They continue:

> "It is difficult to overestimate their importance in this regard. They provided a means of defining logics that were not *a priori* finitely axiomatizable, and this led naturally to the question of the existence of logics that are not finitely axiomatizable. They were used first by Bernays and then extensively by Lukasiewicz to establish the independence of various axiom systems. Finally, they provided a natural way of defining the equivalence of deductive systems with different primitive connectives. All three of these topics proved to be a rich source of problems." [Blok and Pigozzi, 1988, p. 421])

In addition, there is fruitful interplay between the two methods. Corcoran observes that [Tarski, 1930c]

> "contains results achieved by interrelating axiomatic and matrix methods for defining sentential calculi, an interrelation that clearly foreshadows what will later be seen as the interrelation of proof-theoretic (syntactic) and model-theoretic (semantic) methods." [Tarski, 1983a, p. xix]

2.4 The algebraization of logic

Tarski's paper "Foundations of the Calculus of Systems" [1935+1936] extends the metamathematical investigations of [1930b; 1930c] and [1930d]. Like [Tarski, 1930b], the paper considers only those theories "whose construction presupposes a logical basis of a greater or lesser extent, and at least the whole sentential calculus." (p. 342). But the axiom system of [Tarski, 1930b] is replaced by a "simpler", "more natural" one (p. 342). Let 'L' denote the smallest deductive system defined by Axioms 1–10 of [Tarski, 1930b] — that is, the system $C(\varnothing)$, the set of consequences of the empty set. In the axiom system of [Tarski, 1935+1936], the symbol 'L' is taken as a primitive expression instead of '$Cn(X)$' — the concept of *consequence* is no longer a primitive notion.

The new axiom system consists of the following five axioms:

Axiom 1. $0 < \overline{\overline{S}} \leq \aleph_0$ (where S is the set of meaningful sentences)

Axiom 2. If $x, y \in S$, then $\bar{x} \in S$ and $x \to y \in S$ (where '$\bar{x}$' is the negation of x).

Axiom 3. $L \subseteq S$.

Axiom 4. If $x, y, z \in S$, then $(\bar{x} \to x) \to x \in L, x \to (\bar{x} \to y) \in L$ and $(x \to y) \to ((y \to z) \to (x \to z)) \in L$.

Axiom 5. If $x, x \to y \in L$ (where $y \in S$), then $y \in L$.

L can be interpreted as the set of all logically valid sentences. It is straightforward to define the concept of consequence, once we have defined the concepts of the *sum* and the *product* of sentences:

DEFINITION 1. $x + y = \bar{x} \to y, x.y = \overline{(x \to \bar{y})}$ for all $x, y \in S$.[21]

Tarski goes on to show that this system of axioms and primitive concepts is equivalent to that of [Tarski, 1930b].

At the heart of [35+36] are two deductive systems, *each presented as an interpretation of Boolean algebra*. Tarski writes:

> "[T]wo calculi can be constructed which are very useful in metamathematical investigations, namely the *calculus of sentences* and the *calculus of deductive systems*; the first is a complete and the second a partial interpretation of the formal system which is usually called the *algebra of logic* or *Boolean algebra*." (p. 347)

Tarski identifies eight primitive concepts: 'B' is 'the universe of discourse', '$<$' denotes the relation of inclusion, '$=$' denotes identity, '$+$' and '.' denote respectively the *sum of* and *product* operations, '0' denotes the zero (or empty) element, '1' denotes the unit (or universal) element, and '$\bar{x}$' denotes the complement of the element x. There are seven postulates that suffice for Boolean algebra:

Postulate I. (a) If $x \in B$, then $x < x$; (b) if $x, y, z \in B, x < y$ and $y < z$, then $x < z$.

Postulate II. If $x, y \in B$, then $x = y$ if and only if both $x < y$ and $y < x$.

Postulate III. If $x, y \in B$, then (a) $x + y \in B$; (b) $x < x + y$ and $y < x + y$; (c) if $z \in B, x < z$ and $y < z$, then $x + y < z$.

Postulate IV. If $x, y \in B$, then (a) $x.y \in B$; (b) $x.y < x$ and $x.y < y$; (c) if $z \in B, z < x$ and $z < y$, then $z < x.y$.

[21] In order to define the concept of consequence, the definition of the sum and the product of sentences is extended by recursion to an arbitrary finite number of conjuncts and disjuncts:

Definition $\sum_{i=1}^{n} x_i = \prod_{i=1}^{n} x_i = x_1$ if $n = 1$ and $x_1 \in S$; $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n=1} x_i + x_n$ and $\prod_{i=1}^{n} x_i = \prod_{i=1}^{n=1} x_i . x_n$, if n is an arbitrary natural number > 1 and $x_1, x_2, \ldots, x_n \in S$.

The definition of *consequence* is as follows:

Definition For an arbitrary set XS the set $Cn(X)$ consists of those, and only those, sentences $y \in S$, which satisfy the following condition: either $y \in L$, or there exist sentences $x_1, x_2, \ldots, x_n \in X$ such that $(\prod_{i=1}^{n} x_i \to y) \in L$.

Postulate V. If $x, y, z \in B$, then (a) $x.(y+z) = x.y + x.z$ and (b) $x + (y.z) = (x+y).(x+z)$.

Postulate VI. (a) $0, 1 \in B$; (b) if $x \in B$, then $0 < x$ and $x < 1$.

Postulate VII. If $x \in B$, then (a) $\bar{x} \in B$, (b) $x.\bar{x} = 0$, and (c) $x + \bar{x} = 1$.

These are the axioms for the general, abstract theory of Boolean algebra. Boolean algebra is an abstraction from Boolean set algebra, in which B is the universe of sets, '$<$' denotes set-theoretical inclusion, '$+$' and '.' denote respectively set-theoretical union and intersection, 0 and 1 are the empty set and universal set respectively, and '$\bar{x}$' is the set-theoretical complement of x.

Tarski turns first to the *calculus of sentences* (which he also calls "the sentential algorithm" to avoid confusion with the expression 'sentential calculus'). Its universe of discourse is S, and two relations are defined on S: the relation of *implication*, denoted by '$\supset$', and the relation of *equivalence*, denoted by '$\equiv$'.

DEFINITION 2. (a) $x \supset y$ if and only if $x, y \in S$ and $x \rightarrow y \in L$; (b) $x \equiv y$ if and only if both $x \supset y$ and $y \supset x$.

Now Tarski states the following (easily checked) theorem:

THEOREM 3. *Postulates I–VII are satisfied under the following replacements: the symbols 'B', '$<$', and '$=$' are replaced respectively by 'S', '$\supset$', and '$\equiv$'; and '0' is replaced by the variable 'u' and '1' by the variable 'v', where it is assumed that $u \in S, \bar{u} \in L$, and $v \in L$. (See Theorem 4, p. 348.)*

This theorem shows that the Boolean algebra of Postulates I–VII can be derived from the Axioms 2–5.[22] And the result holds in the other direction. Tarski sums up:

> "We can thus assert that Axs. 2-5 form a system of statements which is equivalent to the system of postulates for the ordinary algebra of logic" (p. 349).

In making fully precise the relation between the system L and Boolean algebra, Tarski sets the stage for algebraic logic. The calculus of systems can now be seen as an interpretation of Boolean algebra. As Blok and Pigozzi put it, "...here for the first time can be found all the essential features of modern algebraic logic" [Blok and Pigozzi, 1988, pp. 45–6]; and as Vaught puts it, "...Tarski introduced here the now well-known Boolean algebra $B(L)$ canonically associated with a theory L; and he initiated the still continuing study of the algebras $B(L)$" [Vaught, 1986, p. 873]. Tarski observes that the present interpretation of Boolean algebra is easily modified so as to avoid the replacement of identity by another equivalence relation (viz. $=$). This modification yields a very early instance of an important and now-familiar category of Boolean algebras: the Lindenbaum-Tarski algebras.[23]

[22] Axiom I plays no part in the proof of the theorem.

[23] Tarski details the modification in fn. 1, p. 349. Given any $x \in S$, consider its equivalence

Despite the lasting significance of the calculus of systems, the "chief subject" of Tarski's [Tarski, 1935+1936] is another calculus, the *the calculus of deductive systems*, or for short, the *calculus of systems*. Recall that a deductive system is a system closed under consequence. Given the definition of *consequence* (see fn. 22), the notion of a deductive system can be characterized as follows:

> X is a deductive system iff $L \subseteq X \subseteq$ and if $x, x \to y \in X$ (where $y \in S$), then $y \in X$.

In the calculus of systems, there is a correlate of each primitive concept of Boolean algebra. The universe of discourse is the class of deductive systems. L is the zero system and S is the unit system. Inclusion and identity are interpreted in the ordinary set-theoretical way — and, similarly, the product of systems is interpeted as set-theoretical intersection. However, the addition of systems is not set-theoretical union — the union of two systems does not in general yield a new deductive system. Instead, the addition of systems X and Y, denoted by '$X \dot{+} Y$', is defined as $Cn(X \cup Y)$. Similarly, complementation is not set-theoretical complementation; instead, the complement $\bar{X}$ of system X is the sum of all systems Y disjoint with X (i.e. such that $X \cap Y = L$). The calculus of systems is a *partial interpretation of Boolean algebra*. Consider the postulates I–VII, and replace the variables 'x', 'y', 'z' respectively by 'X', 'Y', 'Z', and the constants 'B', '$<$', '$+$', '0' and '1' respectively by 'D', '$\subseteq$', '$\dot{+}$', 'L' and 'S'. Then every postulate except VIIc is satisfied. And the following consequence of VIIc is also satisfied:

VIId. If $X, Y \subseteq D$, and $X \cap Y = L$, then $Y \subseteq \bar{X}$. (See Theorem 6, p.351.)

So the law of excluded middle — $X \dot{+} \bar{X} = S$ — fails in the calculus of systems, and it is here that the calculus of systems differs essentially from the ordinary calculus of classes or sets. Tarski observes:

> "The formal resemblance of the calculus of systems to the intuitionistic sentential calculus of Heyting is striking: we might say that the formal relation of the calculus of systems to the ordinary calculus of classes is exactly the same as the relation of Heyting's sentential calculus to the ordinary sentential calculus" (p. 352)

In other words, the system of Postulates I–VI and VIIa,b,d is

> "a sufficient basis for a system of the algebra of logic which has the intuitionistic calculus as one of its interpretations" (*ibid.*)[24]

---

class $X \subseteq S$ — viz., the set of all sentences y such that $y \equiv x$. For these equivalence classes the relations and operations $\supset, +, .$, etc. are defined in the appropriate way; for example, '$X \supset Y$' says that $x \supset y$ for all $x \in X$ and $y \in Y$. O is the set of all sentences $x \in S$ such that $\bar{x} \in L$, and 1 is the set L. (Observe that the identity relation is left unchanged.) The set of equivalence classes is a Boolean algebra — the Lindenbaum-Tarski algebra determined by L.

[24] For a more detailed elaboration of these remarks, Tarski refers the reader to his paper [Tarski, 1938b] (in particular section 5) and Stone [1937–38]. (The references are given in [Tarski, 1983a, p. 352, fn. †], where a typographical error appears — 'X and VII" should read "XVII".)

Tarski goes on to state a large number of results about the calculus of systems that are important both for metamathematics and model theory.[25] In section 3, Tarski investigates axiomatizable and non-axiomatizable systems. In section 4, he studies irreducible[26] and complete[27] systems, and shows how to characterize certain classes of systems in terms of cardinality and structure. In section 5 and the appendix, Tarski applies these general results to particular deductive theories.[28]

We have seen, then, that Tarski's metamathematical investigations in [Tarski, 1935+1936] are carried out in the framework of Boolean algebra; they "do not transcend the boundary of the algebra of logic" (p. 350). In a footnote later added to [Tarski, 1935+1936], Tarski observes:

> "In fact the calculus of deductive systems outlined in this paper proves to coincide with what was somewhat later developed as the calculus of Boolean-algebraic ideals..."[29]

Subsequently, it was the study of Boolean algebras that took center stage for Tarski, not the calculus of systems. In his paper [Tarski, 1938b], for example, Tarski refers to "*general metamathematics*, i.e. the *theory of deductive systems*" as "[a]nother important realization of Boolean algebra" [Tarski, 1938b, p. 454].[30] For more on Tarski and algebraic logic, see Section 6 below.

2.5 Metamathematics and models

In Chapter VI of [Tarski, 1941a], Tarski provides a clear, textbook account of the methodology of the deductive sciences which is strikingly semantic in character. As we have just seen, the semantic concepts of *interpretation* and *realization* figure in [Tarski, 1935+1936] — but in [Tarski, 1941a] they are at the heart of Tarski's presentation of metamathematics.

[25]For a summary of some of these results, see Blok and Pigozzi [1988, p. 44], and Vaught [1986, p. 873].

[26]X is an irreducible system iff $X \neq L$ and for every deductive system Y such that $Y \subseteq X, Y = L$ or $Y = X$.

[27]X is a complete system iff X is a deductive system, $X \neq S$ and for all deductive systems Y such that $X \subseteq Y, Y = X$ or $Y = S$.

[28]These theories include the theory of dense orders, the theory of discrete orders, the theory of identity, the theory of atomistic Boolean algebras, the general theory of order, and the general theory of binary relations.

[29]Fn.1, p. 350. Given a Boolean algebra characterized by Postulates I–VII, an *ideal* is a non-empty subset K of B such that (1) if $x \in K$ and $y \in K$, then $x + y \in K$, and (2) if $x \in K$ and $y \in B$, then $x.y \in K$.

Tarski refers the reader to Stone [1936; 1937], with which there is significant overlap or connection.

[30]As Tarski observes (in fn †, p. 352 of [Tarski, 1935+1936]), [Tarski, 1938b] contains "a more exact and detailed formulation" of the relation mentioned above between the calculus of systems and the intuitionistic sentential calculus. But [Tarski, 1938b] proceeds in terms of Boolean algebra, and not at all in terms of the calculus of systems. It is only at the very end of the paper that Tarski remarks that the results shown to hold for the formal system of Boolean algebra also hold for every realization of this system, such as the theory of fields of sets and the calculus of systems. (For a summary of results in [Tarski, 1938b], see Blok and Pigozzi [1988, pp. 46–7].

Tarski starts out with a concrete example of a simple deductive theory, the theory of the congruence of line segments. Let the variables x, y, z range over line segments. There are two primitive terms 'S' and '$\cong$', where 'S' denotes the set of all line segments, and '$\cong$' denotes the relation of congruence. There are two axioms:

Axiom 1. For all $x \in S, x \cong x$.

Axiom 2. For all $x, y, z \in S$, if $x \cong z$ and $y \cong z$, then $x \cong y$.

Tarski goes on to derive two theorems: the first says that the congruence relation is symmetric, the second that the relation is transitive.

Tarski observes that though our knowledge of segments and congruence goes a long way beyond the axioms, this additional knowledge plays no role in the construction of the theory.

> "In particular, in deriving theorems from the axioms, we make no use whatsoever of this knowledge, and behave as though we did not understand the content of the concepts involved in our considerations, and as if we knew nothing about them that had not been expressly asserted in the axioms. We disregard, as it is commonly put, the meaning of the primitive terms adopted by us, and direct our attention exclusively to the form of the axioms in which these terms occur." (p. 122)

Accordingly, we can abstract away from our particular theory by replacing 'S' by a variable 'K' that ranges over all classes, and '$\cong$' by a variable 'R' that ranges over all (2-place) relations, so as to obtain:

Axiom I'. For all $x \in K, xRx$, and

Axiom II' For all $x, y, z \in K$, if xRz and yRz, then xRy.

Now the statements of the theory are logical statements; for example, Axiom I' says that the relation R is reflexive on K. And any theorem about congruent segments is now correlated with a general law in the domain of logic (for example, that any relation R for which the generalized Axioms I' and II' hold is symmetric).

At this point, Tarski introduces the notion of a *model* (or *realization*) *of the axiom system.* If, in a set K, a relation R is reflexive and has the property expressed by Axiom $2'$, then K and R together form a model of the axiom system I and II. One model is, of course, formed by the class of segments and the congruence relation — but this is not in any way privileged. Another model of the axiom system is provided by the universal class and the relation of identity. And another is formed by the set of all numbers — or any set of numbers — and the relation given by 'the difference between numbers x and y is an integer'. Every model of the axiom theory will also satisfy all theorems deduced from these axioms.

These general facts, Tarski says, "have many interesting applications in methodological researches" [Tarski, 1941a, p. 124]. For example, it may be proved from these facts that certain sentences cannot be deduced from the axiom system I and II. Consider the following sentence:

A. There exist two elements x and y of the set S for which it is not the case that $x \cong y$. (There exist two segments that are not congruent).

Though A is true, attempts to prove it from the axioms fail. Can we *show* that the attempt must fail? We can, by applying what Tarski here calls the *method of proof by interpretation.*[31]

> "If sentence A could be proved on the basis of our axiom system, then, as we know, every model of this system would satisfy that sentence; if, therefore, we succeed in indicating such a model of the axiom system which will not satisfy Sentence A, we shall prove thereby that this sentence cannot be deduced from Axioms I and II." (p. 125)

And such a model is easy to find — for example, consider the model formed by the set of integers, together with the relation 'the difference between numbers x and y is an integer'.

Abstracting away from any particular axiom system and its models, Tarski goes on to state "a general law from the domain of the methodology of the deductive sciences".[32]

> "Every theorem of a given deductive theory is satisfied by any model of the axiom system of this theory; and moreover, to every theorem there corresponds a general statement which can be formulated and proved within the framework of logic and which establishes the fact that the theorem in question is satisfied by any such model." (p. 127)

This law — a version of what Tarski calls the *law of deduction* — has "tremendous practical importance" (p. 127). Given an axiom system of a given theory, we will often find that the constants of another deductive theory form a model of that axiom system — that is, we find an interpretation of the axiom system of the original theory within the other theory.[33] And the validity of the theorems of the

[31]As we shall see, the method here is based on Tarski's definition of *logical consequence* in terms of models — see Section 3.4 below. The definition runs as follows:

The sentence X *follows logically* from the sentences of the class K if and only if every model of the class K is also a model of the sentence X. [Tarski, 1936b, p. 417].

We show that A does not follow logically from Axioms I and II by constructing a model of the axiom system which is not a model of A.

[32]In leading up to the statement of this law, Tarski makes the simplifying assumption that logic is the only theory preceding the given theory (see footnote 7 above).

[33]For example, the axiom system of arithmetic may be interpreted within geometry (enabling us to obtain a visual images of various facts in the field of arithmetic); and conversely, the axiom system of geometry has an interpretation within arithmetic (giving rise to analytic geometry, in which geometrical facts are investigated with the help of arithmetical or algebraic methods). Further, arithmetic is interpretable within logic (and hence, so is geometry). See [Tarski, 1941a, pp. 129–130].

first theory carries over to those of the second: '*All theorems proved on the basis of a given axiom system remain valid for any interpretation of the system.*" (p. 128) So in proving a theorem of the original theory we prove any number of others:

> "Every proof within a deductive theory contains — potentially, so to speak — an unlimited number of other analogous proofs." (p. 128)

These facts demonstrate "the great value of the deductive method from the point of view of economy of human thought" (*ibid*). They are also of far-reaching theoretical importance, since

> "they establish a foundation for various arguments and researches within the methodology of deductive sciences" (p. 128).

For example, the law of deduction is the theoretical basis for all proofs by interpretation.

In the final section of Chapter VI, Tarski stresses the broad scope of the methodology of the deductive sciences. As we have seen throughout the present section, it goes beyond the study of methods used in the construction of deductive sciences, and incorporates the study of deductive theories as *wholes*; consider, for example, investigations into the consistency and completeness of theories. Better, says Tarski, to call it "metalogic and metamathematics". Rather than the study of methods, it is become a general science of deductive sciences, requiring its own precise methods. Tarski writes:

> "Methodology has become like those sciences which constitute its own subject matter - it has assumed the form of a deductive discipline". (p. 140)

2.6 Some historical remarks

Understood in this suitably broad way, the methodology of the deductive sciences is, Tarski says, "a very young discipline".

> "Its intensive development began only twenty years ago - simultaneously (and, as it seems, independently) in two different centers: Gottingen under the influence of D. Hilbert and P. Bernays, and Warsaw, where S. Lesniewski and J. Lukasiewicz, among others, worked." [Tarski, 1941a, p. 19]

Its roots, however, go back a very long way. We can find an analysis of the deductive sciences in Aristotle's *Posterior Analytics*.[34] Much more recent, but still of

[34] [Tarski, 1935]. For example, according to Aristotle every demonstrative science has three elements: (1) what is posited, (2) the axioms, and (3) the attributes. Consider for illustration the case of geometry. In geometry, points and lines are posited, and their attributes are demonstrated "by means of the axioms and from previous conclusions as premisses" (76b9–10, Bk.1, Ch.10, *Posterior Analytics*, McKeon [1941, p. 24]).

historical significance for the methodology of the deductive sciences, are Bolzano's *Wissenschaftslehre* [1837], Pasch [1882], Hilbert's *Grundlagen der Geometrie*, and Veblen and Young [1910].

Applications of the deductive method have a very long history too. Of Euclid's *Elements*, Tarski writes:

> "For 2200 years, mathematicians have seen in Euclid's work the ideal and prototype of scientific exactitude." [Tarski, 1941a, p. 120, fn 1]

Tarski continues:

> "An essential progress in this field occurred only in the last 50 years, in the course of which the foundations of the basic mathematical disciplines of geometry and arithmetic were laid in accordance with all requirements of the present-day methodology of mathematics." (p. 120, fn.1)

Tarski mentions in particular Peano's *Formulaire de Mathematiques*, and Hilbert's *Grundlagen der Geometrie*. As they stood, Peano's and Hilbert's theories failed to meet one of the requirements of methodology, that the deductive theory be formalized. The development of formalized deductive theories started with Frege:

> "The first attempts to the present the deductive theories in a formalized form are due to Frege.... A very high level in the process of formalization was achieved in the works of the late Polish logician S. Lesniewski...." (p. 133)

It is clear that Tarski's work in metamathematics continued a long tradition, and that even the "young discipline" of the methodology of the deductive sciences was not original with Tarski. But it is also clear that Tarski took metamathematics to a new level. Corcoran writes:

> "The modern conception, and perhaps the very existence, of the methodology of deductive science as a separate science and in particular as a formalizable deductive science, is largely due to Tarski in that it was he who collected, analyzed, and codified emerging concepts and results, and it was he who most clearly and forcefully articulated and defended the possibility of such a science." (Corcoran, editor's introduction to [Tarski, 1983a, p. xvi]

3 FORMAL DEFINABILITY

As we have seen, one of the tasks of the methodology of the deductive sciences is to "make precise the meaning ... of important metamathematical concepts" [1930e]. In a series of papers published in the 1930s, Tarski carried out this project for a number of fundamental concepts — most notably *definability*, *truth*, and *logical consequence*. Tarski constructed precise definitions of each of these concepts.

Tarski distinguished two kinds of definition:

> "The words 'define', 'definable', etc., are used in two distinct senses: in the first sense it is a question of a *formal* relation of certain expressions to other expressions of a theory...; in the second sense of a *semantical* relation between objects and expressions." [Tarski, 1934, p. 386], [Tarski, 1935]

Broadly speaking, formal definitions involve only 'syntactic' notions, such as derivability, while semantic definitions involve word-world relations. In [Tarski, 1934], Tarski provided a formal definition of *definability.* In his papers [Tarski, 1931a; Tarski, 1933a] and [Tarski, 1936b], Tarski constructed semantic definitions for *definability*, *truth*, and *logical consequence* respectively. In this section, we will explore Tarski's formal treatment of definability, and its ramifications. In Section IV below, we turn to Tarski's semantic treatment of semantic definability, truth, and logical consequence.

The formal sense of 'definability' is investigated in Tarski's paper "Some methodological investigations on the definability of concepts" [Tarski, 1934]. Tarski constructs a formal definition of definability that draws on earlier work of Padoa's [Padoa, 1900].

Padoa takes a deductive theory to be a system of *undefined symbols* (or uninterpeted primitive terms) and a system of *unproved propositions* (or axioms). Padoa's chief aim is to characterize an *irreducible* system of undefined symbols, that is, a system in which no undefined symbol can be defined in terms of the others:

> "We say that *the system of undefined symbols is irreducible with respect to the system of unproved propositions* when *no symbolic definition of any undefined symbol can be deduced from the system of unproved propositions*, that is, when we cannot deduce from the system a relation of the form $x = a$, where x is one of the undefined symbols and a is a sequence of other such symbols (and logical symbols)." (pp. 121–22)[35]

Padoa's definition of an irreducible system contains an informal characterization of formal definability that Tarski makes explicit and precise in [Tarski, 1934]. Tarski considers deductive theories with the simple theory of types as their logical basis, and directs his attention to the extra-logical constants, or *terms*, of a given deductive theory. Let 'a' be a term, and B any set of terms. A *definition of the term 'a' by means of the terms of the set B* is a sentence of the following form:

(I) $(x)[x = a \leftrightarrow \phi(x, b', b'', \ldots)]$

[35]As Padoa observes, there is a parallel definition of an *irreducible (or "absolutely independent") set of propositions* — viz, a set where no member can be derived from the other members. (Padoa points out that a method of proving this irreducibility is already well-known: we can say that a proposition is not a logical consequence of the others if there is an interpretation in which it is false, and the others are true. The method is now standard in independence proofs. Padoa cites Peano [1899, p. 30], which contains a proof of the absolute independence of the axioms of arithmetic.) Like Padoa, Tarski presses the analogy between the concept *derivable sentence* on the one hand and *definable concept* on the other.

where '$\phi(x, b', b'', \ldots)$' is a sentential function with the sole free variable x, and in which no extra-logical constants other than $b', b'', \ldots$ of the set B occurs. Now let X be a set of sentences in which all terms of B occur. Tarski defines the notion of (formal) *definability* as follows:

DEFINITION 4. The term 'a' *is definable by means of the terms of the set B on the basis of the set X of sentences* if a definition of the term 'a' by means of the terms of B is derivable from the sentences of X. (See [Tarski, 1934, p. 299].)

Observe that this notion of definability involves only formal or 'syntactic' notions, in particular those of *derivability*, *sentential function*, and *term*.

Having defined the notion of formal derivability, Tarski goes on to characterize and justify Padoa's method for determining the irreducibility of a system of undefined symbols (or, equivalently, the undefinability of a term by means of other terms). Padoa offers only a sketch of this method:

> "Let us assume that, after an *interpretation* of the system of undefined symbols that verifies the system of unproved propositions has been determined, all these propositions are still verified if we suitably change the meaning of the undefined symbol x *only*. Then, since the meaning of x is not *individualized* once we have chosen an *interpretation* of the *other* undefined symbols, we can assert that it is impossible to deduce a relation of the form $x = a$, where a is a sequence of other undefined symbols, from the unproved propositions." [Padoa, 1900, p. 122]

Tarski describes Padoa's method this way:

> "In order, by this method, to show that a term 'a' cannot be defined by means of the terms of a set B on the basis of a set X of sentences, it suffices to give two interpretations of all extra-logical constants which occur in the sentences of X, such that (1) in both interpretations all sentences of the set X are satisfied and (2) in both interpretations all sentences of the set B are given the same sense, but (3) the sense of 'a' undergoes a change." [Tarski, 1934, p. 300]

Tarski goes on to present some results "which provide a theoretical justification for the method of Padoa" (ibid). Consider again a term a, and a set B of terms $b', b'', \ldots$, where a is not a member of B. Let $c', c'', \ldots$ be the terms other than $a, b', b'', \ldots$ that occur in the sentences of X. The conjunction of all the sentences of X is represented in the schematic form: '$\psi(a; b', b'', \ldots; c', c'', \ldots)$'. Suppose that in every sentence of X, variables are uniformly substituted for some or all terms, where it is assumed that none of these variables become bound. In this way, sentential functions are formed. The conjunction of all these sentential functions is represented schematically by making the same substitutions in '$\psi(a; b', b'', \ldots; c', c'', \ldots)$'. Tarski proves the following theorems:

THEOREM 1. The term 'a' is definable by means of the terms of the set B on the basis of the set X of sentences if and only if the formula

II. $\forall x(x = a \leftrightarrow \exists z', z'', \ldots \psi(a; b', b'', \ldots; z', z'', \ldots))$ is derivable from the sentences of X. (See [Tarski, 1934, p. 301].)

THEOREM 2. The term 'a' is definable by means of the terms of the set B on the basis of the set X of sentences if and only if the formula

III. $\forall x, x', y', y'', \ldots, z', z'', \ldots, t', t'', \ldots (\psi(x'; y', y'', \ldots; z, z', \ldots) \& \psi(x''; y', y'', \ldots; t', t'' \ldots) \rightarrow x' = x'')$ is derivable. (See [Tarski, 1934, p. 303].)

THEOREM 3. The term 'a' is not definable by means of the terms of the set B on the basis of the set X of sentences if and only if the formula

IV. $\exists x, x', y', y'', \ldots, z', z'', \ldots, t', t'', \ldots (\psi(x'; y', y'', \ldots; z, z', \ldots) \& \psi(x''; y', y'', \ldots; t', t'' \ldots) \& x' \neq x'')$ is consistent. (See [Tarski, 1934, p. 304].)

Theorem 3 follows immediately from Theorem 2.[36]

Theorem 3 "constitutes the proper theoretical foundation for the method of Padoa" [Tarski, 1934, p. 305]. For suppose that we want to establish that the term 'a' is not definable by the means of the terms of the set B. We apply the following procedure (see [Tarski, 1934, p. 305]). Take a deductive system Y, either shown to be consistent or assumed to be so. Look for terms $\bar{a}, \bar{\bar{a}}, \bar{b}', \bar{b}'', \ldots, \bar{c}', \bar{c}'', \ldots, \bar{\bar{c}}', \bar{\bar{c}}'', \ldots$ which satisfy the following three conditions:

(i) Replacing the terms $a, b', b'', \ldots c', c'', \ldots$ by $\bar{a}, \bar{b}', \bar{b}'', \ldots, \bar{c}', \bar{c}'', \ldots$ in all sentences of X yields sentences of Y.

(ii) Replacing the terms $b', b'', \ldots$ by the terms $\bar{b}', \bar{b}'', \ldots$, and the terms $a, c', c'', \ldots$ by the terms $\bar{\bar{a}}, \bar{\bar{c}}', \bar{\bar{c}}'', \ldots$ in all sentences of X yields sentences of Y.

(iii) The system Y contains the formula $\bar{a} \neq \bar{\bar{a}}$.

If these conditions are met, then the following conjunction belongs to Y:

$$\psi(\bar{a}; \bar{b}', \bar{b}'', \ldots; \bar{c}', \bar{c}'', \ldots) \& \psi(\bar{\bar{a}}; \bar{b}', \bar{b}'', \ldots; \bar{\bar{c}}', \bar{\bar{c}}'', \ldots) \& \bar{a} \neq \bar{\bar{a}}.$$

Formula IV of Theorem 3 is an easy consequence of this conjunction, and so formula IV belongs to Y. And since Y is consistent, formula IV is consistent. So by Theorem 3, the term 'a' is not definable by means of the terms of B. And here is Padoa's method (somewhat extended, since the procedure accommodates the occurrence of terms in the sentences of X that are different from 'a' and the terms in B.)

Thus Tarski has shown that we are justified in using Padoa's method in the simple theory of types. (As Tarski remarks in a footnote subsequently added to [Tarski, 1934] — fn.2, p.300 — Beth later made a very significant advance, proving

[36]To show that Theorem 3 follows from Theorem 2, Tarski observes that formula (II) is equivalent to the negation of formula (I), and applies the principle that a formula is not provable if and only if its negation is consistent.

Theorems I and II for a much wider class of deductive theories — the theories of first-order logic.)[37] The method has significant practical importance. Suppose we are considering the primitive terms of a given deductive theory. By applying Padoa's method as many times as there are primitive terms, we may determine whether or not the terms are mutually independent (or in Padoa's terminology, whether or not they form an *irreducible system.*) If the terms are not mutually independent, then we may eliminate the unnecessary ones — and this may lead to a simplification of the axiom system.

To illustrate this, Tarski considers n-dimensional Euclidean geometry. We may take the terms 'equidistant' and 'lying between' as the sole primitive terms of an axiom system of geometry. (The term 'equidistant' is the *metrical primitive term* and 'lying between' is the *descriptive primitive term* of geometry.) Tarski announces that by Padoa's method it can be shown that the term 'equidistant' is not definable by means of the term 'lying between'. However, for 2 dimensions and higher, the term 'lying between' is definable by means of the term 'equidistant'[38] — and the definition may be constructed according to the formula II in Theorem 1 above. Accordingly, Padoa's method yields a simpler axiom system for n-dimensional geometry ($n > 1$) in which 'equidistant' is the only primitive term.

So the formally defined notion of definability is the key to Tarski's presentation and justification of Padoa's method. In the second half of [Tarski, 1934], Tarski takes the notion of definability in another direction, using it to characterize what he calls "the problem of the completeness of concepts". This requires the notion of an "essentially richer" set, which rests on the notion of definability. Let Y be a set of sentences, and let $X \subseteq Y$. Then Y is *essentially richer* than X if and only if there are terms in Y (i.e. occurring in the sentences of Y) that are not in X, and these terms are not definable by means of the terms in X (not even on the basis of Y). Now we can define the notion of a set of sentences being *complete* with respect to its terms.

DEFINITION 5. A set X of sentences is *complete with respect to its terms* if it is impossible to construct a categorical set Y of sentences which is essentially richer than X with respect to specific terms. (See [Tarski, 1934, p. 311].)[39]

---

[37] [Tarski, 1953].

[38] See Lindenbaum and Tarski [1927b]. Tarski credits Lindenbaum with the result that the two terms are independent of each other only in the case of one-dimensional geometry.

[39] Notice the restriction here to *categorical* sets of sentences, where a set is *categorical* if any two interpretations of it are isomorphic. Tarski expresses the intuitive significance of the notion of categoricity this way:

> "A non-categorical set of sentences (especially if it is used as an axiom system of a deductive theory) does not give the impression of a closed and organic unity and does not seem to determine precisely the meaning of the concepts contained in it." [Tarski, 1934, p. 311]

Suppose we were to drop this restriction, and say that a set X of sentences is complete with respect to its terms if and only if it is impossible to construct an essentially richer set Y. Tarski points out (pp. 308–9) that, given this unrestricted definition, there will be no complete sets (apart from some trivial cases). For suppose that X is consistent, and does not contain all extra-logical constants, or terms. Now add to X an arbitrary logically provable sentence that contains

The problem of the completeness of concepts can now be stated: given a set of sentences, is it complete or incomplete with respect to its specific terms? For specific examples, Tarski again turns to geometry. Consider *descriptive one-dimensional geometry*, the geometry of points and subsets of a straight line in which the only primitive concept is that of *lying between*. Consider the (easily formulated) categorical system X_1 which contains 'lying between' as its only primitive term. Now we can extend X_1 to the full metrical geometry of the straight line by adding 'equidistant' as a primitive term. Call this extended system X_2. As we saw above, the concept of *equidistance* cannot be defined in terms of the concept of *lying between*. Since X_2 is categorical and essentially richer than X_1, it follows that X_1 is not complete with respect to its terms.

Now the system X_2, like X_1, is not complete with respect to its terms. But we can extend X_2 to a system that *is*. We extend X_2 to an essentially richer categorical set of sentences X_3 by adding two new primitive terms, say '0' and '1', together with an axiom saying that these terms denote two distinct points. It turns out that X_3 is complete with respect to its terms. (X_3 is, in fact, formally identical to the arithmetic of the real numbers.)[40]

As a whole, Tarski's paper [Tarski, 1934] may be regarded as an investigation into the concept of *concept*. The formal definition of *definable concept* is the core of Tarski's solution to two problems: the problem of the definability and mutual independence of concepts, and the problem of the completeness of concepts. As we have seen, this definition turns on *formal* relations between expressions, such as derivability. No semantic word-world relations are involved. The reverse is true for Tarski's *semantic* definitions of definability, truth and logical consequence, to which we now turn.

---

a term not occurring on the sentences of X. This new set Y is an essentially richer set than X. For, as Tarski puts it,

> "...it is obvious that the new extra-logical constant cannot be defined by means of the terms of X. In fact the only sentence of Y in which this constant occurs is logically provable, and thus true independently of the meaning of the specific terms contained in it." [Tarski, 1934, p. 309, fn. 1].

[40]In [Tarski, 1934], Tarski goes on to prove a general theorem concerning completeness with respect to terms. First we need the notion of a *monotransformable* set of sentences:

Definition A set of sentences is *monotransformable* if for any two interpretations of this set there is at most one relation which establishes the isomorphism of the two interpretations.

(Compare and contrast the notion of a categorical set of sentences: a set of sentences is categorical if for any two interpretations of it there is at least one relation which establishes the isomorphism of the two interpretations.)

Tarski's theorem is as follows:

THEOREM 4. Every monotransformable set of sentences is complete with respect to its terms. [Tarski, 1934, p. 314]).

4 SEMANTICS: DEFINABILITY, TRUTH, CONSEQUENCE

In three celebrated papers — "On definable sets of real numbers" [Tarski, 1931a], "The concept of truth in formalized languages" [Tarski, 1933a], and "On the concept of logical consequence" [Tarski, 1936b] — Tarski constructed precise definitions of *definability*, *truth*, and *logical consequence*. We turn to these constructions in Sections 4.1, 4.2 and 4.3 below. As we will see, at the core of each definition is the semantic notion of *satisfaction of a sentential function*. These semantic definitions were important contributions to the development of model theory, and in 4.4 I make some historical remarks about model theory.

At the outset of [Tarski, 1931a], Tarski notes that mathematicians are extremely wary of the notion of definability: "their attitude is one of distrust and reserve" [Tarski, 1931a, p. 110]. Tarski finds this perfectly understandable: the semantical notion of definability is ambiguous and subject to well-known paradoxes, such as Richard's, König's, and Berry's.[41] Tarski makes similar remarks about truth:

> 'For although the meaning of the term 'true sentence' in colloquial language seems to be quite clear and intelligible, all attempts to define this meaning more precisely have hitherto been fruitless, and many investigations in which this term has been used and which started with apparently evident premisses have often led to paradoxes and antinomies.... The concept of truth shares in this respect the fate of other analogous concepts in the domain of the semantics of language". [Tarski, 1933a, p. 152][42]

Small wonder, then, that semantical concepts have an "evil reputation". [Tarski, 1933a, p. 252]

In his papers on definability, truth and logical consequence, Tarski's aim is to find rigorous characterizations of these notions and thereby place them beyond suspicion and 'safe' for theorists to use. Prior to Tarski, it was generally held that these semantic notions were not the business of the mathematician or logician. Of the notion of definability, Tarski writes:

> "The distrust of mathematicians towards the notion in question is reinforced by the current opinion that this notion is outside the proper limits of mathematics altogether." [Tarski, 1931a, p. 110]

Tarski sought to reverse "the current opinion". He identified a metamathematical task here regarding definability, that of

> "making its meaning more precise, of removing the confusions and misunderstandings connected with it, and of establishing its fundamental properties...." [Tarski, 1931a, p. 110]

Tarski continues:

[41] See Richard [1905], König [1905], and for Berry's paradox, Russell [1908].

[42] For related remarks about the concept of logical consequence, see [Tarski, 1936b, p. 409].

> 'I believe I have found a general method which allows us to construct a rigorous metamathematical definition of this notion." [Tarski, 1931a, p. 111]

Tarski later applied the same method to construct his definitions of truth and logical consequence.

According to Tarski, the method allows the construction of definitions that are not only "formally correct" but also "materially adequate", in the sense that they "grasp the current meaning of the notion as it is known intuitively" (p.129 VI). At the core of Tarski's method is the notion of *satisfaction of a sentential function*, which appears in print for the first time in [Tarski, 1931a], Tarski's paper on definable sets of real numbers. We will now focus on [Tarski, 1931a] and the notion of definability.

4.1 Semantic definability

In [Tarski, 1931a], Tarski restricts his analysis of the notion of definability to just one category of objects: sets of real numbers.

> "The problem set in this article belongs in principle to the type of problems which frequently occur in the course of mathematical investigations. Our interest is directed towards a term of which we can given an account that is more or less precise in its intuitive content, but the meaning of which has not at present been rigorously established, at least in mathematics. We then seek to construct a definition of this term which, while satisfying the requirements of methodological rigour, will also render adequately and precisely the actual meaning of the term. It was just such problems that the geometers solved when they established the meaning of the terms 'movement', 'line', 'surface', or 'dimension' for the first time. Here I present an analogous problem concerning the term 'definable set of real numbers'. (pp. 111–112)

Tarski works within the framework of the simple theory of types (where the variables of the first order range over individuals, variables of the second order range over sets of individuals, and so on). The notion of a sentential function is introduced recursively. There are four *primitive sentential functions* denoted by '$\in$', '$\nu(x)$', '$\mu(x, y)$', '$\sigma(x, y, z)$'. The last three are special to the theory of real numbers, where the first-order variables range over reals, and '$\nu(x)$' says that $x = 1$, '$\mu(x, y)$' says that '$x \leq y$', and '$\sigma(x, y, z)$' says that $x + y = z$. So the object of Tarski's investigation is the system of reals with the primitive concepts 1, $\leq$, and $+$, or the system $\langle R, 1, \leq, + \rangle$ for short. There are five *fundamental operations* on sentential functions: negation, logical sum '+' (disjunction), logical product '.' (conjunction), universal quantification, and existential quantification. The set of all sentential functions is the smallest set which contains the primitive functions and is closed under the five operations.

Tarski explicitly distingishes system and metasystem:

> "For each deductive system it is possible to construct a particular science, namely the 'metasystem', in which the given system is subjected to investigation" (p. 116).

Tarski stresses that the notion of definability is *relative* — we can define, in the metasystem, the notion of definability *for the given system*. Definability is relative to a system — and if we ignore this characteristic, we will fall into paradox.[43] In the present case, expressions such as 'variable of the nth order' and 'sentential function' are terms of the metasystem, applying as they do to particular expressions of the system under investigation. We can also introduce into the metasystem arithmetical notions, like *real number*, *set of real numbers*, and so on. The metasystem now has the resources to define the following phrase:

> A finite sequence of objects satisfies a given sentential function.

In [Tarski, 1931a], Tarski makes no attempt to define this phrase precisely — a rigorous treatment is given in [Tarski, 1933a]. Instead, Tarski presents some simple illustrations. For example, the primitive sentential function '$\sigma(x, y, z)$' is satisfied by all sequences of three real numbers x, y, z such that $x = y + z$; the sentential function '$\nu(x).\nu(y).\mu(y, z)$' is satisfied by all sequences of three real numbers x, y, z such that $x = 1 = y \leq z$; and the sentential function '$\exists z \exists u(\sigma(x, y, z).\mu(u, z).\nu(u)$' is satisfied by sequences of two real numbers x and y where $x \geq y + 1$. Tarski notes that these examples indicate the possibility of establishing a 1-1 correspondence between the members of the sequence and the free variables in the sentential function. In the limit case, a sentential function without free variables — that is, a sentence — "is satisfied either by the empty sequence or by no sequence, according as the sentence is true or false" (p. 117).

Of special significance are sentential functions with one free variable, where instead of (unit) sequences, we can speak of *objects* satisfying the given function. We are led naturally to the notion of a definable set:

> "Consequently, a function which contains a variable of order 1 as its only free variable determines a certain set of individuals, which, in particular, may be a certain set of real numbers. The sets thus determined by sentential functions are precisely the *sets definable in the arithmetical system considered*." (p. 118)

[43] On p. 119, Tarski observes that there are only denumerably many definable sets (because there are only denumerably many sentential functions), but non-denumerably many sets of numbers. So there exist undefinable sets. Tarski goes on:

> "More than that, it is known that, with every denumerable family F of sets of numbers, a uniquely determined set $F*$ can be correlated that does not belong to F; taking for F the family of all definable sets, we get $F*$, an example of a *set of numbers defined in terms of the metasystem, but not definable in the system itself*." (p. 119)

If we ignore the system/metasystem distinction, we will be landed in contradiction and paradox. (In a footnote added later, Tarski refers the reader to [Tarski, 1948b, pp. 108f] for a precise definition of $F*$.)

Tarski now states the metamathematical definition of definable sets of real numbers as follows:

(M) A set X is a *definable set* if there is a sentential function which contains some variable of order 1 as its only free variable, and which satisfies the condition that, for every real number $x, x \in X$ if and only if x satisfies this function. (p. 118)

For example, the following sets are definable: the set $\{0\}$, the set of all positive reals, and the set of all reals x such that $0 \leq x \leq 1$; these sets are determined respectively by the sentential functions '$\sigma(x,x,x)$', '$\neg\sigma(x,x,x).\exists y(\mu(y,x).\sigma(y,y,y))$', and '$\exists y \exists z(\mu(y,x).\sigma(y,y,y).\mu(x,z).\nu(z))$'.

So Tarski sketches here, for the first time, a rigorous definition of definability in metamathematical terms - in particular, in terms of the metamathematical notion of satisfaction. But now Tarski takes further step: he offers a (partial) reconstruction of the notion of definability *in purely mathematical terms.*[44] Tarski writes:

> "Under this new definition the notion of definability does not differ from other mathematical notions and need not arouse either fears or doubts; it can be discussed entirely within the realm of normal mathematical reasoning." (p. 111)

This purely mathematical reconstruction rests on the following observation: each sentential function determines a certain set of finite sequences, namely, the set of all finite sequences that satisfy it. So we can replace the metamathematical concept of a sentential function by its mathematical analogue — the notion of a set of sequences.

In particular, we can introduce the *primitive* sets of sequences, those which are determined by the primitive sentential functions. Tarski goes on to define five operations on sets of sequences: complementation, addition, multiplication, and summation and multiplication with respect to the kth members of the sequences, corresponding to the five operations on sentential functions.[45] Then the family (or set) *Df* of definable sets of finite sequences of real numbers may be defined along the following lines:

> *Df* is the intersection of all the families of sets K which satisfy the following conditions: (i) the primitive sets belong to K; (ii) K is closed under the five operations.[46]

[44] No total reconstruction is possible on pain of paradox; Tarski's partial reconstruction provides a mathematical definition of definable sets of individuals of order n, for a given n. The construction is carried out in detail for $n = 1$, and outlined for $n = 2$; this suffices, says Tarski, to make clear the method for all higher levels.

[45] The operations of complementation, addition, multiplication correspond to the sentential operations of negation, disjunction, and conjunction; they also correspond to the usual Boolean operations. The operations of summation and multiplication with respect to the kth terms correspond to the existential and the universal quantifiers respectively.

[46] See Definition 9 (p. 128) for a precise formulation.

The family Df — characterized in purely mathematical terms — is exactly the family of sets of sequences which are determined by sentential functions. We have moved from metamathematical notions (sentential functions, satisfaction) to mathematical notions (sets of sequences, operations on sets of sequences). Now we can move on to a mathematical definition of the notion *definable set of reals*. Among the sets of sequences that belong to Df there are all the sets U of unit sequences. For any such U, form the set $U*$ of the members of U's unit sequences. $U*$ is a *definable set of real numbers*. The set D of all these $U*$s is the family of all *definable sets of real numbers*.[47] The set D is identical to the set of definable sets of real numbers as metamathematically defined by (M). Here, then, we have reached a purely mathematical characterization of the notion of a definable set of real numbers. The fact that it coincides with the metamathematical characterization demonstrates its material adequacy.

Tarski goes on to state two very important theorems about Df and D, and sketches their proofs.

THEOREM 6. *A set S of sequences of real numbers is a member of Df iff S is a finite sum of finite products of elementary linear sets.*[48]

THEOREM 7. *A set X of real numbers belongs to the set D iff X is the sum of a finite number of intervals with rational end-points.*[49]

Theorem 6 has a special significance. Given the correlation between (mathematical) sets of sequences and (metamathematical) sentential functions, and between Boolean operations on sets and Boolean operations on sentential functions, there is a metamathematical analogue of Theorem 6. And this metamathematical theorem leads to metalogical results of great moment: the theory of the reals $\langle R, 1, \leq, + \rangle$ is complete and decidable. More on this in Section 5.1 below.[50]

---

[47]See Definition 10 (p. 128) for a rigorous formulation of the definition of definable sets of individuals.

[48]For the notion of an elementary linear set, we need the notion of a linear polynomial with integral coefficients, exemplified by $3x + 10y + 7z + 2$. Associated with this polynomial is a linear equation

(i) $3x + 10y + 7z + 2 = 0$ and a linear inequality

(ii) $3x + 10y + 7z + 2 > 0$.

Observe that associated with (i) are sequences of reals, each of which is a solution to (i). One such sequence is $\langle 2, -1.5, 1 \rangle$, since $x = 2, y = -1.5$ and $z = 1$ provides a solution to (i). We can make a parallel observation about (ii).

An *elementary linear set* is a set of all sequences of real numbers which are solutions either of a linear equation of the type of (i), or a linear inequality of the type of (ii). (For a more rigorous characterization, see Tarski [Tarski, 1931a, p. 132].)

[49]An *interval with rational end points* is a set of one of the following types, where a and b are arbitrary rational numbers: $\{x|x = a\}, \{x|x > a\}, \{x|x < a\}$, or $\{x|a < x < b\}$.

[50]Theorems 6 and 7 also have significance for analytic geometry. Near the end of [Tarski, 1931a, pp. 141–2], Tarski reports that Kuratowski drew his attention to a geometric interpretation of the concepts and operations of [Tarski, 1931a]. (For example, on this interpretation the operation of summation with respect to the kth members may be taken as projection parallel to the axis X_k.) Interpreted in this way, Theorems 1 and 2 take the form of certain theorems of analytic geometry. This geometric interpretation is investigated in Kuratowski and Tarski's [31b].

4.2 Truth

Tarski's seminal monograph 'The concept of truth in formalized languages' [Tarski, 1933a], and his later informal summary 'The Semantic Conception of Truth' [Tarski, 1944] are devoted to a single problem — the definition of truth. Tarski seeks a definition of truth that does justice to the classical Aristotelian conception of truth:

> "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true."[51]

Tarski writes:

> "Let us start with a concrete example. Consider the sentence "snow is white". We ask the question under what conditions this sentence is true or false. It seems clear that if we base ourselves on the classical conception of truth, we shall say that the sentence is true if snow is white, and that it is false if snow is not white. Thus, if the definition of truth is to conform to our conception, it must imply the following equivalence:
>
> The sentence "snow is white" is true if, and only if, snow is white." [1944, p. 667]

The account is readily generalized:

> "Let us consider an arbitrary sentence: we shall replace it by the letter 'p'. We form the name of this sentence and we replace it by another letter, say 'X'. We ask now what is the logical relation between the two sentences "X is true" and 'p'. It is clear that from the point of view of our basic conception of truth these sentences are equivalent. In other words, the following equivalence holds:
>
> (T) X is true if, and only if, p. [Tarski, 1944, p. 668]

Tarski continues:

> "Now at last we are able to put into a precise form the conditions under which we will consider the usage and definition of the term "true" as adequate from the material point of view: we wish to use the term "true" in such a way that all equivalences of the form (T) can be asserted, and *we shall call a definition of truth "adequate" if all these equivalences follow from it.*" [Tarski, 1944, p. 668]

Here, then, is an informal expression of Tarski's famous adequacy condition on a definition of truth.

However, Tarski stresses that

[51] Aristotle, *Metaphysics*, Γ, 7, 27.

> "*the attempt to construct a correct semantical definition of the expression 'true sentence' meets with very real difficulties.*" [Tarski, 1933a, p. 162]

This is primarily because of the well-known Liar paradox. Tarski presents the Liar along the following lines. Consider the following sentence:

> The sentence printed in this chapter on p. 43, lines 18–19, is not true.

Let the letter 's' be an abbreviation of this sentence. Now, we may assert the following instance of schema (T):

> 's' is true iff the sentence printed in this chapter on p. 43, lines 10–11, is not true.

We can establish empirically that

> 's' is identical to the sentence printed in this chapter on p. 43, lines 10–11, is not true.

By the logic of identity, we obtain

> 's' is true iff 's' is not true.

And we arrive at a contradiction.

Tarski identifies two assumptions that are essential to the construction of the Liar. The first assumption is this:

> "(I) We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term "true" referring to sentences of this language; we have also assumed that all sentences which determine the adequate usage of this term can be asserted in the language." [Tarski, 1944, p. 20]

Tarski calls a language with these properties "semantically universal" [Tarski, 1969, p. 89],[52] or "semantically closed" [Tarski, 1944, p. 20]. The second assumption is this:

> "(II) We have assumed that in this language the ordinary laws of logic hold." [Tarski, 1944, p. 20]

A possible response is to reject II and give up classical logic. For Tarski, this is not an option:

> "It would be superfluous to stress here the consequences of rejecting the assumption II, that is, of changing our logic (supposing this were possible) even in its more elementary and fundamental parts. We thus consider only the possibility of rejecting the assumption I." [Tarski, 1944, p. 21]

[52][Tarski, 1969] is a semi-popular account of the notions of truth and proof.

And so Tarski concludes that we should seek a definition of truth for languages that are not semantically universal.

This leads to a distinction between *object language* and *metalanguage*:

> "The first of these languages is the language which is 'talked about' and which is the subject-matter of the whole discussion; the definition of truth which we are seeking applies to the sentences of this language. The second is the language in which we 'talk about' the first language, and in terms of which we wish, in particular, to construct the definition of truth for the first language." [Tarski, 1944, pp. 22–23]

Tarski points out that the terms 'object language' and 'metalanguage' apply only in a relative way:

> "If, for instance, we become interested in the notion of truth applying to sentences, not of our original object language, but of its metalanguage, the latter becomes automatically the object language of our discussion; and in order to define truth for this language, we have to go to a new metalanguage — so to speak, to a metalanguage of a higher level. In this way we arrive at a whole hierarchy of languages." [Tarski, 1944, p. 22]

So in Tarski's view we can hope to construct a definition of truth for a language L only if L is not semantically universal, only if there is a metalanguage for L in which we can couch the definition. But now, according to Tarski colloquial or natural languages (such as English or Polish) *are* semantically universal. In fact, according to Tarski, we can say something stronger: they are universal, in the sense that they can say everything that can be said. Tarski writes:

> "A characteristic feature of colloquial languages (in contrast to various scientific languages) is its universality. It would not be in harmony with the spirit of this language if in some other language a word occurred which could not be translated into it; it could be claimed that 'if we can speak meaningfully about anything at all, we can also speak about it in colloquial language" [Tarski, 1933a, p. 164]

In the same vein, Tarski remarks on the "all-comprehensive, universal character" of natural language and continues

> "The common language is universal and is intended to be so. It is supposed to provide adequate facilities for expressing everything that can be expressed at all, in any language whatsoever; it is continually expanding to satisfy this requirement." [Tarski, 1969, p. 89]

It is the universality of natural languages, says Tarski, "which is the primary source of all semantical antinomies" [Tarski, 1933a, p. 164]. For if a language can say everything there is to say, then in particular, it can say everything there is to say about its own semantics. Such a language is semantically universal. But a semantically universal language, together with the usual laws of logic, leads to semantic paradox (see [Tarski, 1933a, pp. 164–5]). Tarski concludes:

> "The very possibility of a consistent use of the expression 'true sentence' which is in harmony with the laws of logic and the spirit of everyday language seems to be very questionable, and consequently the same doubt attaches to the possibility of constructing a correct definition of this expression." ([Tarski, 1933a, p. 165]. See also p. 267)

So Tarski turns away from natural language.[53] He writes:

> "I now abandon the attempt to solve our problem for the language of everyday life and restrict myself henceforth entirely to *formalized languages*." [Tarski, 1933a, p. 165]

For the sake of clarity, Tarski constructs a definition of truth for one particular formal language — though as he points out, the method applies generally to an extensive group of formalized languages. Tarski chooses as his object language the *calculus of classes*. This calculus is one interpretation of Boolean algebra, and a particularly simple fragment of mathematical logic. There are four constants — negation ('N'), disjunction or logical sum ('A'), the universal quantifier 'Π'), and inclusion ('I') — and an infinite supply of variables ('$x_{\prime}$', '$x_{\prime\prime}$', '$x_{\prime\prime\prime}$') which range over classes, and which may be arranged in a sequence (so that we may speak of the kth variable). Composite expressions, such as $Ix_{\prime\prime}x_{\prime\prime\prime}$, are formed in the obvious way.

The concept of truth for the calculus of classes is to be defined in a metalanguage. Accordingly,

> "the proper domain of the following considerations is not the language of the calculus of classes itself but the corresponding metalanguage. Our investigations belong to the *metacalculus of classes* developed in this metalanguage." [Tarski, 1933a, p. 169]

The metalanguage includes the logical expressions 'not', 'or', 'for all', and 'is included in', which have the same meanings as the constants of the object language — so the object langage can be translated into the metalanguage. The metalanguage includes further expressions of a general logical character, including set-theoretical expressions (such as 'is an element of', 'class' and 'natural number') and expressions from the logic of relations (such as 'domain' and 'ordered n-tuple'). There is one further important category of expressions in the metalanguage: the expressions of a *structural-descriptive character*. These are the names

[53]Despite Tarski's own misgivings, others have been less pessimistic — for example, Davidson in [Tarski, 1967], and a number of authors who offer a 'Tarskian' approach to the Liar in the sense that they attribute a hierarchical structure to natural languages (see, for example, Parsons [1974]], Burge [1979], Barwise and Etchemendy [1987], Gaifman [1992], Koons [1992]], Glanzberg [2001]).

of expressions of the object language. Tarski mentions the following names in particular: 'the negation sign' (abbreviated by 'ng'), 'the sign of logical sum' ('sm'), 'the sign of the universal quantifier' ('un'), 'the inclusion sign' ('in'), 'the kth variable' ('v_k'), and 'the expression which consists of two successive expressions x and y' ('$x^\frown y$'). Tarski observes that the metalanguage contains both an individual name and a translation of every expression (and so every sentence) of the object language; this is crucial for the construction of the definition of truth.

Tarski next turns to the metatheory, and presents the axiom system of the metalanguage. The axioms divide into two kinds. First there are the *general logical axioms* which provide a sufficiently comprehensive system of mathematical logic (we may take, for example, the axioms of *Principia Mathematica*). Axioms of the second kind deal with the structural-descriptive concepts. In his introduction to [83a], Corcoran stresses the significance of these axioms.

> "In the 'Wahrheitsbegriff', Tarski isolates as a primitive notion the fundamental operation of concatenation of strings and he presents, employing concatenation, the first axiomatic codification of string theory, thereby providing deductive foundations of scientific syntax." [Tarski, 1983a, p. xxi]

Tarski writes:

> "What we call metatheory is, fundamentally, the *morphology of language* — a science of the form of expressions - a correlate of such parts of traditional grammar as morphology, etymology, and syntax." (p. 251)

The axioms are as follows:

Axiom 1. ng, sm, un, and in are expressions, no two of which are identical.

Axiom 2. v_k is an expression if and only if k is a natural number distinct from 0; v_k is distinct from ng, sm, un, in, and also from v_l if $k \neq l$.

Axiom 3. $x^\frown y$ is an expression if and only if x and y are expressions; $x^\frown y$ is distinct from ng, sm, un, in, and from each of the expressions v_k.

Axiom 4. If x, y, z, and t are expressions, then we have $x^\frown y = zt$ if and only if one of the following conditions is satisfied: (a) $x = z$ and $y = t$; (b) there is an expression u such that $x = zu$ and $t = u^\frown y$; (c) there is an expression u such that $z = x^\frown u$ and $y = u^\frown t$.

Axiom 5. (The principle of induction) Let X be a class which satisfies the following conditions: (a) ng $\in X$, sm $\in X$, un $\in X$, and in $\in X$; (b) if k is a natural number distinct from 0, then $v_k \in X$; (c) if $x \in X$ and $y \in X$, then $x^\frown y \in X$. Then every expression belongs to the class X.

Axiom 5 tells us in a precise way that every expression consists of a finite number of signs.

In a series of definitions that follow these axioms (pp. 175–6), Tarski introduces three fundamental operations — negation ('$\bar{x}$' denotes the negation of x), disjunction ('$x + y$' denotes the disjunction of x and y), and universal quantification (denoted by '$\cap$'). Conjunction is defined in the usual way in terms of negation and disjunction, and the existential quantifier — denoted by 'U' — is defined in terms of the universal quantifier and negation. By means of these operations, compound expressions are formed from simpler ones. This leads to a definition of the notion of a sentential function. The simplest sentential function is an *inclusion*, an expression of the form $(\mathrm{in}^\frown v_k)^\frown v_l$, or $\iota_{k,l}$ for short. Compound sentential functions are formed by applying the three operations to inclusions any number of times. Sentences are then defined as a special case of sentential functions: they are sentential functions with no free variable. Corcoran remarks that Tarski here provides

> "the first formal presentation of a generative grammar" [Tarski, 1983a, p. xxi][54]

Tarski now turns to the axioms for the object language, the calculus of classes. Among these are axioms that suffice for the sentential calculus, in which negation and disjunction are the only constants. The remaining axioms form a system for the calculus of classes, a simplified version of Huntington's system [Huntington, 1904, p. 297]. Tarski completes the preparations for his definition of truth by defining a number of basic metamathematical notions, including *consequence*, *theorem*, *deductive system*, *consistency*, and *completeness* (pp. 180–85).

The definition of *true sentence* in the language of the calculus of classes is carried out in Section 3 of [Tarski, 1933a]. At the outset of this section, Tarski returns to his adequacy condition and now states it in a more precise way. (In what follows, S is the class of all meaningful sentences of the language of the calculus of classes.)

> "*Convention T.* A formally correct definition of the symbol 'Tr', formulated in the metalanguage, will be called an *adequate definition of truth* if it has the following consequences:
>
> (α) all sentences which are obtained from the expression

[54]Corcoran continues:

> 'It is to be regretted that many linguists, philosophers, and mathematicians know so little of the history of the methodology of deductive science that they attribute the basic ideas of generative grammar to linguists working in the 1950s rather than to Tarski (and other logician/methodologists) working in the early 1930s." [Tarski, 1983a, p. xxi]

Corcoran also remarks that although other authors had used ideas about scientific syntax prior to [Tarski, 1933a], Tarski was the first to present these ideas in a rigorous way — another instance where Tarski makes formally precise ideas already informally in use (see [Tarski, 1983a, p. xxi, fn. 13].

'$x \in \mathrm{Tr}$' if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage;

(β) the sentence 'for any x, if $x \in \mathrm{Tr}$ then $x \in S$' (in other words '$\mathrm{Tr} \subseteq S$')." (pp. 187–8)

Tarski observes that if the object language contained only a finite number of sentences, it would be straightforward to produce an adequate definition. In schematic form, it would look like this:

$$x \in \mathrm{Tr} \text{ iff } (x = x_1 \text{ and } p_1) \text{ or } x = x_2 \text{ and } p_2 \text{ or } \ldots \text{or } x = x_n \text{ and } p_n$$

where 'x_1', 'x_2', ...'x_n' are replaced by structural-descriptive names of sentences of the object language, and 'p_1', 'p_2', ...'p_n' by the corresponding translations of these sentences into the metalanguage. But if the object language contains infinitely many sentences, as is the case with the language of the calculus of classes, then we cannot proceed in this 'list-like' way — we cannot formulate infinitely long definitions in the metalanguage. Consequently, Tarski turns to the recursive method. Truth is to be defined in terms of the more basic notion of satisfaction, and satisfaction is defined recursively.

We saw above that in [Tarski, 1931a] Tarski sketches the notion of satisfaction. But now, for the first time in print, Tarski provides a rigorous definition of satisfaction (specific, of course, to the language of the calculus of classes).

"*Definition 22* The sequence f *satisfies* the sentential function x if and only if f is an infinite sequence of classes and x is a sentential function and if f and x are such that either (α) there exist natural numbers k and 1 such that $x = \iota_{k,l}$ and $f_k \subseteq f_l$; (β) there is a sentential function y such that $x = \bar{y}$ and $f$does not satisfy the function y; (γ) there are sentential functions y and z such that $x = y + z$ and f either satisfies y or satisfies z; or finally (δ) there is a natural number k and a sentential function y such that $x = \cap_k y$ and every infinite sequence of classes which differs from f in at most the kth place satisfies the function y." (p. 193)

It is remarkable how close this definition is to contemporary textbook definitions — that is a tribute to the precision and clarity of Tarski's work.

The definition of truth now follows. Members of sequences correspond to variables with respect to their indices. Whether or not a sequence satisfies a sentential function depends only on the members of the sequence that correspond to the free variables of the sentential function. In the limit case, where the sentential function has no free variables, the members of the sequence do not matter at all. As Tarski puts it:

> "Thus in the extreme case, when the function is a sentence, and so contains no free variable (which is in no way excluded by Def. 22), the satisfaction of a function by a sequence does not depend on the properties of the terms of the sequence at all. Only two possibilities then remain: either every infinite sequence of classes satisfies a given sentence, or no sequence satisfies it.... The sentences of the first kind ... are the *true sentences*; those of the second kind ... can correspondingly be called the *false sentences*." (p. 194)

So we arrive at Tarski's definition of truth:

> "*Definition 23* x is a true sentence — in symbols $x \in \mathrm{Tr}$ — if and only if $x \in S$ and every infinite sequence of classes satisfies x." (p. 195)

This definition is clearly formally correct - so the question now is whether it is materially adequate. The answer is in the affirmative: the definition is adequate in the sense of Convention T. Tarski points out that a rigorous proof of this would require a meta-metatheory, since it is a result about the metatheory. To avoid any such shift of levels, Tarski adopts an alternative method — the "empirical method", according to which we confirm the adequacy of the definition by way of concrete examples. The example Tarski considers is the sentence $\cap_1 U_2 \iota_{1,2}$. By successive applications of the relevant clauses of Definition 22, it is straightforward to derive:

$\cap_1 U_2 \iota_{1,2} \in \mathrm{Tr}$ iff for every class a there is a class b such that $a \subseteq b$.[55]

This is a so-called 'T-sentence', one of the sentences mentioned in clause (α) of Convention T. And there is nothing special about this derivation - we can proceed in an analogous way with any sentence of the object language. Tarski concludes:

> "*We have succeeded in doing for the language of the calculus of classes what we tried in vain to do for colloquial language: namely, to construct*

[55] Following Tarski (p. 196), observe first that the sentential function $\iota_{1,2}$ is satisfied by exactly those sequences f such that $f_1 \subseteq f_2$. So the negation $\overline{\iota_{1,2}}$ is satisfied by exactly those sequences f such that $f_1 \subseteq f_2$. Consequently,

> "a sequence f satisfies the [sentential] function $\cap_2 \iota_{1,2}$ if every sequence g which differs from f in at most the 2nd place satisfies the function $\overline{\iota_{1,2}}$ and thus verifies the formula $g_1 \subseteq g_2$. Since $g_1 = f_1$ and the class g_2 may be quite arbitrary, only those sequences f satisfy the function $\cap_2 \overline{\iota_{1,2}}$ which are such that $f_1 \subseteq b$ for any class b. If we proceed in ananalogous manner, we reach the result that the sequence f satisfies the function $\cup_2 \iota_{1,2}$, i.e. the negation of the function $\cap_2 \overline{\iota_{1,2}}$ only if there is a class b for which $f_1 b$ holds. Moreover the sentence $\cap_1 U_2 \iota_{1,2}$ is only satisfied (by an arbitrary sequence f) if there is for an arbitrary class a, a class b for which $a \subseteq b$. Finally by applying Def. 23 we at once obtain one of the theorems which were described in the condition (α) of the convention T: $\cap_1 U_2 \iota_{1,2} \in \mathrm{Tr}$ iff for every class a there is a class b such that $a \subseteq b$.
>
> From this we infer without difficulty, by using the known theorems of the calculus of classes, that $\cap_1 U_2 \iota_{1,2}$ is a true sentence." (p. 196)

> *a formally correct and materially adequate semantical definition of the expression 'true sentence'.*" (pp. 208–9, Tarski's emphasis)

Although Tarski has defined truth for a specific language, the method of construction applies to 'many other formalized languages" (p. 209), as Tarski shows in Section 4. For each particular object language and object theory, we will need to construct a corresponding metalanguage and metatheory. The construction will follow the pattern indicated in the case of the calculus of classes. The metalanguage will always contain three groups of expressions: expressions of a general logical kind, expressions with the same meaning as the constants of the object language, and expressions of the structural-descriptive type. And the axiom system of the metatheory will include three groups of axioms: axioms of a general logical kind, axioms which provide for the translations into the metalanguage of all theorems of the object theory, and axioms for the structural-descriptive concepts. We will go on to define the notions of 'sentential function', 'sentence' and 'satisfaction', and define the notion of 'true sentence' in terms of satisfaction. Tarski concludes:

> "A. For every formalized language a formally correct and materially adequate definition of true sentence can be constructed in the metalanguage with the help only of general logical expressions, of expressions of the language itself, and of terms from the morphology of language - but under the condition that the metalanguage possesses a higher order than the language which is the object of investigation." (postscript to [Tarski, 1933a, p. 273])
>
> This condition on the metalanguage is crucial. For:
>
> "B. If the order of the metalanguage is at most equal to that of the language itself, such a definition cannot be constructed." (p. 273)[56]

In particular, if the metalanguage is identical to, or interpretable in the object language, it is impossible to construct an adequate definition of truth. In Section

[56]The order of a language is the smallest ordinal number which is greater than the orders of all variables occurring in the language (see p. 270). Variables ranging over individuals are assigned the order 0; variables ranging over classes of individuals and over relations between individuals are assigned the order 1; variables ranging over classes of classes of individuals and over relations between classes of individuals are assigned the order 2; and so on. The order of the language of the calculus of classes is 2, since its variables (ranging over classes of individuals) are of order 1. The language of the general theory of classes (introduced in Section 5) is a language of infinite order. Its variables range over classes, classes of classes, classes of classes of classes, and so on — and so its order is ω. (For the notion of the order of an expression, see the Postscript pp. 269–272, which revises Tarski's earlier account in section 4, p. 218).

Tarski's summary conclusion A from the Postscript is a significant revision of his summary in Section 6. In Sections 4 and 5, Tarski works with Husserl's notion of *semantical category*, introduced by Lesniewski into metamathematical investigations (Tarski cites Lesniewski [1929], especially pp. 14 and 68.) In the Postscript Tarski dispenses with the theory of semantical categories, and as a consequence finds that "the range of the results obtained has been essentially enlarged, and at the same time the conditions for their application have been made more precise" (p. 273).

5, Tarski takes as object language the language of the general theory of classes, and considers a metalanguage which may be interpreted in the object language — so that every sentence of the metalanguage is correlated with an equivalent sentence of the object language. Suppose that under these circumstances we can construct in the metalanguage a correct definition of truth. Tarski writes:

> "It would then be possible to reconstruct the antinomy of the liar in the metalanguage, by forming in the language itself a sentence x such that the sentence of the metalanguage which is correlated with x asserts that x is not a true sentence."(p. 248)

Let us take a close look at Tarski's reconstruction of the liar in the metalanguage. We start with the language of the general theory of classes. The variables of this language range over individuals, classes of individuals, classes of classes of individuals, and so on. Following Tarski, let 'n' be a variable ranging over classes of classes of individuals, and treat numbers in the Russell–Whitehead way. Then the range of 'n' includes the natural numbers.[57] Tarski observes that we may construct a sentential function ι_k with sole free variable 'n' which says, given a natural number k, that the class whose name is represented by 'n' is identical with k (see p. 249).

Tarski further observes that it is straightforward to set up a correspondence between the expressions of the language and the natural numbers. That is, we can define in the metalanguage an infinite sequence $\varnothing$ of expressions in which every expression occurs exactly once; we will let '$\varnothing_n$' denote the nth expression in the sequence. Consequently, we can correlate with every operation on expressions an operation on natural numbers, with every class of expressions a class of natural numbers, and so on. "In this way", Tarski writes,

> "the metalanguage receives an interpretation in the arithmetic of the natural numbers and indirectly in the language of the general theory of classes." (pp. 249–50)[58]

That is, the metalanguage is interpretable in the object language.

Now suppose that we have defined in the metalanguage the class Tr of true sentences of the object language. Let the symbol '$U * y$' denote the existential quantification of the sentential function y with respect to the variable 'n'. Consider the expression:

$$U * (\iota_n.\varnothing_{n)} \notin \text{ Tr.}$$

The expression '$\iota_n.\varnothing_n$' denotes the conjunction of the sentential functions denoted by ι_n and $\varnothing_n$. And '$U * (\iota_n.\varnothing_n)$' denotes the existential quantification of this

[57] See Tarski p. 233, fn.1. For example, the number 1 is the class of all classes which have exactly one individual as a member.

[58] The theory of natural numbers can be developed in the general theory of classes. Tarski writes that the general theory of classes "suffices for the formulation of every idea which can be expressed in the whole language of mathematical logic" (pp.241–2).

conjunction with respect to 'n'. As a whole, the expression '$U * (\iota_n.\varnothing_{n)} \notin \mathrm{Tr}$' says that this existential quantification is not a true sentence of the object language.

The expression '$U * (\iota_n.\varnothing_{n)} \notin \mathrm{Tr}$' itself is a sentential function of the metalanguage with 'n' as its sole free variable. Given that we can always correlate sentential functions with arithmetical functions, this sentential function may be correlated with another function equivalent to it but expressed solely in the terms of arithmetic. Let us schematically refer to this arithmetical function by '$\psi(n)$'. We have:

1. for any $n, U * (\iota_n.\varnothing_n) \notin \mathrm{Tr}$ if and only if $\psi(n)$.

Since the general theory of classes provides a foundation for the theory of natural numbers, it follows that '$\psi(n)$' is a sentential function of the object language. So it will appear in the sequence $\varnothing$ — let it appear at the kth place. Now instantiate (1) to 'k':

2. $U * (\iota_k.\varnothing_k) \notin \mathrm{Tr}$ if and only if $\psi(k)$.

Remember that we have assumed that the definition of Tr is adequate. Accordingly, condition (α) of Convention T holds. Now $U * (\iota_k.\varnothing_k)$ is a structural-descriptive name of a sentence of the object language. The sentence that it names says the following: "there is an n such that $n = k$ and $\psi(n)$", or more simply, "$\psi(k)$". So by condition (α) we obtain:

3. $U * (\iota_k.\varnothing_k) \notin \mathrm{Tr}$ if and only if $\psi(k)$.

And (2) and (3) yield a contradiction.

Notice that '$\psi(k)$', a sentence of the object language, is correlated with '$U * (\iota_k.\varnothing_k) \notin \mathrm{Tr}$', a sentence of the metalanguage, and the latter says of the former that it is not true. Thus, to repeat Tarski's words, we have formed in the object language itself "a sentence x such that the sentence of the metalanguage which is correlated with x asserts that x is not a true sentence." In reconstructing the Liar in this way, Tarski has used the so-called *diagonal argument*; the distinctive 'diagonal' step is the inference to (2), where we replace the variable 'n' in the correlates '$U * (\iota_n.\varnothing_n) \notin \mathrm{Tr}$' and '$\psi(n)$' by k, the index of '$\psi(n)$' in $\varnothing$.[59] One can also see close analogies with the heterological paradox, generated by the phrase 'is not true of itself' or 'heterological' for short. Tarski pursues this analogy in [Tarski, 1944, p. 371, fn. 11]].

We now have a sketch of the proof of Tarski's undefinability theorem for the specific case of the general theory of classes. Tarski states the result as follows:

> "*Theorem 1* (α) In whatever way the symbol 'Tr', denoting a class of expressions, is defined in the metatheory, it will be possible to derive from it the negation of one of the sentences which were described in the condition (α) of the convention T;

[59] For more on the diagonal argument, the Liar, and Tarski's theorem, see Simmons [Simmons, 1993, Chapter2].

> (β) assuming that the class of all provable sentences of the metalanguage is consistent, it is impossible to construct an adequate definition of truth in the sense of convention T on the basis of the metatheory." (p. 247)

Clearly, (β) follows immediately from (α). For a general statement of the theorem, recall Tarski's summary conclusion in the Postscript:

> "B. If the order of the metalanguage is at most equal to that of the language itself, such a definition cannot be constructed." (p. 273)

A typical contemporary formulation of Tarski's theorem is likely to dispense with any mention of a metalanguage; for example,

> Given a language L of a theory that includes first-order number theory, no adequate definition of truth in L can be given in L.

Of his proof of Theorem 1, Tarski writes "We owe the method used here to Gödel" (fn. 1, p. 247). Indeed, (α) does not really involve truth at all.[60] If we replace the symbol 'Tr' by the more neutral symbol 'ϕ', (α) tells us that there is a sentence p with a structural-descriptive name x such that the sentence '$\neg(\phi(x) \leftrightarrow p)$' and hence '$\phi(x) \leftrightarrow \neg p)$' is derivable from the metatheory. If we interpret ϕ not as 'true' but as 'provable', (α) is essentially an instance of Gödel's first incompleteness theorem. Originally, Tarski's paper did not contain Theorem 1, but instead "certain suppositions in the same direction" (fn. 1, p. 247), based partly on Tarski's own work and partly on a short report of Gödel's [Tarski, 1930b]. Theorem 1 and the sketch of its proof were added in press, after Tarski became acquainted with Gödel [Gödel, 1931] and was convinced that Gödel's results carried over to the general theory of classes.[61]

In notes added to [Tarski, 1933a], Tarski is at pains to set the historical record straight. There was some need for this. For example, Carnap erroneously wrote that Tarski's investigations were carried out "in connection with those of Gödel" (Carnap [Tarski, 1935]). In fact Tarski arrived at the final formulation of his definition of truth in 1929. But the paper was not published until 1933 (in Polish) — and only received more general attention when it was published in German in 1935. Tarski emphasizes that the methods and results of [Tarski, 1933a] were entirely his own, with two noted exceptions — Section 2 (on the problems of defining truth in natural language) was directly influenced by Lesniewski, and the method of the proof of Theorem 1 was drawn from Gödel [Gödel, 1931]. It is noteworthy that Tarski's discussion of the arithmetization of the metalanguage and metatheory proceeded quite independently of Gödel's work (though Gödel developed the method far more completely).

[60] This point is stressed by Vaught in [Tarski, 1986, p. 871].

[61] These negative results about truth and provability are informally discussed in Tarski [Tarski, 1969].

****

Tarski's theory of truth has received far more philosophical attention than anything else in Tarski's corpus. It is without doubt the work for which Tarski is best known to philosophers. It is also the subject of Tarski's most sustained philosophical discussion, which comprises the second half of [Tarski, 1944] (under the title of 'Polemical Remarks'). We turn now to some of the philosophical issues raised by Tarski's work on truth.

(i) There is disagreement over the *point* of Tarski's definition of truth. What is Tarski's aim? To this question, I believe there is a clear answer in Tarski's writings. On the one hand, truth is a notion that is fundamental to science, logic and metamathematics; on the other hand, it leads to paradoxes and antinomies, earning itself an "evil reputation". We must then find a way of defining truth that will preserve its intuitive content but place it beyond suspicion. Tarski seeks a definition of truth in terms that no one could question. And for Tarski, those terms are the terms of logic, the terms of the object language, and the structural-descriptive terms:

> "... the semantical concepts are ... reduced to purely logical concepts, the concepts of the language being investigated and the specific concepts of the morphology of language" [Tarski, 1936a, p. 406]

Moreover, this reduction is strongly suggested by the T-sentences. According to Tarski, the T-sentence

'snow is white' is true if and only if snow is white is a partial definition of truth — and it contains a logical term, a sentence of the object language, and a structural-descriptive name of that sentence. As we saw earlier, Tarski s†uggests that for an object language with finitely many sentences, we need only list all the associated T-sentences for a complete definition of truth. This will not work, of course, for languages with infinitely many sentences — for those, Tarski uses the recursive method. But still, says Tarski, his recursive definition effects the kind of reduction suggested by the T-sentences (see for example [Tarski, 1933a, pp. 251–3]).

Suppose now that in the metalanguage we introduce semantic concepts (referring to the object language) by way of Tarski's definition. Then:

> "the definition of truth, or of any other semantic concept, will fulfil what we intuitively expect from every definition; that is, it will explain the meaning of the term being defined in terms whose meaning appears to be completely clear and unequivocal. And, moreover, we have then a kind of guarantee that the use of semantic concepts will not involve us in any contradictions." [Tarski, 1944, pp. 674–5]

Here is the point of Tarski's definition. In the metalanguage, we may use the notion of truth (for the object language) without fear of paradox, safe in the knowledge that 'true' can always be eliminated via its *definiens* in a precise, uncontroversial way, in favour of terms which are not semantical and thereby immune to paradox.[62]

Tarski's definition, then, legitimates the use of the concept of truth (*for* a given object language, *in* a suitable metalanguage). And Tarski was optimistic that the concept of truth thus defined would prove fruitful, both for philosophy and the special sciences. Tarski points out that the problem of defining truth "has often been emphasized as one of the fundamental problems of the theory of knowledge" [Tarski, 1936a, p. 407]. He suggests that the notion of truth is essential to the following constraint on empirical theories:

> "As soon as we succeed in showing that an empirical theory contains (or implies) false sentences, it cannot be any longer considered acceptable." [Tarski, 1944, p. 691]

And in regard to mathematics and metamathematics, Tarski suggests that we can already see the valuable results contributed by the theory of truth and the semantic method:

> "These results concern the mutual relations between the notion of truth and that of provability; establish new properties of the latter notion (which, as well known, is one of the basic notions of metamathematics); and throw some light on the fundamental problems of consistency and completeness. ...
>
> Furthermore, by applying the method of semantics we can adequately define several important metamathematical notions which have been used so far only in an intuitive way - such as, e.g., the notion of definability or that of a model of an axiom system; and thus we can undertake a systematic study of these notions. In particular, the investigations on definability have already brought some interesting results, and promise even more in the future." (p. 693)

[62] These remarks are in broad agreement with Etchemendy's discussion in section 1 of Etchemendy [Etchemendy, 1988a]. They are opposed to Field's claim in Field [1986] that Tarski aimed for a physicalistic reduction of truth. Field rests his textual case on one brief passage from [Tarski, 1936a]:

> "... it would then be difficult to bring this method into harmony with the postulates of the unity of science and of physicalism (since the concepts of semantics would be neither logical nor physical concepts)." [Tarski, 1936a, p. 406]

The method in question is the axiomatic method of defining truth, according to which truth is introduced as a primitive concept whose basic properties are established by axioms. Tarski's concern here is that a primitive semantic concept would be in tension with the scientific, physicalist outlook. Tarski's theory produces no such tension: truth is eliminable, not primitive. Immediately after the passage just quoted, Tarski remarks that his preferred method reduces the concept of truth to logical concepts, concepts of the object language, and structural-descriptive concepts. There is no mention of physicalistic concepts or the need for a specifically physicalistic reduction. Tarski thought it an advantage of his theory that it harmonized with the scientific outlook — but there is no indication that Tarski sought an explicitly physicalist reduction of truth.

(ii) What kind of theory is Tarski's theory of truth? Is it, for example, a *correspondence* theory of truth? At several places, Tarski says that he wishes to capture the intuitions behind the classical Aristotelian conception of truth - and Tarski formulates that conception in terms of correspondence. Consider:

> "If we wished to adapt ourselves to modern philosophical teminology, we could perhaps express this conception by means of the following formula:
>
> The truth of a sentence consists in its agreement with (or correspondence to) reality." ([Tarski, 1944, p. 667]. See also [Tarski, 1933a, p. 153] and [Tarski, 1936a, p. 401].)

However, we would expect a full-blown correspondence theory to provide an account of the correspondence relation and its relata — and Tarski's theory does not. Such notions as *correspondence* and *state of affairs* are not the subject of Tarski's theory — for Tarski, these are vague notions that gesture towards a certain conception of truth. It is noteworthy that Aristotle's dictum (quoted above) makes no mention of a correspondence relation, or facts, or states of affairs. And Tarski finds Aristotle's formulation preferable to those that talk vaguely of 'correspondence' and 'reality' (see [Tarski, 1944, p. 667]). Still, Aristotle's formulation, like the others, cannot serve as a definition of truth, and

> "[i]t is up to us to look for a more precise expression of our intuitions." [Tarski, 1944, p. 667]

According to Tarski, those intuitions are adequately accommodated by the T-sentences. Take the sentence 'snow is white'. To what fact or state of affairs or condition does this sentence correspond? To the fact, state of affairs or condition that snow is white. That is, the sentence is true if snow is white, and false if snow is not white — or to put it in familiar terms, 'snow is white' is true if and only if snow is white. A definition of truth conforms to the correspondence conception if it implies such T-sentences. (See [Tarski, 1944, p. 667]). So we can attribute to Tarski a correspondence *conception* of truth — but we should not say that Tarski's *theory* is a correspondence theory in any robust sense.

If Tarski's theory is not a correspondence theory, is it a kind of *deflationary* theory?[63] Deflationists, in particular disquotationalists, have drawn heavily from Tarski's account. According to the disquotationalist (for example Quine [70], chapter 1), saying

> 'snow is white' is true

is just an indirect way of saying something about the world, namely, that snow is white. All there is to say about the truth of 'snow is white' is encapsulated by its associated T-sentence. According to the disquotationalist, truth is not a

[63] For a recent collection of classic and contemporary papers on deflationism see Blackburn and Simmons 1999.

substantive property; rather, the term 'true' is a "device for disquotation" (Quine [Quine, 1970, p. 12]). A natural disquotational definition of 'true' for a language L with finitely many sentences is this:

$$x \text{ is true iff } (x = x_1 \text{ and } p_1) \text{ or } (x = x_2 \text{ and } p_2) \text{ or } \ldots \text{or } x = x_n \text{ and } p_n$$

where 'x_1', 'x_2', ...'x_n' are replaced by quote-names of sentences of L, and 'p_1', 'p_2', ...'p_n' by the sentences themselves. And this definition looks very like the one Tarski suggests for languages with finitely many sentences (see above). So there appear to be some close connections between Tarski's theory and disquotationalism.

Of course, Tarski and disquotationalists are concerned with languages that contain infinitely many sentences. For that reason, as we saw, Tarski employs the recursive method. The disquotationalist might follow suit: given a language with a finite stock of names and predicates, reference may be disquotationally defined by a finite list of sentences of the form ' "a" refers to a', and satisfaction by a finite list of sentences of the form 'x satisfies "F" if and only if x is F'. But such a recursive disquotationalism is restricted to sentences that have the appropriate kind of logical form. And there is an array of truths that are notoriously hard to fit into the Tarskian mould: belief attributions, counterfactuals, modal assertions, statements of probability, and so on. Since disquotationalism is an account of the truth predicate of an entire natural language and not of some restricted portion, the recursive route seems unattractive.

The disquotationalist might prefer to extend the above finitary definition to an infinitary one:

$$x \text{ is true iff } (x = x_1 \text{ and } p_1) \text{ or } (x = x_2 \text{ and } p_2) \text{ or } \ldots.$$

where 'p_1', 'p_2' ... abbreviate the sentences of the language.[64] So now 'true' is not just a device of disquotation; it is also a device for expressing infinite disjunctions. Here the disquotationalist and Tarski part company; Tarski writes:

> "Whenever a language contains infinitely many sentences, the definitions constructed automatically according to the above scheme would have to consist of infinitely many words, and such sentences cannot be formulated either in the metalanguage or in any other language" [Tarski, 1933a, p. 188–9][65]

But Tarski's infinitary qualms aside, might it be claimed that disquotationalism captures the deflationary spirit of Tarski's theory?

[64]Such a definition is suggested by remarks in Leeds [1978], and versions of it are presented explicitly in Field [1986], Resnik [1990], David [1994].

For an illustration, consider the sentence: 'What Claire said yesterday was true'. According to the definition, the sentence is equivalent to What Claire said='aardvarks amble' and aardvarks amble or What Claire said = 'antelopes graze' and antelopes graze or ....

[65]Later Tarski investigated infinitary languages — see [Tarski, 1958a; Tarski, 1958b; Tarski, 1961b; Tarski, 1961c].

Such a claim might be further encouraged by the observation that Tarski's definition is eliminative. According to the disquotationalist we have just characterized, 'true' is always in principle eliminable by disquotation. The definiens contains, apart from logical terms, only the quote-names of the sentences of the given language, and their disquotations. And Tarski is quite clear that his definition of truth is eliminative (see [Tarski, 1944, p. 683]). Tarski points out that the elimination may not be as straightforward as it is with " 'snow is white' is true" — consider the sentences "All consequences of true sentences are true" and "The first sentence written by Plato is true", where the elimination of 'true' is not a matter of simple disquotation (see [Tarski, 1944, p. 83]). Nevertheless, elimination is always possible in principle:

> "Of course, since we have a definition for truth and since every definition enables us to replace the definition by its definiens, an elimination of the term 'true' in its semantic sense is always theoretically possible." [Tarski, 1944, p. 83]

However, despite the close connections between Tarski and the disquotationalist, and the shared eliminative nature of their definitions, I think Tarski would strongly resist the label 'deflationist'. First, unlike the disquotationalist, Tarski is not offering a general theory of truth in natural languages. His theory is a limited to certain regimented languages — including fragments of a natural language, but never the whole language. Second, for Tarski truth is a substantive and fruitful concept. According to the disquotationalist, there is no substantive concept of truth — only the term 'true' that serves as a logical device. But, according to Tarski, as we have seen, truth is a concept that has important work to do — in philosophy and in metamathematics, for example. Third, Tarski thinks it is a mistake to think that the concept of truth is sterile on the grounds that the word 'true' may be eliminated on the basis of its definition. That would lead, Tarski says, to the absurd conclusion that all defined notions are sterile. Tarski writes:

> "In fact, I am rather inclined to agree with those who maintain that the moments of greatest creative achievement in science frequently coincide with the introduction of new notions by means of definition." (p. 683)

Tarski reserves the label 'the semantic conception of truth' for his account. And the label is useful, at least for distinguishing Tarski's theory from correspondence and deflationary theories. Tarski's theory does not fit comfortably into either of those categories. Tarski himself emphasizes the neutrality of his theory:

> "we may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naive realists, critical realists or idealists, empiricists or metaphysicians — whatever we were before. The semantic conception is completely neutral toward all these issues" [Tarski, 1944, p. 686]

Epistemological and metaphysical disputes about truth are not Tarski's concern. Rather, Tarski is after a definition of 'true sentence' that was precise, fruitful, immune to paradox, and consonant with ordinary usage.

(iii) There are many other philosophical issues raised by Tarski's work on truth. Here we indicate just some of them.

(a) Is truth immanent or transcendent? It is striking that Tarski defines 'true' only relative to a given language:

> "We can only meaningfully say of an expression that it is true or not if we treat this expression as a part of a concrete language. As soon as the discussion concerns more than one language the expression 'true sentence' ceases to be unambiguous." [Tarski, 1933a, p. 263]

The question arises whether truth must always be 'immanent' in this way, or whether there is a more general notion that transcends particular languages.[66]

(b) Are natural languages universal in the way that Tarski says they are? At issue here is the expressive capacity of natural languages.[67]

(c) Are natural languages inconsistent? Tarski is often taken to give an affirmative answer to this question, and he certainly comes close. He says that the semantical antinomies

> "seem to provide a proof that every language which is universal in the above sense, and for which the normal laws of logic hold, must be inconsistent." [Tarski, 1933a, pp. 164–5]

But note the "seem to". In [Tarski, 1944], Tarski is more explicitly cautious — since a natural language has no exactly specified structure, "the problem of consistency has no exact meaning with respect to this language" [Tarski, 1944, p. 673]. Tarski also notes in [Tarski, 1933a, p. 267] that bringing exact methods to bear on natural language would require a reform of the language, including its division into a sequence of object languages and metalanguages.

> "It may, however, be doubted whether the language of everyday life, after being 'rationalized' in this way, would still preserve its naturalness and whether it would not rather take on the characteristic features of the formalized languages." [Tarski, 1933a, p. 267]

There is, then, a delicate question as to how to read Tarski here. My preferred reading is this: it is inappropriate to bring exact methods to bear on natural language, and if one insists on doing so, the result will either be confusions and contradictions or an artificial regimentation of natural language.[68]

(d) How is Tarski's definition of truth related to the truth-conditions theory of meaning? According to Davidson [Davidson, 1967], the relationship is very close:

[66] For discussions of this issue, see for example Resnik [1990] and Field [1994].
[67] For further discussion see, for example, Martin [1976], Herzberger [1981] and Simmons [1993].
[68] For further discussion see, for example, Soames [1999, pp. 49–56 and 150–51].

> "There is no need to suppress, of course, the obvious connection between a definition of truth of the kind Tarski has shown how to construct, and the concept of meaning. It is this: the definition works by giving necessary and sufficient conditions of the truth of every sentence, and to give truth conditions is a way of giving the meaning of a sentence. To know the semantic concept of truth for a language is to know what it is for a sentence — any sentence — to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language." (p. 76)

Others would deny the connection that Davidson maintains here.[69] It is clear that Tarski would not regard his definition of truth as the basis for a theory of meaning for natural languages. For one thing, Tarski restricts himself to formal languages. And for another, truth, not meaning, is Tarski's target. The meanings of the sentences of the object language and metalanguage are taken as given from the outset — indeed, it is a requirement of Tarski's definition that the object language be translatable into the metalanguage.

4.3 Logical consequence

We have just seen how Tarski took up the challenge of providing a precise definition of truth that conforms to ordinary usage. In his paper "On the concept of logical consequence" [Tarski, 1936b], Tarski takes up the parallel challenge for the concept of *logical consequence*. At the outset of the paper, Tarski rejects the idea that logical consequence can be captured exhaustively via formal rules of inference. The idea had great appeal among logicians: after all, all exact reasonings in mathematics were formalizable using a small stock of simple rules of inference. But Tarski showed that there is a clear and unbridgeable gap between the ordinary notion of logical consequence and the logicians' formalized notion.

As early as 1927, Tarski showed the existence of so-called *ω-incomplete theories* (the terminology, but not the concept, is due to Gödel [Gödel, 1931]).[70] Such a theory includes the normal rules of inference (such as substitution and detachment), and among its theorems are the following:

A_0. 0 possesses the property P

A_1. 1 possesses the property P

and in general every sentence of the form

[69]See for example Etchemendy [1988a, pp. 56–62], and Soames [1999, pp. 102–107]. Davidson's own position on this issue has evolved over the years– - for example, one may compare and contrast Davidson [1967] and Davidson [1996].

[70]Tarski reports (in [Tarski, 1933b, p. 279, fn. 2] that in 1927 he had delivered a paper entitled 'Remarks on some notions of the methodology of the deductive sciences' to the Second Conference of the Polish Philosophical Society in Warsaw in which he stressed the importance of the concepts of ω-consistency and ω-completeness, and communicated an example of an ω-incomplete theory. A slightly altered form of this theory appears in [Tarski, 1933b].

A_n. n possesses the property P.

But the following sentence cannot be proved in the theory:

A. Every natural number possesses the property P.

Now it clear that, according to our ordinary notion, A *is* a logical consequence of $A_0, A_1, \ldots A_n \ldots$. So:

> "the formalized concept of consequence, as it is generally used by mathematical logicians, by no means coincides with the common concept." (p. 411)

Tarski goes on to point out that this gap cannot be closed by the introduction of new (infinitary) rules of inference. For Gödel's results show that

> "[i]n every deductive theory (apart from theories of a particularly elementary nature), however much we supplement the ordinary rules of inference by new purely structural rules, it is possible to construct sentences which follow, in the usual sense, from the theorems of this theory, but which nevertheless cannot be proved in this theory on the basis of the accepted rules of inference." [Tarski, 1936b, p. 413]

Consequently,

> "In order to obtain the proper concept of consequence, which is close in essentials to the common concept, we must resort to quite different methods and apply quite different conceptual apparatus in defining it." (*ibid.*)

The new methods are Tarski's semantic methods. While Tarski makes "no very high claim to complete originality" [Tarski, 1935, p. 414], it is clear that he regards his account as the first fully precise semantic treatment of logical consequence. Tarski's positive account in [Tarski, 1936b] starts out with two guiding intuitions. Suppose we have a class K of sentences and a sentence X which follows from K. The intuitions are these:

(i) It can never happen that every member of K is true and X is false.

(ii) The subject matter of K and X (the objects to which reference is made) have no effect on the consequence relation — only the *form* of the sentences matters.

(i) and (ii) suggest the following condition on the consequence relation:

(F) Given K and X, if the nonlogical constants are replaced uniformly, yielding K' and X', then X' must be true if all the sentences of K' are true.

Now Tarski points out that while (F) can serve as necessary condition, it is not sufficient. The difficulty is this: we may not assume that the language contains designations for all possible objects.[71] (Here, Tarski's remarks are aimed at Carnap's definition of logical consequence.)[72] To avoid this difficulty, Tarski turns to the notions of sentential function and satisfaction. Assume a correspondence between variables and extralogical constants. Given a class L of sentences, replace each occurrence of an extralogical constant by its corresponding variable. We will obtain a class of sentential functions — call it L'. Given the concept of the satisfaction of a sentential function by a sequence of objects, we can define the notion of a *model* of L as follows:

> An arbitrary sequence S of objects is a *model* of the class L of sentences if and only if S satisfies every member of L'.

The notion of a *model of a sentence* is to be understood in the the obvious way. And now Tarski states the definition of logical consequence:

> "The sentence X *follows logically* from the sentences of the class K if and only if every model of the class K is also a model of the sentence X." [Tarski, 1936b, p. 417]

This definition, says Tarski, agrees with common usage, and meets the conditions (i) and (ii) — for it follows from the definition that every consequence of true sentences is true, and that the consequence relation is independent of the meanings of extra-logical constants. Notice that a definition of *logical truth* is immediately forthcoming: a sentence Y is logically true if every interpretation of Y is a model of Y.

There has been far less philosophical discussion of Tarski's notion of logical consequence than of his notion of truth. (Perhaps, as Gómez-Torrente suggests in [Gómez-Torrente, 1996, p. 126], this is a sign of its success — it is generally regarded as a standard fixture of modern logic.) But there has been some discussion, most notably the critical work of Etchemendy in [1983; 1988a; 1988b; 1990]. Etchemendy challenges the model-theoretic treatment of logical consequence, a treatment that is now quite standard.[73] Etchemendy also challenges the widespread attribution of this standard treatment to Tarski (see [Etchemendy, 1988a]). He argues that Tarski's definition of logical consequence diverges from the standard one.

---

[71] Tarski points out that it may, and does, happen that X does not follow in the ordinary sense from the sentences of K, and yet the condition F is satisfied — where F is satisfied because the language with which we are dealing does not have a sufficient stock of extra-logical symbols. Tarski writes that the condition is sufficient

> "only if the designations of all possible objects occurred in the language in question. This assumption, however, is fictitious and can never be realized" [Tarski, 1936b, p. 416]

[72] In Carnap [1934, pp. 137ff], Carnap provided definitions of logical consequence and further derivative concepts. Tarski regarded them as inadequate because the defined concepts are dependent on the richness of the language under investigation.

[73] For a defence of the model-theoretic account against Etchemendy's attack, see Soames [1999, pp. 117–136].

According to Etchemendy, there are two main points of divergence. The first concerns the ω-rule of inference, the inference that allows us to move from $A_1, A_2, \ldots A_n, \ldots$ to A. Given a first-order formalization of arithmetic (where the numerals are nonlogical constants) the ω-rule is invalid according to the standard model-theoretic account — for there are interpretations of 'natural number' and '0', '1', '2', … for which $A_1, A_2, \ldots A_n, \ldots$ are all true and A is false. But Tarski claims that the inference is intuitively valid, and that his definition respects this intuition. If so, then it follows that Tarski's definition and the standard model theoretic definitions are different. And Etchemendy goes on to argue that the price Tarski pays for the validation of the ω-rule is a trivialization of his analysis.[74]

The second main point of divergence concerns domain variation: according to Etchemendy, Tarski's account of logical consequence, unlike the standard account, does not require that we vary the domain of quantification. But now Tarski's account gets into trouble, Etchemendy argues. Suppose the intended interpretation of a given language has a domain with two or more individuals. Now consider the sentence $\exists x \exists y(x \neq y)$. If we do not require domain variation, this sentence will be a logical truth.

If Etchemendy is right, Tarski's definition — to its detriment — diverges from the standard model-theoretic account in two important respects. But it is possible to resist Etchemendy's reading of Tarski. According to Gómez-Torrente [Gómez-Torrente, 1996], the supposed divergences are illusory. First, Tarski did not think that first-order versions of the ω-rule are valid — only certain versions in broader logical frameworks are valid. (Of course, an argument may be valid under some but not all of its formalizations). Gómez-Torrente argues that in [Tarski, 1936b] Tarski has in mind a presentation of arithmetic in a broad logical framework such as the simple theory of types or the calculus of classes. (We saw above in fn. 71 that Tarski discusses ω-incompleteness in [Tarski, 1933b], where the logical framework is the general theory of classes.) In such a broad framework, arithmetical expressions (like the numerals) are defined expressions, not primitives or nonlogical constants. Now consider the ω-rule of inference again. If there are any nonlogical constants, they will appear in P only — nowhere else are there any constants subject to reinterpretation. And since the class of logically defined natural numbers will coincide with the extension of the predicate 'natural number', A will follow from $A_1, A_2, \ldots A_n, \ldots$ (to borrow Etchemendy's words, quoted in fn. 74, any set that contains each natural number contains every natural number).

Second, although it is true that [Tarski, 1936b] makes no mention of domain variation, Gómez-Torrente provides several reasons for thinking that it is implicit

[74] According to Etchemendy, Tarski allows a nonstandard choice of logical constants, which may include the numerals and the quantifier 'every natural number'. Then A will be a consequence of $A_1, A_2, \ldots A_n, \ldots$, since "any set that contains each natural number contains every natural number" [Etchemendy, 1988a, p. 73]. Etchemendy argues that this flexibility with regard to the choice of logical constants leads to a certain trivialization of Tarski's analysis. However, below we will consider another reading of Tarski, according to which Tarski's definition does not diverge from the standard model-theoretic one, Tarski does not endorse nonstandard choices of logical constants, and no such trivialization results.

in Tarski' definition. Perhaps the most compelling reason is the following. Tarski often specifies a domain (or universe of discourse) by way of a nonlogical predicate. For example, recall the 'textbook' introduction of a model in [Tarski, 1941a, Chapter VI] — detailed in 2.5 above. We start with a simple deductive theory, the theory of congruence. The nonlogical predicate 'S' denotes the set of all line segments — the domain or universe of discourse. Now the two axioms are *relativized to* S; for example, Axiom 1 says that for all $x \in S$, $x \cong x$. Now we abstract away from our particular theory. In particular, we replace 'S' by 'K', so that now the axioms are fully abstract, and relativized to 'K'. Now 'K' may be interpreted as the set of all line segments — but it may also be interpreted as the universal class, or the set of all numbers, or any set of numbers, and so on. In each case, the interpretation of 'K' will be the domain of the model. Here, and elsewhere in Tarski's work,[75] domain variation is accommodated by the reinterpretation of nonlogical predicates. But Tarski's test for logical consequence just is a matter of interpretation and reinterpretation — and so Tarski's definition as it stands accommodates domain variation.[76]

We have good reason, then, to maintain the received view: Tarsi provided us with the first precise model-theoretic treatment of logical consequence — and Tarski's treatment remains standard to this day.

4.4 Model theory: some historical remarks

Tarski's seminal work on definability, truth and logical consequence were of central importance to the development of *model theory* — the study of the relation between formal languages and their interpretations. Chang and Keisler point out that model theory is a young subject, "not clearly visible as a separate area of research until the early 1950s" [Chang and Keisler, 1973, p3]. Lowenheim [1915] is generally regarded as the first specific contribution to model theory. In this paper, Lowenheim introduces *the method of elimination of quantifiers* (more on this

[75] There are many other places where Tarski specifies a domain via a nonlogical predicate. For another example, consider Tarski's presentation of abstract Boolean algebra in [Tarski, 1935; **?**] — detailed in 2.4 above. As we saw, the meaning of the predicate 'B' is given by 'the universe of discourse'. In the course of [Tarski, 1935; **?**], 'B' is variously interpreted as the universe of sets (Boolean set algebra), the set S of all sentences (the calculus of sentences), and the class D of deductive systems (the calculus of systems). Again, domain variation is accommodated by the reinterpretation of a nonlogical predicate.

As further examples, Gomez-Torrente also cites all the first order theories, and the second-order theory of real arithmetic, presented in [Tarski, 1936c].

[76] Gomez-Torrente mentions two other reasons for thinking that domain-variation is implicit in Tarski's definition. First, as both Etchemendy and Gomez-Torrente point out, Tarski had already defined the notion of 'true sentence in an individual domain a' in [Tarski, 1933a], showing an explicit concern with domain variation prior to [Tarski, 1936b]. Why unnecessarily attribute to Tarski an abrupt change of mind? Second, [Tarski, 1936b] is a summary of an address given to a philosophical audience (it was delivered at the International Congress of Scientific Philosophy, in Paris, 1935). The paper is indeed informally written and brief — so it would not be surprising if Tarski suppressed certain more technical matters. (Gomez-Torrente [1996] is a careful and convincing paper which contains more argumentation and historical information that I am able to cover here.)

below). Lowenheim also proves the theorem that bears his name: If a first-order sentence A is true in every finite domain but not every domain, then there is a denumerable domain in which A is not true. Lowenheim's result was improved by Skolem in [**?**], yielding the Lowenheim-Skolem theorem: Any set of first-order sentences which has a model has a denumerable model. In [Skolem, 1922], Skolem went on to improve the result further: Any model of a set of first-order sentences has a denumerable submodel.[77]

This citation is for Skolem 1920.

Tarski's earliest contributions to model theory were presented in the university lectures and seminars that he gave at Warsaw University in the years 1926–1928 (see [Tarski, 1948a, p. 50, fn. 11]). It is reported that in 1927-8 Tarski produced further improvements to the Lowenheim-Skolem theorem as follows:

(i) Any denumerable set of first-order sentences that has an infinite model has an uncountable model.[78]

Please check reference in footnote

(ii) Any denumerable set of first-order sentences that has an infinite model has a model in each infinite power.[79]

Vaught reports that Tarski never published a proof of (i) or (ii); Tarski's proofs remains a mystery.[80] Tarski also explored the method of eliminating quantifiers, a powerful method for establishing metamathematical properties of theories, especially decidability. We will take a closer look at the method in Section 5 below. A number of key notions of model theory were introduced in the seminar — among them, Vaught reports, the notion of *elementary equivalence*.[81]

Gödel's landmark completeness theorem was published in 1930. During the 1930s, Tarski produced his greatest contributions to model theory — the work on definability, truth and logical consequence. As we have seen, Tarski stresses semantical methods in Chapter VI of his 1941 textbook [Tarski, 1941a], and the notion of model takes center-stage.

It was not until the postwar period that model theory became an autonomous area of study. In 1954–5, Tarski published a series of three articles under the title "Contributions to the theory of models" [**?**; **?**; **?**]. In the first of these, Tarski writes:

[77]Given a model M with domain D, a submodel M' of M has for its domain a subset of D. M' assigns to the expressions of the given first-order language restrictions of the extensions that M assigns — restrictions tailored to M''s smaller domain.

[78]According to the editors of Fundamenta Mathematicae, Tarski proved this result in 1927–8. See the editors' Note after the paper Skolem [**?**].

[79]In [Tarski, 1957b], Tarski and Vaught write:

> "A proof, along these lines, of Theorem 2.2 is known, but is by no means simple. It is essentially the same proof which was originally found by Tarski, in 1928, for the generalized Lowenheim-Skolem theorem." [Tarski, 1957b, p. 666, fn. 8]

[80]Vaught refers the reader to his speculations in Vaught [Vaught, 1965, p. 398], and [Vaught, 1974, p. 160]. The first published proof of (ii) appeared in Mal'cev [36] (which also contains a generalization of Godel's completeness and compactness theorems to uncountable languages).

[81]Vaught [1986, p. 870]. Two models are elementarily equivalent iff every sentence that is true in one is true in the other, and vice versa.

> "Within the last few years a new branch of metamathematics has been developing. It is called the *theory of models* and can be regarded as a part of the semantics of formalized theories." [**?**, p. 517]

This appears to be the first time that the phrase 'theory of models' appears in the literature.[82] The subject developed quickly in the 1950s, stimulated by work of Tarski, Malcev, Henkin and Abraham Robinson, among others. Tarski continued to make important contributions: for example, he introduced the notions of *elementary substructures* and *elementary chains*, and proved some fundamental results involving these notions, including "Tarski's union theorem" (see [Tarski, 1957b], with Vaught); with Frayne, Morel and Scott, he showed the importance of the role of ultraproducts in model theory (see Frayne *et al.* [**?**]); and with Hanf, he initiated work on measurable cardinals (Tarski [Tarski, 1962] and Hanf [Hanf, 1963–4] — and see Section 7 below).[83] Tarski also influenced model theory a great deal through his students and through his collaborations with other logicians.[84]

These references are to Tarski, 1954a, 1954b and 1955, none of which appear in the bibliography.

No reference for Frayne et al. in bibliography

5 DECIDABILITY AND UNDECIDABILITY

Tarski casts the decision problem sometimes in terms of proof, and sometimes in terms of truth. In [Tarski, 1953], for example, the notion of proof is central:

> "By a *decision procedure* for a given formalized theory T we understand a method which permits us to decide in each particular case whether a given sentence formulated in the symbolism of T can be proved by means of the devices available in T (or, more generally, can be recognized as valid in T). The *decision problem* for T is the problem of determining whether a decision procedure for T exists (and possibly of exhibiting such a procedure). A theory is called *decidable* or *undecidable* according as the solution of the decision problem is positive or negative." [Tarski, 1953, p. 3]

In [Tarski, 1948a], the notion of truth takes centre-stage:

[82] This claim is made by Chang and Keisler [1973, p. 3], and by Vaught [1986, p. 876].

[83] For a more detailed survey of Tarski's postwar contributions to model theory, see Vaught [1986] and Chang and Keisler [1973, pp. 515–531]. Addison, Henkin and Tarski [1965b] contains a large bibliography of work on model theory up to that time.

[84] Consider the roster of logicians that Vaught provides:

> "Tarski influenced model theory not only with his papers but also through his PhD students, his correspondence, and his conversations with people, many of whom came to Berkeley to see him. This was especially so in the two decades after the war, His PhD students active in model theory included Mostowski, Julia Robinson, Wanda Szmielew, [Robert Vaught], C. C. Chang, S. Feferman, Montague, H. J. Keisler, H. Gaifman, W. Hanf, and others. Some people (not PhD students) active in model theory, who were close to Tarski and received inspiration from him in their work, were Beth, Fraisse, Henkin, Los, Lyndon, and D. Scott. One could add J. Ax, W. Craig, A. Ehrenfeucht, Y. Ershov, S. Kochen, M. Rabin, A. Robinson, and many, many others." [Vaught, 1986, p. 877]

> "The most important kind of decision problems is that in which K [a class of sentences] is defined to be the class of true sentences of a certain theory. When we say that there is a decision method for a certain theory, we mean that there is a decision method for the class of true sentences of the theory" [Tarski, 1948a, p. 1]

Tarski was in no doubt about the significance of the decision problem:

> "As is well known, the decision problem is one of the central problems of contemporary metamathematics". [Tarski, 1953, p. 3]

In the present section, we turn to Tarski's work on decidability (subsections 1 and 2) and undecidability (subsection 3).

5.1 Decidability of the theory of the reals

Tarski famously proved the decidability of the first-order theory of the real numbers. The proof was first published in full detail in *A Decision Method for Elementary Algebra and Geometry* [Tarski, 1948a]:

> "In this monograph we present a method ... for deciding on the truth of sentences of the elementary algebra of real numbers." (p. 2)

By 'the elementary algebra of real numbers" is to be understood

> "that part of the general theory of the real numbers in which one uses exclusively variables representing real numbers, constants denoting elementary operations on and relations between real numbers, like '+', '.', '-', '<', '>', and '=', and expressions of elementary logic such as 'and', 'or', 'not', 'for some x', and 'for all x'." (p. 2)

The primitive operations are '+' and '.', and so we shall refer to this theory as $\langle R, +, . \rangle$.

Here decidability is understood in the semantic sense — the intuitive notion of truth plays a fundamental role in Tarski's presentation. (Tarski refers the reader to the formal definition in [Tarski, 1933a].) But Tarski points out that truth could be eliminated from the entire discussion by proceeding axiomatically, and replacing the notion of truth by the notion of provability (see footnote 9, pp. 48–50, where Tarski provides a list of axioms). Under this new interpretation, Tarski's results lead easily to the conclusion that the axiomatic system of elementary algebra is decidable, consistent, and complete (see fn. 15, p. 53). Moreover, since the axioms are satisfied not only by the reals but also by the elements of any real-closed field, it follows that the theory of real-closed fields is also decidable, consistent, and complete (see fn. 15, p. 54).[85] Of these various related results, the decision method for real-closed fields is particularly celebrated.

[85] For a full characterization of the theory of real-closed fields see, for example, Chang and Keisler [Chang and Keisler, 1973, p. 41].

Tarski established the decidability of the first-order theory of the reals by "the method of eliminating quantifiers" [Tarski, 1948a, p. 15]. Following Tarski [Tarski, 1948a, p. 50, fn. 11], we can trace the history of this method through Lowenheim [1915, Section 3], Skolem [1919, Section 4], Langford [?], and Presburger [?]. Tarski mentions that in his university lectures for the years 1926–8 the method was developed in a general and systematic way (see [Tarski, 1948a, p. 50, fn. 11].

No entry in bib for Langford 1910

In outline, the method is as follows. Given a theory T, we identify certain *basic formulas*, and prove that every formula of T is T-equivalent to a Boolean combination of basic formulas. A *Boolean combination* of basic formulas is a formula obtained from basic formulas by repeated applications of $\neg$ and &; and two formulas ϕ and ψ are *T-equivalent* iff $T|\phi \leftrightarrow \psi$ (i.e. $\phi \leftrightarrow \psi$ is a semantic consequence of T). The proof — and in particular the step where we 'eliminate quantifiers' — yields a decision procedure for the theory T.

For illustration, consider the theory D of dense simple order without endpoints. We take D to be given by the set of its axioms,[86] and we take the true sentences of D to be its axioms and their semantic consequences. We will apply the method of quantifier elimination to show that D is decidable (a result published by Langford in 1927).[87] We take as basic formulas the following: '$v_m = v_n$' and '$v_m \leq v_n$'. (Notice that the Boolean combinations of basic formulas are exactly the open formulas.) Let '$v_m < v_n$' abbreviate '$v_m < v_n \& v_m \neq v_n$'. We consider $n+1$ variables $v_0, v_1, \ldots, v_n$ for $n > 0$. We associate with these variables certain conjunctions $\theta_0 \& \theta_1 \& \ldots \theta_{n-1}$ called *arrangements.* Each conjunct θ_i is either $u_i < u_{i+1}$ or $u_i = u_{i+1}$, where $u_0, u_1, \ldots, u_n$ is a renumbering of $v_0, v_1, \ldots, v_n$.

The main result to be proved is the following:

Quantifier Elimination Theorem for D Every formula is D-equivalent to an open formula.

[86] The theory of dense simple order without endpoints has the following axioms:

1. $\forall x \forall y \forall z(x = y \& y = z \rightarrow x = z)$ (transitivity)
2. $\forall x \forall y(x = y \& y = x \rightarrow x = y)$ (antisymmetry)
3. $\forall x(x = x)$ (reflexivity)
4. $\forall x \forall y(x = y \vee y = x)$ (comparability)
5. $\forall x \forall y(x = y \& x \neq y) \rightarrow \exists z(x = z \& z \neq x \& z = y \& z \neq y)$
6. $\exists x \exists y(x \neq y)$
7. $\forall x \exists y(x = y \& x \neq y)$
8. $\forall x \exists y(y = x \& x \neq y)$.

(Axioms 1,2,3 give the theory of *partial orders*; if we add Axiom 4, we obtain the theory of *simple order* (or *linear order*); if we further add Axioms 5 and 6, we obtain the theory of *dense simple order*; and with the final addition of Axioms 7 and 8, we arrive at the theory of dense simple order without endpoints.)

[87] In his papers [1927; 1928], Langford proved that the following theories are decidable by the method of quantifier elimination: dense linear orders without endpoints, dense linear orders with a first element but no last, dense linear orders with first and last elements, and the system of the natural numbers ordered by '<'.

We can easily show that to prove this theorem it is sufficient to prove:

(1) If $\psi(v_0, \dots v_n)$ is an open formula, then $\exists v_m \psi$ is D-equivalent to an open formula.[88]

As we are about to see, the proof of (1) involves the elimination of quantifiers, and yields a decision procedure for the theory D.[89] First, some preliminaries. Note that if $m > n$, then the quantifier is vacuous, and (1) is immediate. So we may assume $m \leq n$. It is convenient to further assume a suitable renaming of the variables so that $m = n$. We will also assume the following lemma:

LEMMA 8. *Every open formula $\psi(v_0, \dots, v_n)$ is D-equivalent to $v_0 < v_0$, or $v_0 = v_0$, or the disjunction of finitely many arrangements of the variables $v_0, \dots, v_n$.*[90]

So for the purposes of proving (1), we can assume that ψ is

(i) $v_0 < v_0$, or

(ii) $v_0 = v_0$, or

(iii) the disjunction of finitely many arrangements of the variables $v_0, \dots, v_n$.

(1) is immediate for the cases (i) and (ii), since in each case $\exists v_n \psi$ (either $\exists v_0(v_0 < v_0)$ or $\exists v_0(v_0 = v_0)$) is D-equivalent to the open formula ψ (either $v_0 < v_0$ or $v_0 = v_0$ respectively).

---

[88] Here is a sketch of a proof of this result, that to prove the theorem it is sufficient to prove (1). First we establish a general result for any theory T (this is Lemma 1.5.1 in [Chang and Keisler, 1973, p. 50]]:

> To show that every formula is T-equivalent to a Boolean combination of basic formulas, it is sufficient to show:
>
> (i) every atomic formula is T-equivalent to a Boolean combination of basic formulas, and
>
> (ii) if ψ is a Boolean combination of basic formulas then $\exists v_m \psi$ is equivalent to a Boolean combination of basic formulas.

This general result is proved by induction on the complexity of formulas. If ϕ is an atomic formula, then ϕ is T-equivalent to a Boolean combination of basic formulas by (i). The cases for negation and conjunction are obvious. If ϕ is $\exists v_m \psi$, where ψ is a Boolean combination θ of basic formulas, then $\exists v_m \psi$ is T-equivalent to $\exists v_m \theta$, and so, by (ii), T-equivalent to a Boolean combination of basic formulas. (For more details, see Chang and Keisler, *ibid.*).

Now apply this general result to theory D. Here, (i) is immediate: the atomic formulas are $v_m = v_n$ and $v_m = v_n$, and these are the basic formulas (and so vacuously equivalent to a Boolean combination of basic formulas). So it remains only to prove (ii). Since the open (or quantifier-free) formulas are exactly the Boolean combinations of basic formulas, (ii) is equivalent to:

If ψ is an open formula, then $\exists v_m \psi$ is D-equivalent to an open formula.

But this is just (1) — and so it remains only to prove (1). So to prove the quantifier elimination theorem it is sufficient to prove (1) — which is what we wanted to show.

[89] For the proof that follows, I follow Chang and Keisler [Chang and Keisler, 1973, pp. 52–4]]. Full details can be found there.

[90] This is Lemma 1.5.2, [Chang and Keisler, 1973, p. 51]. For the proof, see pp. 51–2.

So we turn to case (iii), and let

$$\psi = \theta_0 \ldots v\theta_p$$

where each θ_i is an arrangement of $v_0, \ldots, v_n$. It follows that

$$D \vDash \exists v_n \psi \leftrightarrow \exists v_n(\theta_0 v \ldots v\theta_p),$$

and so

$$D \vDash \exists v_n \psi \leftrightarrow \exists v_n \theta_0 v \ldots v \exists v_n \theta_p.$$

We now show that $\exists v_n \psi$ is D-equivalent to an open formula. First, suppose that $n = 1$. Then there are only three possibilities for each disjunct $\exists v_n \theta_i : \exists v_1(v_0 < v_1)$ or $\exists v_1(v_0 = v_1)$ or $\exists v_1(v_1 < v_0)$. Each of these is a consequence of D, and so the entire disjunction is a consequence of D. That is:

$$D \vDash \exists v_n \theta_0 v \ldots v \quad v_n \theta_p.$$

It follows that

$$D \vDash \exists v_n \psi.$$

So $\exists v_n \psi$ is D-equivalent to $v_0 = v_0$, an open formula.

Second, suppose that $n > 1$. We now describe a procedure for finding an open formula D-equivalent to $\exists v_n \psi$. (We will later see that this procedure is the core of the decision procedure for D.) Consider each arrangement θ_i of $v_0, \ldots, v_n$. There is an associated arrangement $\theta_i{}^*$ of $v_0, \ldots, v_{n-1}$ formed from θ_i by leaving out v_n.[91] It is straightforward to check that

$$D \vDash \exists v_n \theta_i \leftrightarrow \theta_i * .$$[92]

So it follows that

$$D \vDash \exists v_n \psi \leftrightarrow \theta_0 * v \ldots v\theta_p * .$$

[91] For example, suppose $n = 2$ and the arrangement θ_i of v_0, v_1 and v_2 is:

$$v_0 < v_1 \& v_1 = v_2.$$

Then the associated arrangement $\theta_i *$ of v_0 and v_1 is:

$$v_0 < v_1.$$

[92] For an illustration, consider θ_i and θ_{i*} from the previous footnote. Here the biconditional $\exists v_n \theta_i \leftrightarrow \theta_i^*$ is

$$\exists v_2(v_0 < v_1 \& v_1 = v_2) \leftrightarrow v_0 < v_1.$$

which is logically equivalent to

$$v_0 < v_1 \& \exists v_2(v_1 = v_2) \leftrightarrow v_0 < v_1.$$

This is clearly a semantic consequence of D as required (since $Q\exists v_2(v_1 = v_2)$ is an obvious semantic consequence of D).

So $\exists v_n\psi$ is D-equivalent to an open formula. This completes the proof of (1), and so we have proved the quantifier elimination theorem for D.

We are also in a position to describe a decision procedure for D. Given an arbitrary sentence ϕ, we wish to determine whether ϕ is a true sentence of D. We proceed as follows. We put ϕ into prenex normal form $\mathrm{Q}_0\mathrm{v}_0\ldots\mathrm{Q}_n\mathrm{v}_n\psi$, where each Q_i is a quantifier expression, and ψ is an open formula. We can assume without loss of generality that Q_n is an existential quantifier, since otherwise we may start with $\neg\phi$. We now eliminate the existential quantifier Q_n. Since ψ is an open formula, by the above lemma it is D-equivalent to $\mathrm{v}_0 < \mathrm{v}_0$ or $\mathrm{v}_0 = \mathrm{v}_0$ or a disjunction $\theta_0\mathrm{v}\ldots\mathrm{v}\ \theta_p$ of finitely many arrangements of $\mathrm{v}_0,\ldots,\mathrm{v}_n$. In the first case, we replace $\mathrm{Q}_n\mathrm{v}_n\psi$ by $\mathrm{v}_0 < \mathrm{v}_0$; in the second case we replace $\mathrm{Q}_n\mathrm{v}_n\psi$ by $\mathrm{v}_0=\mathrm{v}_0$. In the third case, we replace $\mathrm{Q}_n\mathrm{v}_n\psi$ by $\mathrm{v}_0=\mathrm{v}_0$ if n=1, and if n>1, we replace $\mathrm{Q}_n\mathrm{v}_n\psi$ by a D-equivalent open formula $\theta_0{}^*$ v ... v $\theta_p{}^*$ by the procedure just described in the previous paragraph. We repeat this procedure for Q_{n-1}, Q_{n-2}, $\ldots\mathrm{Q}_1$, until a sentence of the form $\mathrm{Q}_0\chi(\mathrm{v}_0)$ remains. And then it is obvious whether or not $\mathrm{Q}_0\chi(\mathrm{v}_0)$ is a true sentence of D. (It is easy to see that $\mathrm{Q}_0\chi(\mathrm{v}_0)$ must take one of four forms: $\exists v_0(v_0 = v_0), \forall v_0(v_0 = v_0), \exists v_0(v_0 < v_0)$ or $\forall v_0(v_0 < v_0)$, the first two of which are clearly true sentences of D, while the second two clearly are not.) From this we can decide whether or not ϕ is a true sentence of D.

For a simple example, consider the sentence

(i) $\exists v_0\forall v_1 v_2(v_0 < v_1 \& v_1 < v_2)$.

Notice that this sentence is already in prenex normal form, and that the formula '$v_0 < v_1 \& v_1 < v_2$' is already an arrangement of the variables v_0, v_1, v_2. Following our procedure, we replace the formula '$\exists v_2(v_0 < v_1 \& v_1 < v_2)$' by '$v_0 < v_1$', and obtain the sentence:

(ii) $\exists v_0\forall v_1(v_0 < v_1)$.

Since the innermost quantifier expression of (ii) is a universal quantifier, we work instead with its negation, which is logically equivalent to

(iii) $\forall v_0\exists v_1\neg(v_0 < v_1)$.

Since here $n = 1$, our procedure tells us to replace '$\exists v_1\neg(v_0 < v_1)$' by '$v_0 = v_0$' to obtain the sentence

(iv) $\forall v_0(v_0 = v_0)$.[93]

[93]Notice how this step connects to the case of $n = 1$ in the proof above. The formula '$\neg(v_0 < v_1)$' is D-equivalent to the following disjunction of arrangements of v_0, v_1:

$$v_0 = v_1 \vee v_1 < v_0.$$

Now $\exists v_1(v_0 = v_1 \vee v_1 < v_0)$ is logically equivalent to

$$\exists v_1(v_0 = v_1) \vee \exists v_1(v_1 < v_0).$$

It is clear that each disjunct is a logical consequence of D. And so we may replace '$\exists v_1\neg(v_0 < v_1)$' by '$v_0 = v_0$'.

We readily see that (iv) and hence (iii) are true sentences of D. And since (iii) is logically equivalent to the negation of (ii), we conclude that (ii) and hence (i) are not true sentences of D.

This completes our illustration of a decision procedure generated by the proof of a quantifier elimination theorem. Tarski's decision procedure for the first-order theory $\langle R, +, .\rangle$ follows the same general pattern. The central result of [Tarski, 1948a] is a quantifier elimination theorem for $\langle R, +, .\rangle$ (Theorem 31, p. 39).[94] The decision method is a straightforward consequence of the theorem, and follows soon after (Theorem 37, p. 42).

The proofs and results of [Tarski, 1948a] have a long history. Tarski reports that he had obtained partial results tending in the same direction — such as the decidability of elementary algebra with addition as the only operation, and of the geometry of the straight line — in his university lectures of 1926–8 (see footnote 4 of [Tarski, 1948a], where Tarski refers the reader to to Presburger [**?**, p. 95, fn. 4], and Tarski [1921, p. 324, fn. 3].) Tarski remarks that the decision procedure for the first-order theory of the reals was found in 1930 (see the Preface of [Tarski, 1948a] and also p. 2, especially footnote 4). Tarski's paper [Tarski, 1931a] on the definability of the reals (which we have discussed in Section 3 above) implicitly mentions a quantifier elimination result for $\langle R, 1, \leq, +\rangle$, where the basic formulas are '$x = 1$', '$x \leq y$)', '$x+y = z$'. Recall Theorem 1 from [Tarski, 1931a] (mentioned in Section 3 above):

THEOREM 9. *A set S of sequences of real numbers is a member of Df if and only if S is a finite sum of finite products of elementary linear sets.*

Theorem 9 is a theorem of mathematics, but as we noted in Section 3.2 there is a correlation between (mathematical) sets of sequences and (metamathematical) sentential functions, and between their respective Boolean operations — and so there is a metamathematical analogue of Theorem 9. Tarski writes:

> We can easily formulate (and prove) a metamathematical theorem which is an exact analogue of Th. 1. This metamathematical result leads us to a conclusion that, in the system of arithmetic described in Sec 1 [viz. $\langle R, 1, \leq, +\rangle$], every sentence of order 1 can be proved or disproved. Moreover, by analysing the proof of this result, we see that there is a mechanical method which enables us to decide in each particular case whether a given sentence (of order 1) is provable or disprovable. [Tarski, 1931a, p. 134]

The metamathematical result that Tarski mentions here is a quantifier elimination theorem for $\langle R, 1, \leq, +\rangle$ expressed in terms of definability. Tarski also mentions that the decidability of $\langle R, 1, \leq, +\rangle$ (as well as its completeness) follows from the proof of the theorem.

[94] Tarski remarks that, from a purely mathematical point of view, the results leading up to Theorem 31, and Theorem 31 itself, are closely related to Sturm's theorem. Indeed, Tarski writes that "Theorem 31 constitutes an extension of Sturm's theorem. . . to arbitrary systems of equations and inequalities with arbitrarily many unknowns." (p.52, fn 12).

Immediately after the passage just quoted, Tarski points out that the theory of the real numbers can be based on different systems of primitive concepts — for example, it can be based on the concepts of sum and product. Then a modified form of Theorem 9 holds for $\langle R, +, . \rangle$. Here we can see an implicit mention of the quantifier elimination theorem for $\langle R, +, . \rangle$, since the modified version of Theorem 9 will have its metamathematical analogue too. Tarski first *explicitly* mentions the quantifier elimination theorem for $\langle R, +, . \rangle$ in *The completeness of elementary algebra and geometry* [Tarski, 1967], a monograph that was scheduled for publication in 1940.[95] Tarski also provides an outline of the proof. In this work the emphasis is on the *completeness* of $\langle R, +, . \rangle$ (see Theorem 2.1, in Sec 2). Referring to [Tarski, 1967] and [Tarski, 1948a], Tarski writes:

> "A comparison of the titles of the two monographs reveals that the center of scientific interest has been shifted to the decision problem from that of completeness". (Foreword to [Tarski, 1967])

Indeed, from [Tarski, 1931a] to [Tarski, 1967] and on to [Tarski, 1948a] we see a shift in emphasis from definability to completeness to decidability. In [Tarski, 1948a], mention of definability and completeness is relegated to footnotes. In footnote 13 of [Tarski, 1948a] Tarski remarks that the quantifier elimination theorem "gives us a simple characterization of those sets of real numbers, and relations between real numbers, which are arithmetically definable", and refers the reader back to [Tarski, 1931a]. Completeness is mentioned in footnote 15. The emphasis on decidability is at least in part explained by the fact that the RAND corporation supported the publication of [Tarski, 1948a]. Tarski writes:

> "Within a few months the monograph was published. As was to be expected, it reflected the specific interests which the RAND Corporation found in the results. The decision method for elementary algebra and geometry — which is one of the main results of the work - was presented in a systematic and detailed way, thus bringing to the fore the possibility of constructing an actual decision machine. Other, more theoretical aspects of the problems discussed were treated less thoroughly, and only in notes." (Preface to [Tarski, 1948a])

5.2 The decidability of elementary geometry

As Tarski mentions here, and as its title suggests, the monograph [Tarski, 1948a] contains a proof not only of the decidability of the first-order theory of the reals

[95] The monograph was to appear in 1940 in *Actualites Scientifiques et Industrielles*, Hermann & Cie, Paris. In the Foreword to [Tarski, 1967], Tarski reports that the process of publication reached the stage of page proofs, but then, during the war, the whole setting was destroyed. All that remained were two sets of page proofs in Tarski's possession. Several years later, J. C. C. McKinsey re-wrote the monograph, and this version was published as [Tarski, 1948a]. In 1967, the original page proofs were reprinted, largely due to the efforts of Rene de Possel, Director of the Institut Blaise Pascal in Paris. Sczcerba remarks (in [**?**, p. 908]) that Tarski first saw the reprint of the original proofs in 1967 on his 65th birthday.

but also of the decidability of elementary (that is, first-order) geometry. Tarski outlines a decision method for the specific case of 2-dimensional Euclidean geometry, but points out that the method can be adapted to Euclidean geometry of any number of dimensions, as well as to various systems of non-Euclidean and projective geometry.[96] Tarski observes in footnote 18 (pp. 55–57) that, just as in the case of elementary algebra, one may proceed axiomatically, in terms of provability rather than truth. Tarski presents a list of axioms for 2-dimensional Euclidean geometry, which is easily modified to form a basis for elementary geometry of any number of dimensions. One of the fruits of the decision method in this axiomatic setting is a "constructive consistency proof for the whole of elementary geometry" (fn. 18, p. 57)[97]

The decision method for 2-dimensional Euclidean geometry is readily obtained from the decision method for elementary algebra, since each sentence of elementary geometry can be suitably correlated with a sentence of elementary algebra. Following Tarski (pp. 43–45), we will first describe the system of 2-dimensional Euclidean geometry and then show how to set up the correlation.

The language of 2-dimensional Euclidean geometry contains infinitely many variables ranging over points of the Euclidean plane, and three predicate constants: the identity sign '$=$', the 3-place predicate '$B(x, y, z)$' to be read as 'y is between x and z', and the 4-place predicate '$D(x, y, z, w)$' to be read as 'the distance from x to y is equal to the distance from z to w'. The atomic formulas are '$x = y$', '$B(x, y, z)$', and '$D(x, y, z, w)$'; complex formulas are built from the atomic formulas via the usual sentential connectives and the quantifiers. It is noteworthy that in Tarski's formalization only points are treated as individuals, and there are no set-theoretical devices or second-order variables. This is in contrast to Hilbert's influential *Grundlagen der Geometrie*, where certain geometrical figures are treated as individuals, and the others are treated as second- order point sets.[98]

[96] In footnote 19, Tarski mentions that ordinary projective geometry may be treated as a branch of the theory of modular lattices, referring the reader to Birkhoff [Birkhoff, 1948]. Tarski goes on to say that the decision method also applies to this branch of lattice theory.

[97] Tarski observes that this improves a result to be found in Hilbert-Bernays [Hilbert and Bernays, 1939, vol 2, pp. 38ff].

[98] Tarski draws the contrast in [Tarski, 1959, fn. 3] and the associated text.

> "Thus, in our formalization of elementary geometry, only points are treated as individuals and are represented by (first-order) variables. Since elementary geometry has no set-theoretical basis, its formalization does not provide for variables of higher orders and no variables are available to represent or denote geometrical figures (point sets), classes of geometrical figures, etc. It should be clear that, nevertheless, we are able to express in our symbolism all the results which can be found in textbooks of elementary geometry and which are formulated there in terms referring to various special classes of geometrical figures, such as the straight lines, the circles, the segments, the triangles, the quadrangles, and, more generally, the polygons with a fixed number of vertices, as well as to certain relations between geometrical figures in these classes, such as congruence and similarity. This is primarily a consequence of the fact that, in each of the classes just mentioned, every geometrical figure is determined by a fixed finite number of points." [Tarski, 1959, p. 20]

We can translate a sentence ϕ of elementary geometry into a sentence $\phi*$ of elementary algebra by fixing a coordinate system. We make the following replacements:

(i) If x is a (geometric) variable in ϕ, we replace it by two (algebraic) variables x_1 and x_2. (Intuitively, x represents a point in the Euclidean plane, and x_1 and x_2 represent its real coordinates.) Distinct geometric variables are replaced by distinct algebraic variables.

(ii) The quantifier phrases $\exists x$ and $\forall x$ are replaced by $\exists x_1 \exists x_2$ and $\forall x_1 \forall x_2$ respectively.[99]

(iii) '$x = y$' is replaced by '$x_1 = y_1 \& x_2 = y_2$'.

(iv) '$B(x, y, z)$' is replaced by

$$`(y_2 - x_2).(z_1 - y_1) = (z_2 - y_2).(y_1 - x_1) \& 0 \leq (x_1 - y_1).(y_1 - z_1) \\ \& 0 \leq (x_2 - y_2).(y_2 - z_2)'.$$

(v) '$D(x, y, z, w)$' is replaced by

$$'(x_1 - y_1)^2 + (x_2 - y_2)^2 = (z_1 - w_1)^2 + (z_2 - w_2)^2'.$$

For example, the true sentence

$$\forall x \forall y \forall z \exists w D(x, y, z, w$$

of elementary geometry is correlated with the true sentence

$$\forall x_1 \forall x_1 \forall y_1 \forall y_2 \forall z_1 \forall z_2 \exists w_1 \exists w_2 [(x_1 - y_1)^2 + (x_2 - y_2)^2 = (z_1 - w_1)^2 + (z_2 - w_2)^2]$$

of elementary algebra. Tarski makes the general observation:

> It is now obvious to anyone familiar with the elements of analytic geometry that whenever ϕ is true then $\phi*$ is true, and conversely. [Tarski, 1948a, p. 45]

It follows that 2-dimensional elementary geometry is decidable:

> And since we can always decide in a mechanical way about the truth of $\phi*$, we can also do this for ϕ. (*ibid*)

Tarski proved a number of other decidability results that we have not yet mentioned, all by the method of quantifier elimination. They include decidability results for the theory of dense linear orderings (proved in Tarski's lectures of 1926–8); for the theory of wellorderings (work started with Mostowski in the 1930s, and culminating in [Tarski, 1978]); and for the theory of Boolean algebras (announced

[99] Alternatively, we can work just with '$\exists$', as Tarski does, and take '$\forall$' to abbreviate '$\neg\exists\neg$'.

in [Tarski, 1949f]).[100] Although the method of quantifier elimination did not originate with Tarski, it was in his hands an extremely powerful systematic tool, producing a number of important results about a wide range of theories. None were more significant than the decision method for real-closed fields, described by Doner and Hodges as "an astonishingly fruitful mathematical result" [Doner and Hodges, 1988, p. 23]; for a detailed account of the impact of this result of Tarski's, see van den Dries [van den Dries, 1988, esp. pp. 10–16].

5.3 Undecidability

In his paper "A general method in proofs of undecidability" [Tarski, 1953, pp. 3–35][101], Tarski distinguishes two different methods for solving the decision problem.

1. *The direct method* The 'direct' method is based on ideas developed in Gödel [Gödel, 1931]. As Tarski summarizes it:

   > "A theory T is called *decidable* if the set of all its valid sentences is recursive, and otherwise *undecidable*." [Tarski, 1953, p. 14]

   This is the method used by Church to show the undecidability of Peano's arithmetic, and by Rosser to show that every consistent extension of Peano's arithmetic is also undecidable.

2. *The indirect method* This method

   > "consists in reducing the decision problem for a theory T_1 to the decision problem for some other theory T_2 for which the problem has previously been solved." [Tarski, 1953, p. 4]

   In the original form of the indirect method, there are two ways to proceed:

   (i) Where T_2 is undecidable, we show that T_1 can be obtained from T_2 by deleting finitely many of T_2's axioms.

   (ii) Where T_2 is essentially undecidable, we show that T_2 is interpretable in T_1.

Church applied procedure (i) to prove that first-order predicate logic is undecidable, taking T_2 to be a fragment of Peano arithmetic (PA). Tarski observes that

[100] For more details and further references, see Doner and Hodges [Doner and Hodges, 1988, pp. 21–23]. Several of Tarski's students — Presburger, Szmielew, and Doner — proved decidablity results under Tarski's supervision (see [Doner and Hodges, 1988, pp. 23–4]). In particular, Presburger proved the decidability of the arithmetic of the natural numbers with addition in [Presburger, 1930] — his master's thesis, supervised by Tarski.

[101] This paper is the first of three that compose Tarski [Tarski, 1953] (written in collaboration with Mostowski and A. Robinson). Tarski reports that the observations contained in the paper were made in 1938–9, presented to a meeting of the Association for Symbolic Logic in 1948, and summarized in [Tarski, 1949b].

procedure (ii) often proceeds with PA as T_2 — in this way, for example, "various axiomatic systems of set theory have turned out to be undecidable." (p. 4)

However, Tarski also observes that the indirect method is limited:

> "The indirect method in its original form was rather restricted in applications. Only in exceptional cases can a theory for which the decision problem is discussed be obtained from another theory, which is known to be undecidable, simply by omitting finitely many sentences from the axiom system of the latter. On the other hand, one could hardly expect to find an interpretation of Peano's arithmetic in various simple formalized theories, with meager mathematical contents, for which the decision problem was open. With regard to theories of this kind both the direct and indirect methods seem to fail." (p. 4)

Tarski's novel contribution was to extend and modify the indirect method, significantly enlarging the scope of its application. Here the key notion is that of an *essentially undecidable theory*: a theory is *essentially undecidable* if it and every consistent extension of it is undecidable. The observation at the heart of Tarski's method is this:

(I) "[I]n order to establish the undecidability of a theory T_1, it suffices to show that some essentially undecidable theory T_2 can be interpreted, not necessarily in T_1, but (what is much easier) in some consistent extension of T_1 — provided only that T_2 is based upon a finite axiom system." (p. 5)

Since T_2 must be finitely axiomatizable, PA can no longer serve as T_2. However,

> "examples of essentially undecidable theories which are based upon finite axiom systems and are readily interpretable in other theories have been found (by the direct method) among fragments of Peano's arithmetic. Using this fact and applying the extended indirect method, many formalized theories — like the elementary theories of groups, rings, fields, and lattices — have recently been shown to be undecidable." (p. 5)

Not sure which reference this should be

No item for Robinson 1950 in bib.

A particularly useful example is the theory Q, a finitely axiomatizable and essentially undecidable fragment of PA found by Tarski and Mostowski using the direct method in 1939 (see [Tarski, 1949c]), and considerably simplified by Robinson (in [?]).[102] Tarski writes:

> "Theory Q turned out to be very suitable for our method; its mathematical content is meager, and it can easily be interpreted or at least weakly interpreted in many different theories. Hence Theory Q has become a powerful instrument in the study of the decision problem;

[102] The construction of Q is carried out in "Undecidability and Essential Undecidability in Arithmetic", by Mostowski, Robinson and Tarski, the second paper in [Tarski, 1953]. For historical information about the construction, see [Tarski, 1953, pp. 39–40, fn. 1].

> with its help it has proved to be possible to obtain a negative solution of this problem for a large variety of theories for which the problem had previously been open." (p. 32)

In "A general method in proofs of undecidability", Tarski lays out the theoretical foundations for his indirect method. In sections 2 and 3, Tarski defines a number of basic concepts and proves six theorems about undecidable and essentially undecidable theories, with no mention yet of the notion of interpretability. We first review some of Tarski's definitions:

> "A theory T_1 is called a *subtheory* of a theory T_2 if every sentence which is valid in T_1 is also valid in T_2; under the same conditions T_2 is referred to as an *extension* of T_1." (p. 11)
>
> "An extension T_2 of T_1 is called *inesssential* if every constant of T_2 which does not occur in T_1 is an individual constant and if every valid sentence of T_2 is derivable in T_2 from a set of valid sentences of T_1." (p. 11)
>
> "Two theories T_1 and T_2 are said to be *compatible* if they have a common consistent extension." (p. 12)
>
> A theory is *axiomatizable* "if there is a recursive set A of valid sentences of T such that every valid sentence of T is derivable from the set A; if the set A is assumed to be finite, the theory T is called *finitely axiomatizable*." (p. 14)

One useful theorem Tarski proves in section 3 is the following:

THEOREM 10. *Let T_1 and T_2 be two compatible theories such that every constant of T_2 is also a constant of T_1. If T_2 is essentially undecidable and finitely axiomatizable, then T_1 is undecidable, and so is every subtheory of T_1 which has the same constants as T_1.*

Later in the paper (p. 32), Tarski gives an example of an application of Theorem 6. Consider theory N, the arithmetic of the natural numbers, with non-logical constants + and ., and possibly others, such as 0 and 1 and <. By definition, a sentence of N is valid if it is true under the standard intepretation. Let us observe the following:

(i) N is undecidable, and since it is also complete, it is essentially undecidable and not axiomatizable.[103]

[103] Theorem 9 states:

> For a complete theory T the following three conditions are equivalent: (i) T is undecidable, (ii) T is essentially undecidable, and (iii) T is not axiomatizable. (See [Tarski, 1953, p. 14].)

The proof is straightforward: for a complete theory, (i) implies (iii), and the other parts of the theorem are immediate from the definitions.

(ii) PA is a axiomatic subtheory of N with the same constants, and though axiomatizable, PA is not finitely axiomatizable.

(iii) In turn, Q is a subtheory of PA which is finitely axiomatizable, and essentially undecidable.

Given (iii), PA has a finitely axiomatizable, essentially undecidable subtheory. And so it follows from our observations that N has a finitely axiomatizable, essentially undecidable subtheory.

Now if we apply Theorem 10

> "we arrive at once nt the conclusion that every theory which is compatible with Q and has the same constants as Q is undecidable. Hence, in particular, every subtheory of N in which the set of constants includes $+$ and . is undecidable ... this generalizes the results known from the literature which concern various special subtheories of N." (p. 32)

The theorems Tarski proves in section 3 — including Theorem 10 — make no mention of interpretability. In section 4, Tarski widens the scope of these theorems by introducing the notion of interpretability. It is here in section 4 that Tarski explicitly provides the grounds for the method encapsulated in (I) above. The notion of intepretability rests on the prior notion of *definition*, or more specifically the notion of a *possible definition of a given constant in a theory T*. Tarski introduces the notion by way of examples, one of which is the 2-place predicate symbol $<$:

> "A possible definition of $<$ in T is any sentence of the form
>
> (i) $\forall x \forall y (x < y \leftrightarrow \phi)$
>
> where ϕ is a formula of T (x and y being any two different variables such that no variable different from both of them occurs free in ϕ)." (p. 20)

Notice that while (i) is not a sentence of T, it is a sentence in every extension of T which contains $<$ as a constant. It is straightforward to extend the notion of a possible definition to n-place predicate and operation letters.

Now consider two theories T_1 and T_2, and assume for the moment that T_1 and T_2 have no non-logical constants in common. Tarski characterizes the notion of interpretablity of one theory in another as follows:

> "... T_2 is *interpretable* in T_1 if we can extend T_1, by including in the set of valid sentences some possible definitions of the nonlogical constants of T_2, in such a way that the resulting extension of T_1 turns out to be an extension of T_2 as well." (p. 21)

The notion of interpretability is easily extended to the general case in which T_1 and T_2 may have some non-logical constants in common.[104] Tarski also introduces the notion of *weak interpretability*:

[104] In the general case

> "A theory T_2 is said to be *weakly interpretable* in T_1 if T_2 is interpretable in some consistent extension of T_1 which has the same constants as T_1." (p. 21)

Tarski goes on in section 4 to prove several theorems that turn on the notions of interpretability and weak interpretability. One of these theorems is the following:

THEOREM 11. *Let T_1 and T_2 be two theories such that T_2 is weakly interpretable in T_1 or in some inessential extension of T_1. If T_2 is essentially undecidable and finitely axiomatizable, then:*

(i) T_1 is undecidable and every subtheory of T_1 which has the same constants as T_1 is undecidable;

(ii) there exists a finite extension of T_1 which has the same constants as T_1 and is essentially undecidable. (pp. 23–4)

Theorem 11 is a generalization of Theorem 10, for if T_1 and T_2 are compatible and every constant of T_2 is also a constant of T_1, then T_2 is weakly interpretable in T_1. Tarski remarks that, from the point of view of providing theoretical foundations for his general indirect method,

> "Theorem 11 is especially important ...; for, given a finitely axiomatizable and essentially undecidable theory, this theorem enables us to establish the undecidability of various other theories which may be very distant in their mathematical content from the original theory." (p. 30)

Theorem 11 captures in a precise way the observation in (I) above.

For an example of the application of theorems from section 4 (including Theorem 11), we need one more notion, that of *relative interpretability.* Given a theory T and a 1-place predicate P, the theory $T^{(P)}$ is obtained by relativizing the quantifiers in T to P; that is, each subformula $\forall x\psi$ (or $\exists x\psi$) is replaced by $\forall x(Px \rightarrow \psi)$ (or $\exists x(Px \& \psi)$). Now, a theory T_2 is *relatively interpretable* (*relatively weakly interpretable*) in a theory T_1 if $T_2^{(P)}$ is interpretable (weakly interpretable) in T_1. Among the theorems concerning relative interpretability is the following:

THEOREM 12. *Let T be any theory and P a unary predicate which is not a constant of T. Then $T^{(P)}$ is essentially undecidable if and only if T is essentially undecidable. (p. 27)*

Now consider again the theories N and Q, and note that, by Theorem 12, $Q^{(P)}$ is essentially undecidable. Tarski writes:

> "we first replace the non-logical constants in T_2 by new constants not occurring in T_1 (different symbols by different symbols), without changing the structure of T_2 in any other respect; if the resulting theory T_2' proves to be interpretable in T_1, we say that T_2 is also interpretable in T_1." (p.21)

> "N is known to be relatively interpretable in the arithmetic of integers J, a theory constructed analogously to N, but in which the variables are assumed to range over the set of all integers. Hence, Q is also relatively interpretable in J. By applying Theorems 11–12 we conclude that every subtheory of J in which $+, .,$ and possibly some other symbols occur as nonlogical constants is undecidable. As interesting examples of such subtheories we may mention the elementary theories of rings, commutative rings, and ordered rings." (p. 33)

No entry for Robinson 1949 in bib.

A parallel result was established by Julia Robinson for the arithmetic A of rational numbers (see [**?**]). Robinson proved that N, and so Q, is relatively interpretable in A — and that consequently all the subtheories of A (with + and . as non-logical constants) are undecidable. These subtheories include the elementary theories of fields and ordered fields.

As Tarski mentions (pp. 33–34), many other results were established using the indirect method, by Tarski and by others.[105] Tarski established the undecidability of the elementary theories of groups, lattices, modular lattices, complemented modular lattices, and abstract projective geometries (see [Tarski, 1949d; Tarski, 1949e]). Tarski also established undecidability results for equational theories of relation algebras,[106] and for certain branches of elementary geometry.[107] W. Szmielew and Tarski proved that all known axiomatic systems of set theory, with the 1-place predicate S ('is a set') and '$\in$' ('is a member of') as nonlogical constants, are undecidable; they proved this general result by showing that Q is interpretable in a small axiomatic, essentially undecidable fragment of set theory.[108] R. M. Robinson showed that the elementary theories of various special rings are undecidable (see [**?**]). Grzegorczyk proved the undecidability of distributive lattices, Brouwerian algebras, and related algebraic and geometric systems (see [**?**]).

No work with Szmielew in bib for 1950.

In his proofs of decidability results, Tarski took a known method — the elimination of quantifiers — to a new level. But in his proofs of undecidability results, Tarski established a new method, an indirect method that turned on the semantic notions of *definition* and *interpretability*. Indeed, Tarski's method is often referred to as the "method of interpretation". So at the heart of Tarski's original contributions to the decision problem lies his groundbreaking work on fundamental semantic notions.

[105] For a survey of more recent results inspired by Tarski's method, see [McNulty, 1986, esp. pp. 893–6].

[106] These results can be found in [Tarski, 1987a, section 8.5, pp. 251–258]. The history of these results is reviewed in section 8.7, pp. 268–271. Relation algebras are characterized on pp. 235–236. As we will see in Section 6 of this chapter, Tarski initiated the study of relation algebras in his paper [Tarski, 1941b]. An equational theory is a restricted elementary theory, restricted to those universal sentences whose quantifer-free subformulas are equations between terms.

[107] See Tarski [1959] and Tarski and Szczerba [1979].

[108] These and related results are stated without proof in W. Szmielew and Tarski [**?**].

6 MORE ON LOGIC AND ALGEBRA

Tarski's work on algebra and logic flows in two directions, from algebra to logic, and back again. As we saw in 2.4 above, Tarski increasingly 'algebraized' logic, investigating logical systems as interpretations of algebras. In the other direction, Tarski brings metalogical questions to bear on algebras; we saw, for example, that Tarski established the decidability of Boolean algebras (5.1 above) and the undecidability of equational theories of relation algebras (5.3 above). In this section, we say more about Tarski's work on the connections between algebra and logic.

6.1 Boolean algebra and the calculus of classes

With the first direction in mind, recall that in [Tarski, 1935+1936] Tarski treated the calculus of sentences and the calculus of deductive systems as two realizations of Boolean algebra (see 2.4 above). In his 1935 paper "On the Foundations of Boolean Algebra" [Tarski, 1935], Tarski shows that another logical theory, the calculus of classes, can be treated in the same way.[109] Tarski opens his paper as follows:

> "Boolean algebra, also called the *algebra of logic*, is a formal system with a series of important interpretations in various fundamental departments of logic and mathematics. The most important and best known interpretation is the *calculus of classes*." [Tarski, 1935, p. 320]

Tarski goes on to investigate several systems of Boolean algebra, and shows that one of them — the 'atomistic system of Boolean algebra' - is equivalent with the calculus of classes. So, along with [Tarski, 1935+1936], [Tarski, 1935] is a contribution to the algebraization of logic.

Recall Postulates I–VII (presented in 2.4 above, pp. 13–14. These form what Tarski calls the *ordinary system of Boolean algebra*. To these axioms, Tarski adds three more, each of an infinite character. These axioms contain two new primitive operations: $\sum_{y\in x} y$ (the sum of all elements of the set X) and $\prod_{y\in x} y$ (the product of all elements of the set X). The additional axioms are as follows (where, recall, B is the universe of discourse):

Postulate 8 If $X \subseteq B$, then (a) $\sum_{y\in x} y \in B$; (b) $x < \sum_{y\in X} y$ for every $x \in X$; (c) if $z \in B$ and $x < z$ for every $x \in X$, then $\sum_{y\in x} y < z$.

Postulate 9 If $X \subseteq B$, then (a) $\prod_{y\in X} yB$; (b) $\prod_{y\in X} y < x$ for every $x \in X$; (c) if $z \in B$ and $z < x$ for every $x \in X$, then $z < \prod_{y\in X} y$.

Postulate 10 If $x \in B$ and $X \subseteq B$, then (a) $x.\sum_{y\in X} y = \sum_{y\in X}(x.y)$; (b) $x + \prod_{y\ inX} y = prod_{y\ inX}(x + y)$.

[109] We presented the calculus of classes in 4.2 above (it was the object language for Tarski's definition of truth in [Tarski, 1933a]). Variables range over classes, and there are four constants: negation, disjunction, the universal quantifier, and inclusion.

Postulates 1–10 form the *extended* (or *complete*) *system of Boolean algebra*.[110]

In the second half of [Tarski, 1935], Tarski constructs the *atomistic system of Boolean algebra*. The set of atoms, At, is defined as follows:

DEFINITION 13. $x \in At$ (1) if and only if $x \in B$ and $x \neq 0$, and (2) for every element $y \in B$, the formulas $y < x$ and $y \neq 0$ imply $y = x$.

Tarski now adds one further axiom to Postulates 1–10:

Postulate D If $x \in B$ and $x \neq 0$, then there is an element $y \in At$ such that $y < x$.

Without Postulate D, the following questions are all open: Do atoms exist? Are atoms included in every element (other than 0)? Is every element the sum of the atoms included in it? With Postulate D, all these questions are answered in the affirmative. The system of postulates 1-10+D is the *atomistic system of Boolean algebra*. It is this system of algebra that Tarski shows to be equivalent with the calculus of classes (see Theorem 6, p. 340).[111]

6.2 The calculus of relations and relation algebras

In a paper published a few years later — "On the Calculus of Relations" [Tarski, 1941b] — Tarski shows that we can take the algebraic view of another logical theory, the *calculus of (binary) relations*. Tarski traces the beginnings of this theory back to de Morgan, but "[t]he title of creator of the theory of relations was reserved for C. S. Peirce" [Tarski, 1941b, p. 73]. Peirce's work "was continued and extended in a very thorough and systematic way by E. Schröder" (*ibid.*) in *Algebra und Logik der Relative* ([Schröder, 1895]).[112] Schröder's thorough account

[110] In the first half of [Tarski, 1935], Tarski shows that this axiomatization can be simplified — first to a system of four axioms, and then to a system of two (see pp. 323–333). Tarski reports (in a footnote added later to [Tarski, 1935, on p. 333]) that these simplifications were influenced by work of Lesniewski, specifically the deductive theory that Lesniewski called *mereology* (see Lesniewski [1916] and [1927–31]).

[111] Tarski remarks that Postulate D "brings with it a whole series of far-reaching consequences" [Tarski, 1935, p. 335]. Among these are the general distributive laws for addition and multiplication:

(E) Let R be any class of subsets X of the set $B(\sum_{X\in R} X \subseteq B)$, and let S be the class of all those sets Y which are included in the sum of all sets of the class R and have at least one element in common with every set of the class R ($Y \subseteq \sum_{X\in R} X$, and if $X \in R$, then $X.Y \neq 0$). Then:

(a) $\prod_{X\in R}\sum_{Z\in X} z = \sum_{Y\in S}\prod_{Z\in Y} z$, and

(b) $\sum_{X\in R}\prod_{Z\in X} z = \prod_{Y\in S}\sum_{Z\in y} z$.

The sentences (a) and (b) are generalizations of Postulates 5 and 10 respectively. Tarski and Lindenbaum showed that (E) cannot be derived from Postulates 1–10 alone; indeed, given Postulates 1–10, (E) is equivalent to Postulate D.

Tarski's work on general distributive laws in Boolean algebras, started here in [Tarski, 1935], was developed further in Tarski and Smith [1957a].

[112] This work of Schröder's was the third volume of *Vorlesungen uber die Algebra der Logik* [Schröder, 1890–1905].

of the calculus of relations contains a large number of unsolved problems, and indicates the direction for further investigations — consequently, Tarski expresses his amazement that there was no specific development of this rich logical theory in the decades that followed. Tarski sought "to awaken interest in a certain neglected logical theory" [Tarski, 1941b, p. 89]

Tarski contrasts two methods for constructing the calculus of relations. The first constructs the calculus as part of a larger logical theory (the functional calculus developed by Hilbert and Ackermann). The second method is specific to the calculus of relations, and as a result it is simpler and more elegant. This method is an algebraic construction which originates with Tarski, and we turn to it now.

The vocabulary of Tarski's construction comprises relation variables, the usual sentential connectives, the familiar constants and operations from Boolean algebra ($1, 0, ^{-}, +$, and .), and five further symbols peculiar to the calculus of relations: $1'$ (the identity element, or Peircean unit), $0'$ (the diversity element, or Peircean zero), $\smile$ (the unary operation of conversion), ; (the binary operation of relative product), and + (the binary operation of relative addition).[113] To fix ideas, think for the moment of relations set-theoretically, as sets of ordered pairs. Then the identity element is the set of identity pairs, the converse of a relation R is given by:

$$\breve{R} = \{\langle x, y\rangle | \langle y, x\rangle \in R\},$$

and the relative product of R and S is given by

$$R; S = \{\langle x, z\rangle | \langle x, y\rangle \in R \& \langle y, z\rangle \in S\}.$$

Peircean zero and relative addition are dispensable, since each can be defined in terms of the other symbols (as we shall see shortly).

The axioms of Tarski's algebra fall into three groups: the usual axioms for the sentential connectives, the axioms for Boolean algebra (with class variables replaced by relation variables), and the axioms governing the further 'Peircean' constants and operations. Tarski lists eight axioms in this third group:

(i) $\overline{\overline{R}} = R.$

(ii) $\smile R; S = \breve{S}; \breve{R}.$

(iii) $R; (S; T) = (R; S); T$

(iv) $R; 1' = R.$

(v) $R; 1 = 1v1; \bar{R} = 1.$

(vi) $(R; S).\breve{T} = 0 \rightarrow (S; T).\breve{R} = 0.$

[113] Perhaps more familiar now than $1', 0'$, ;, and $\pm$ are the symbols $\mathring{1}$, $\mathring{0}$, $\odot$, and $\oplus$, respectively. See for example, [Tarski, 1987a, pp. 235–6], where a characterization of a *relation algebra* may also be found.

(vii) $0' = \bar{1}'$.

(viii) $R \pm S = \overline{\bar{R}; \bar{S}}$.

Axioms (i)–(iv) involve exclusively $1', \smile$, and ;. Though fully abstract, they are natural and obvious when we read them in the more concrete setting of set theory. Axioms (v) and (vi) establish connections between Boolean and 'Peircean' concepts. Axioms (vii) and (viii) can be understood as definitions of O' and $\pm$.[114]

Here we have the first presentation of the calculus of relations as an axiomatic algebraic theory. Moreover, Tarski's paper [Tarski, 1941b] initiated the study of *relation algebras*. Tarski's work on the arithmetic of relation algebras at Berkeley during the 1940s culminated with the paper "Distributive and modular laws in the arithmetic of relation algebras" [Tarski, 1951a], which Tarski wrote with his student Louise Chin. In [Tarski, 1951b] and [Tarski, 1952], Parts 1 and 2 of "Boolean algebras with operators", Tarski and Jonsson investigated algebraic systems enriched by new operations, and relation algebras figure prominently in this work.[115]

[114]Tarski points out that these axioms can be represented geometrically. Suppose that the relation variables denote relations between real numbers, and consider a rectangular coordinate system in the plane. Then every relation R may be represented as the set of all points (x, y) such that x has the relation R to y. The relations $1, 0, 1', 0'$ are respectively: the whole plane, the empty set of points, the identity relation represented by the straight line with equation $x = y$, and the diversity relation by the set of all points not on this straight line. The Boolean operations are represented in the usual set-theoretic way. To obtain R, take the set of points - or point set — corresponding to R and rotate it (in three dimensions) through an angle of 180^o about the line $x = y$. It is rather more complicated to explain relative product geometrically. We need to supplement the coordinate system with a z-axis perpendicular to the xy-plane. Then for the representation of $R; S$ (in the xy-plane) proceed as follows: rotate the point set corresponding to R through a right angle about the x-axis, draw through every point of the resulting set a straight line parallel to the y-axis, and take the union of all these straight lines — this gives us the 'cylindrical' point set $R*$. Similarly, rotate the set corresponding to S through a right angle about the y-axis, draw the lines parallel to the x-axis, and obtain the 'cylindrical' point set $S*$. Finally, take the intersection of $R*$ and $S*$ and project it orthogonally on the xy-plane. This projection is the geometrical representation of $R; S$. The representation of relative sum is readily obtained from that of the relative product, given axiom (viii).

This geometric representation provides an intuitive reading of the axioms. For example, axiom (i) corresponds to the obvious geometric fact that if we rotate any point set through 180^o about a given straight line, and then again by 180^o about the same straight line, we obtain the original point set. Or consider axiom (v): we can see that this axiom is true under the geometric representation. Observe first that $R; 1$ holds between x and y if and only if there is a z such that xRz (the relation is independent of y). So to obtain the geometric representation of $R; 1$ consider the set corresponding to R, draw through every point of this set a straight line parallel to the y-axis, and take the union of all these straight lines. Now axiom (v) says that either the set representing $R; 1$ is the whole plane or the set representing $1; R$ is the whole plane. If $R; 1$ is not the whole plane then there must be a line parallel to the y-axis which does not contain any point in the set corresponding to R. So this line is composed only of points belonging to the set corresponding to R. If through every point of this line we draw a straight line parallel to the x-axis and take the union of all these parallel lines, we obtain the whole plane. That is, the relation $1; R$ is represented by the whole plane.

[115]For more on the work by Tarski and others on relation algebras, see Monk [1986, esp. pp. 901–2] and Jonsson [1986, esp. p. 884].

Tarski's paper [Tarski, 1941b] is another clear example of Tarski's algebraic treatment of logical theories. But it also shows Tarski working in the reverse direction, raising metalogical questions for his algebraic theory. For example, Tarski raises a question as to the equivalence of the two methods of construction. Declaring this a difficult open question, Tarski continues:

> "I can only say that I am practically sure that I can prove with the help of the second method all of the hundreds of theorems to be found in Schroder's *Algebra und Logik der Relative*." (p. 88)

For a second example, Tarski turns to the so-called "representation problem":

> "Is every model of the axiom system of the calculus of relations isomorphic with a class of binary relations which contains the relations $1, 0, 1', 0'$ and is closed under all the operations considered in this calculus?" (p. 88)

M. H. Stone had proved the analogous result for Boolean algebra in the affirmative (see [Stone, 1936]). In 1941 Tarski's question was an open one — but Lyndon proved that the answer is in the negative (see [Lyndon, 1950], with a correction in [Lyndon, 1956]). Tarski maintained his interest in the metalogic of relation algebras throughout the rest of his career — Tarski's undecidability results for a variety of relation algebras, mentioned above and in 5.2, are contained in his final published monograph [Tarski, 1987a] (with Givant).[116]

6.3 Predicate logic and cylindric algebras

Thus far we have seen that Tarski provides algebraic treatments of a number of logical theories: the calculus of sentences, the calculus of deductive systems, the calculus of classes, and the calculus of (binary) relations. We turn now to Tarski's algebraic treatment of the predicate calculus. To this end, Tarski developed the theory of *cylindric algebras*:

> "This theory . . . was originally designed to provide an apparatus for an algebraic study of first-order predicate logic." [Tarski, 1971, Foreword, p.1]

Tarski collaborated with Henkin and Monk to produce *Cylindric Algebras* Part 1 [Tarski, 1971] and Part 2 [Tarski, 1985].

In the Foreword to [Tarski, 1971], Tarski, Henkin and Monk introduce cylindric algebras by way of *cylindric set algebras.* Just as Boolean algebra is an abstraction

[116] Tarski discusses other metalogical problems in [Tarski, 1941b], specific to the first method of constructing the calculus of relations. He announces two results: (i) the calculus of relations is undecidable, and (ii) there are sentences of the elementary theory of relations with only relation variables free that cannot be transformed into an equivalent sentence of the calculus of relations. Tarksi leaves unsolved a new decision problem regarding the transformations mentioned in (ii): is there a procedure for deciding in a particular case whether such a transformation is possible?

from Boolean set algebra (see Section 2.4 above, p. 12), so cylindric algebra is an abstraction from cylindric set algebra. So let us begin with cylindric set algebras. In the broadest terms, a *cylindric set algebra of dimension* α is a structure

$$A = \langle S, \cup, \cap, \sim, 0,^{\alpha} U, C_\kappa, D_{\kappa\lambda} \rangle_{\kappa,\lambda<\alpha}$$

where S is the universe of the algebra (the set of all its elements), $\cup$ and $\cap$ are binary operations, $\sim$ and $C_\kappa(\kappa < \alpha)$ are unary operations, and $0,^{\alpha} U, D_{\kappa\lambda}(\kappa, \lambda < \alpha)$ are distinguished elements. We now describe these ingredients in more detail.

It is helpful to describe cylindric set algebras using terminology drawn from analytic geometry (see [Tarski, 1971, pp. 1–2]). We start with the set ${}^{\alpha}U$. The elements of ${}^{\alpha}U$ are sequences $x = \langle x_0, x_1, \ldots x_k, \ldots \rangle$ of length α (α an ordinal), where each x_k is a member of an arbitrary set U. The set ${}^{\alpha}U$ is called the *α-dimensional Cartesian space with base* U, and each element x is called a *point*, with *coordinates* $x_0, x_1, \ldots x_k, \ldots$. The elements of the universe S are subsets of ${}^{\alpha}U$. (Just as the elements of a Boolean set algebra may be subsets of an arbitrary set V, so the elements of a cylindric algebra are subsets of a certain Cartesian power of an arbitrary set U.) The operations $\cup, \cap$ and are the usual Boolean operations; indeed, the structure

$$A = \langle S, \cup, \cap, \sim 0,^{\alpha} U \rangle$$

is a Boolean set algebra, with S closed under the Boolean operations.

The distinguished element $D_{\kappa\lambda}$ and the unary operation C_κ are specific to cylindric set algebras. $D_{\kappa\lambda}$ is a certain subset of ${}^{\alpha}U$ — its members are the points of ${}^{\alpha}U$ whose kth coordinate is identical to its λth coordinate. The point set $D_{\kappa\lambda}$ is called the *κ, λ-diagonal set*. The aptness of the name is easily seen if we consider the case $\alpha = 2$ (the case where the Cartesian space with the base U is 2-dimensional) — here $D_{\kappa\lambda}$ is the main diagonal line of the coordinate system. In general, for $k \neq l, D_{\kappa\lambda}$ is the hyperplane defined by $x_\kappa = x_\lambda$.

The unary operation C_κ is called the κth *cylindrification*. Intuitively, given a subset X of the space ${}^{\alpha}U, C_\kappa(X)$ is the cylinder obtained if each point in X is 'stretched out' parallel to the κth coordinate axis. More precisely,

> a point y is in $C_\kappa(X)$ iff there is a point in X which differs from y only in its κth coordinate.

(It is easy to check that if C_κ is applied to the diagonal set $D_{\kappa\lambda}$, it yields ${}^{\alpha}U$.)[117] The universe S is closed under c_κ.

We now move to the general notion of a *cylindric algebra* by abstracting from cylindric set algebras. We consider the algebraic identities that hold in *all* cylindric set algebras, and select some of them as axioms for the general, abstract theory of

[117] In general, $C_\kappa(D_{\kappa\lambda}) = C_\lambda(D_{\kappa\lambda}) =^{\alpha} U$. For example, consider the case $\alpha = 2$. If the 01 diagonal is stretched out parallel to either the 0 axis or the 1 axis, the result is the entire 2-dimensional Cartesian space with base U.

cylindric algebras. Changing the terminology, we take a cylindric algebra to be a structure

$$\mathfrak{A} = \langle A, +, ., -, 0, 1, c_\kappa, d_{\kappa\lambda} \rangle_{\kappa,\lambda<\alpha}$$

where A is an arbitrary set, $+, ., -, 0, 1$ are the familiar Boolean operations and distinguished elements, c_κ (for $\kappa < \alpha$) is a unary operation (still called the *κth cylindrification*), and $d_{\kappa\lambda}$ (for $\kappa, \lambda < \alpha$) is a distinguished element (called the *$\kappa\lambda$-diagonal element*). The axioms of a cylindric algebra fall into two groups. The first group of axioms characterize the Boolean algebra $\langle A, +, ., -, 0, 1\rangle$ (For example, we could use Postulates I–VII from 2.4 above.) The second group is composed of equational axioms that are special to cylindric algebras. These axioms are as follows, where x, y are arbitrary elements of A, and κ, λ, μ are arbitrary ordinals less than α:

1. $c_\kappa(0) = 0$.
2. $x.c_\kappa(x) = x$.
3. $c_\kappa(x.c_\kappa(y)) = c_\kappa(x).c_\kappa(y)$.
4. $c_\kappa(c_\lambda(x)) = c_\lambda(c_\kappa(y)).\mu$
5. $c_\kappa(d_{\kappa\lambda}) = 1$.
6. $c_\mu(d_{\kappa\mu}.d_{\mu\lambda}) = d_{\kappa\lambda}$, provided that $\mu \neq \kappa, \lambda$.
7. $c_\kappa(d_{\kappa\lambda}.x).c_\kappa(d_{\kappa\lambda}.-x) = 0$, provided $\kappa \neq \lambda$.

These axioms are natural enough if we keep in mind cylindric set algebra.[118] But in order to grasp the motivation for them, and for cylindric algebras generally, we must remember their original purpose: to serve "as an instrument for the algebraization of predicate logic" [Tarski, 1971, p. 4]. Cylindric algebras can be viewed not only as an abstraction from cylindric set algebra, but also from first-order logic:

> "The notion of a cylindric algebra can be considered as a common, algebraic abstraction from its two sources." [Tarski, 1985, p. v]

So consider first-order predicate logic with identity, together with Tarski's definitions of *model* and (semantic) *consequence*. Let ϕ be the denumerable set of all formulas of (first-order) predicate logic. Let Σ be a set of sentences — or a *theory* — from the language of predicate logic. We now define the notion of equivalence relative to a theory Σ as follows:

DEFINITION 14. ϕ and ψ are equivalent under set Σ, written $\phi \equiv_{sum} \psi$ iff $\phi \leftrightarrow \psi$ is a consequence of Σ.

[118] For example, consider Axiom 2 in the concrete setting of cylindrical set algebra. It is obvious that if we consider the set of points x and its cylindrification $c(x)$, then the points they share in common are just the points of x. And we have already seen the intuitive backing for Axiom 5.

It is easy to see that $\equiv_{\sum}$ is an equivalence relation on ϕ. So for each formula ϕ, there is an equivalence class $[\phi]$ of all formulas ψ such that $\phi \equiv_\Sigma \psi$, and the union of these mutually exclusive equivalence classes is ϕ (that is, $\equiv_\Sigma$ partitions ϕ). Let ϕ_Σ be the set of these equivalence classes.

Now we form a *cylindric algebra for a theory* Σ:

$$\mathfrak{A}_\Sigma = \langle \Sigma, +, ., -, 0, 1, c_\kappa, d_{\kappa\lambda}\rangle_{\kappa,\lambda<\omega}$$

where

$$\begin{aligned}
&[\phi] + [\psi] = [\phi \vee \psi]\\
&[\phi].[\psi] = [\phi \& \psi]\\
&-[\phi] = [\neg\phi]\\
&c_\kappa[\phi] = [\exists v_\kappa \phi]\\
&d_{\kappa\lambda} = [v_\kappa = v_\lambda]
\end{aligned}$$

and the distinguished elements 0 and 1 are $[v_\kappa \neq v_\kappa]$ and $[v_\kappa = v_\kappa]$ respectively. It is straightforward to check that this algebra satisfies all the axioms of ω-dimensional cylindric algebras, in particular Axioms 1–7.[119]

The significance of cylindrical algebras for the algebraic study of predicate logic "is rather obvious" [Tarski, 1971, p. 8]. To paraphrase what Henkin, Monk and Tarski go on to say, theories are basic entities in metalogical discussions just as algebraic structures are basic entities in algebraic research. By correlating the algebra $_\Sigma$ with any theory Σ, we have established a correspondence (which turns out to be one-one) between first-order theories and cylindric algebras of formulas. The basic metalogical problems for a fixed theory Σ are problems of the type: "Is a given sentence ϕ a consequence of the theory Σ?". *Each such problem clearly reduces to an algebraic problem concerning the associated algebra of formulas*: "Does the equation $[\phi]$=1 hold in the algebra $\mathfrak{A}_\Sigma$?"

For Tarski, then, cylindric algebras (like Boolean and relational algebras) were a tool for algebraizing logic. But Tarski also brought logic to bear on the algebras - like Boolean and relational algebras, cylindrical algebras are themselves a fit subject for metalogical investigation. Chapter 4 of [Tarski, 1985] applies logical notions to cylindrical algebras: the authors consider the model theory of cylindrical algebras in the first part of the chapter, and various decision problems in the second. Decidability and undecidability results are obtained for various cylindrical algebras.

In the Foreword to [Tarski, 1971], Tarski, Henkin and Monk trace the path that led to their collaboration:

> "The theory of cylindric algebras was founded by Tarski, in collaboration with his former students Louise H. Chin (Lim) and Frederick B. Thompson, during the period 1948-52. Soon thereafter Henkin became interested in the subject and began to work with Tarski on its further

[119]For example, Axiom 1 holds since the existential generalization of any member of $[v_\kappa \neq v_\kappa]$ is a member of $[v_\kappa \neq v_\kappa]$; Axiom 2 holds since $[\phi] = [\phi \& \exists x_\kappa \phi]$; Axiom 4 turns on the equivalence of $\exists v_\kappa \exists v_\lambda \phi$ and $\exists v_\lambda \exists v_\kappa \phi$; and Axiom 5 holds since $c_\kappa(d_{\kappa\lambda}) = [\exists v_\kappa(v_\kappa = v_\kappa)] = [v_\kappa = v_\kappa] = 1$.

> development. In 1961 they published a fairly extensive outline of their research, and the plan was first formulated to prepare a detailed monograph on the subject. Subsequently Monk's substantial contributions to the theory made a joining of efforts desirable, and thus the present team of authors was finally formed." [Tarski, 1971, p. 23]

Part 2 of *Cylindric Algebras* [1985m] was published in 1985; its first draft was two-thirds complete when Tarski died. The monograph *Cylindric set algebras* [Tarski, 1981], authored by Tarski, with Andreka, Henkin, Monk and Nemeti, was another major contribution to the theory. The theory of cylindric algebras has been an active field of research: for a survey of major results and further references, see Monk [Monk, 1986, esp. 903–5].

7 MORE ON GEOMETRY

7.1 The metamathematics of elementary geometry

We have already seen one of Tarski's most celebrated metamathematical result in geometry: the decidability of elementary (or first-order) Euclidean geometry, discovered in 1930 but not published in full until [Tarski, 1948a] (see section 5.2 above). Tarski returned to the metamathematics of geometry in "What is elementary geometry?" [Tarski, 1959]. Recall that in fn 18 of [Tarski, 1948a] Tarski briefly considers an axiomatic theory of Euclidean geometry — it is this theory of elementary geometry that Tarski investigates further in [Tarski, 1959].

In [Tarski, 1959], Tarski is fully explicit about the reference of the term 'elementary geometry'. Loosely it refers to the geometry which is based on Euclid's *Elements* and which forms the subject matter of secondary school geometry. But for metamathematical investigations, we need a more precise description; for one thing, we must specify the language in which the sentences of elementary geometry are formulated. As we saw in Section 5.2 above, the language of elementary geometry is the *first-order predicate calculus* — the first-order variables range over individuals (points), and there are no second-order or set-theoretical devices. The language contains two primitive predicates: a three-place predicate $Bxyz$ ('y is between x and z'), and the 4-place predicate $Dxyzw$ ('the distance from x to y is equal to the distance from z to w').

Tarski goes on to present a list of 12 axioms and one axiom schema. (See [Tarski, 1959, pp. 22–23]. This list supersedes the one in [Tarski, 1948a, fn. 18], which was found to contain superfluous axioms. Such refinements aside, Tarski's axiom system essentially dates back to his university lectures in the years 1926–27, as Tarski reports in [Tarski, 1967, p. 341, fn 34]. These axioms form the basis for elementary geometry, or G for short. They are the universal closures of the following:

A1. $Bxyx \rightarrow x = y$. (Identity Axiom for Betweenness)

A2. $Bxyu\&Byzu \rightarrow Bxyz$. (Transitivity Axiom for Betweenness)

A3. $Bxyz\&Bxyu\&x \neq y \rightarrow Bxzu \vee Bxuz$. (Connectivity Axiom for Betweenness)

A4. $Dxyyx$. (Reflexivity Axiom for Equidistance)

A5. $Dxyzz \rightarrow x = y$. (Identity Axiom for Equidistance)

A6. $Dxyzu\&Dxyvw \rightarrow Dzuvw$. (Transitivity Axiom for Equidistance)

A7. $Bxzw\&Byuz \rightarrow \exists v(Bxvy\&Bxuv)$. (Pasch's Axiom)[120]

A8. $Bxut\&Byuz\&x \neq u \rightarrow \exists v \exists w(Bxyv\&Bxzw\&Bvtw)$. (Euclid's Axiom)[121]

A9. $Dxyx'y'\&Dyzy'z'\&Dxux'u'\&Dyuy'u'\&Bxyz\&Bx'y'z'\&x \neq y \rightarrow Dzuz'u')$. (Five-Segment Axiom)

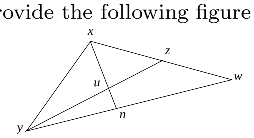
A10. $\exists z(Bxyz\&Dyzuv)$. (Axiom of Segment Construction)

A11. $\exists x \exists y \exists z(\neg Bxyz\&\neg Byzx\&\neg Bzxy)$. (Lower Dimension Axiom)

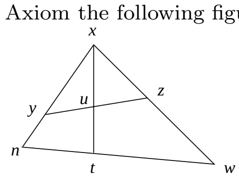
A12. $Dxuxv\&Dyuyv\&Dzuzv\&u \neq v \rightarrow BxyzvByzxvBzxy$. (Upper Dimension Axiom)

---

[120] In [79], Tarski and Sczcerba provide the following figure for Pasch's axiom:



[121] We can associate with Euclid's Axiom the following figure:



A8 says in essence: there is always a straight line through a point t inside an angle (here, angle yxz) that touches both sides of the angle (see [Tarski, 1983b, p. 13]).

Notice that the line vw meets the lines yv and zw in such a way that the angles yvt and twz add up to less than two right angles. Compare A8 with Euclid's original formulation of his Fifth Postulate:

> "That, if a straight line [vw] falling on to straight lines [yv and zw] make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet [at x] on that side on which are the angles less than two right angles."

A8 is provably equivalent to the Parallel Axiom (Axiom IV in Hilbert's [Hilbert, 1922]):

> If A is a line and a is a point that does not belong to A, then there is exactly one line which is parallel to A to which a belongs.

For a proof of the Parallel Axiom (Hilbert's IV) from A8, see [**?**, p. 123].

A13. (Elementary Continuity Axioms)

$$\exists z\forall x\forall y(\phi\&\psi \rightarrow Bzxy) \rightarrow \exists u\forall x\forall y(\phi\&\psi \rightarrow Bxuy)]$$

where ϕ and ψ are formulas, z and y do not occur free in ϕ, and z and x do not occur free in ψ.

Axiom 8 is crucial to the formation of Euclidean geometry. If we omit it, we have the theory of absolute geometry; and if we admit instead its negation, we have Bolyai–Lobachevsky geometry. The system comprising A1–A12 is worthy of study in its own right, since much of Euclidean geometry proceeds without A13.[122]

A13 is a first-order analogue of the second-order axiom of continuity. Tarski notes that standard axiomatizations of G include a continuity axiom that contains second-order variables $X, Y, \ldots$ ranging over sets of points. A typical formulation is:

$$\forall X\forall Y\{\exists z\forall x\forall y[x \in X\&y \in Y \rightarrow Bzxy] \rightarrow \exists u\forall x\forall y[x \in X\&y \in Y \rightarrow Bxuy]\}.$$

The first-order schema contained within A13 is obtained by replacing '$x \in X$' by an arbitrary first-order formula in which x occurs free, and '$y \in Y$' by an arbitrary first-order formula in which y occurs free. A13 can be viewed as a restriction of the second-order axiom to the definable sets.

Tarski goes on to address four fundamental metamathematical problems. The first is the *representation problem.* In general, the representation problem for a theory is the problem of characterizing all of the theory's models. In the case of G, the answer is given by the following theorem:

THEOREM 15 (Representation). *M is a model of G if and only if M is isomorphic with the Cartesian space over some real closed field F.*[123]

The second and third problems are the *completeness problem* and the *decidability problem.* Tarski provides positive answers to both: G is complete and consistent [Tarski, 1959, p. 25, theorem 2], and decidable [Tarski, 1959, p. 26, Theorem 3]. The fourth problem is the *problem of finite axiomatizability.* Here the question is: Can the axiom system be replaced by an equivalent finite system? The answer is negative: G is not finitely axiomatizable [Tarski, 1959, p. 26, Theorem 4].[124]

---

[122]For more in this vein, see [Szczerba, 1986, p. 909].

[123]This is [Tarski, 1959, p. 24, Theorem 1]. A model of G is a structure $\langle A, B, D\rangle$, where A is an arbitrary non-empty set, B is a 3-place relation and D is a 4-place relation, and where all the axioms of G hold if the variables are taken to range over elements of A, and the constants B and D are taken to refer to B and D respectively. Let F be a real closed field $\langle F, +, ., =\rangle$. (For a definition of *real closed field*, see for example [Chang and Keisler, 1973, p. 41].) Consider the set $A_F = F \times F$ of all ordered pairs $x = \langle x_1, x_2\rangle$ with x_1 and x_2 in F. We define the relations B_F and D_F as follows: $B_F(x, y, z)$ iff $(x_1 - y_1).(y_2 - z_2) = (x_2 - y_2).(y_1 - z_1)\&0 = (x_1 - y_1).(y_1 - z_1)\&0 = (x_2 - y_2).(y_2 - z_2).D_F(x, y, z, w)$ iff $(x_1 - y_1)^2 + (x_2 - y_2)^2 = (z_1 - w_1)^2 + (z_2 - w_2)^2$.

(Compare clauses (iv) and (v) in Section 5.1) The structure $C_2 = \langle A_F, B_F, D_F\rangle$ is called the *Cartesian space over F*. We now have all the ingredients of Theorem 9. Tarski provides an outline of a proof in [Tarski, 1959, pp. 24–25].

[124]Tarski provides an outline of the proof on pp. 26–7.

This negative result might be considered a drawback of Tarski's treatment of elementary geometry. Further, there are notions of textbook geometry that cannot be expressed in G — for example, the notions of the circumference and area of a circle, and the notion of a polygon with arbitrarily many vertices. In [Tarski, 1959], Tarski considers two other ways of interpreting the term 'elementary geometry' by way of alternative axiomatizations. The first is of greater expressive power than G, and can accommodate the geometric notions just mentioned; however, Tarski notes that this theory is undecidable, and that the other three metamathematical problems remain open. The second alternative is finitely axiomatizable; however, it is weaker than G and consequently captures fewer validities, and its decision problem is open. Tarski concludes:

> "The problem of deciding which of the various formal conceptions of elementary geometry is closer to the historical tradition and the colloquial usage of this notion seems to be rather hopeless and deprived of broader interest. The author feels that, among these various conceptions, the one embodied in [G] distinguishes itself by the simplicity and clarity of underlying intuitions and by the harmony and power of its metamathematical implications." (p. 31)

A recent, detailed study of Tarski's theory G can be found in [Tarski, 1983b]. This, and its bibliography give a good sense of Tarski's extensive influence on research in foundational geometry and metamathematics.

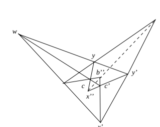
Tarski's interest in the metamathematics of geometry was not restricted to Euclidean geometry. As we noted in Section 5.2, Tarski proved that certain systems of non-Euclidean and projective geometry are decidable; and as we noted in Section 5.3, Tarski showed that abstract projective geometries are undecidable (see [Tarski, 1949e]). Tarski also investigated the metalogic of *affine geometry* in [Tarski, 1965a] and [Tarski, 1979], in collaboration with L.W. Szczerba. As Szczerba puts it, affine geometry is roughly the theory of the betweenness relation on the open and convex subsets of the Euclidean plane.[125] A precise axiomatic characterization of general affine geometry (GA) is presented in [Tarski, 1979]. The system of axioms for GA consists of A1, A2, A3, A7, A11, and A13, together with three further axioms specific to affine geometry.[126] Tarski and Szczerba solve

[125] See [**?**, p. 910].

[126] The three further axioms are as follows:

Extension Axiom The universal closure of $xBxyz\&x?y$.

Desargues' Axiom The universal closure of $Bwxx'\&Bwyy'\&Bwzz'\&Bxyz''\&x'y'z''\&Byzx''\&By'z'x''\&Bxzy''\&Bx'z'y''\&\neg Bwxy\&\neg Bxyw\&\neg Bywx\&\neg Bwyz\&\neg Byzw\&\neg Bzwy\&\neg Bwzx\&\neg Bzxw\&\neg Bxwz?Bx''y''z''$.



Upper Dimension Axiom The universal closure of $v[\{Byvz\&(BxvwvBvwx \vee Bwxv)\} \vee \{Bzvx\&(Byvw \vee Bvwy \vee Bwyv)\} \vee \{Bxvy\&(Bzvw \vee Bvwz \vee Bwzv)\}]$.

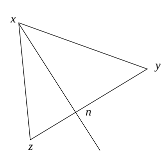
the representation problem for GA, and go on to show that GA is incomplete (in fact GA has $2^{\aleph}$ complete extensions), undecidable (in fact hereditarily undecidable),[127] and not finitely axiomatizable. In [Tarski, 1979] Tarski and Szczerba also investigate the metamathematical properties of various extension of GA. One of these — Euclidean affine geometry (EA) — is obtained by adding Euclid's Axiom A8 to the axioms for GA. EA is the main subject of [Tarski, 1965a], where it is shown, like GA, to be incomplete, (hereditarily) undecidable, and not finitely axiomatizable.

7.2 Primitive concepts of geometry

Another aspect of Tarski's metamathematical treatment of geometry is his work on primitive notions in geometry. We have seen that Tarski's formalization of Euclidean geometry requires just two primitive concepts — *betweenness* and *equidistance*. For Tarski, this economy is desirable and metamathematically significant. (In this respect, Tarski's formalization is superior to Hilbert's.) The question arises: How economical can we be? Can we find a single geometric notion in terms of which all the concepts of Euclidean geometry can be defined?

More narrowly, we can ask whether there is a relation between points that can serve this purpose. This is the leading question of Tarski's papers [Tarski, 1956a] (with E.W.Beth) and [Tarski, 1956b]. This search is constrained by a result established by Tarski and Lindenbaum: No *binary* relation between points can serve this purpose.[128] Still, Pieri showed that a certain three-place relation I comes close, where $I(x, y, z)$ holds iff the distance from x to y is the same as the distance from x to z (see [Pieri, 1908]). (In the 2-dimensional case, $I(x, y, z)$ holds iff x, y and z are the vertices of an isosceles triangle, with the line from y to z as the base; in the 1-dimensional case, x is the midpoint of the line from y to z.) The relation I *can* serve as the only primitive for n-dimensional Euclidean geometry, *provided that* $n \geq 2$. The equidistance relation denoted by D also comes close, since for $n \geq 2$ the betweenness relation B is definable in terms of D.

Can we do better than I and D? In [Tarski, 1956a], Tarski considers the 'equilaterality' relation $E : E(x, y, z)$ holds if the distances between x, y and z are equal. (In the 2-dimensional case, $E(x, y, z)$ holds if x, y and z are the vertices of an equilateral triangle.) Can E serve as the only primitive notion of Euclidean geometry? Tarski shows that the answer is affirmative for n-dimensional Euclidean



[127] A theory T is *hereditarily undecidable* if not only T but also every subtheory of T with the same constants is undecidable.

[128] See Lindenbaum and Tarski [1936c]. Indeed, Lindenbaum and Tarski announce that in elementary Euclidean geometry no 2-place relation can be defined "apart from the universal relation, the empty relation, identity and diversity." [Tarski, 1936c, p. 389]

geometry as long as $n \geq 3$; the answer is negative for $n = 1$ or $n = 2$. So E joins the ranks of other ternary relations, like I, which can serve as the sole primitive concept of n-dimensional Euclidean geometry for certain n, but not all n. In [Tarski, 1956b], Tarski considers in full generality the question of which ternary relations can serve as the sole primitive of Euclidean geometry for all n, and establishes a property that any such relation must have (and which E and I in particular fail to have, as Tarski shows).

Tarski's search for a single primitive relation in [Tarski, 1956a] and [Tarski, 1956b] was limited to relations between points. Much earlier, in the paper 'Foundations of the geometry of solids' [Tarski, 1929a], Tarski developed a geometry of solids, in which the primitive individuals are not points but spheres — or better, *balls*.[129] Tarski was responding to a challenge set by Lesniewski: to establish the foundations of a geometry of solids. In laying these foundations, Tarski works within the framework of Lesniewski's deductive theory of mereology (see Lesniewski, [1916] and [1927–31]). Just one 2-place relation — *the relation of part to whole* — is needed for Lesniewski's theory. Tarski showed in [Tarski, 1929a] that we can define all the concepts of 3-dimensional Euclidean geometry in terms of just the notions of *ball* and *the part-whole relation*.

The key to Tarski's demonstration is this: we may define the notion of *point* and the relation I above in terms of balls and the part-whole relation. The definitions are as follows:

DEFINITION 16. A *point* is the class of all balls which are concentric with a given ball. (See Definition 6, p. 27).[130]

DEFINITION 17. $I(x, y, z)$ iff there exists a ball W which belongs to the point x and which satisfies the following condition: no ball V belonging to the point y or the point z is a part of W or disjoint from W. (See Definition 7, p. 27).[131]

Given Pieri's result (above), it follows that every concept of 3-dimensional Euclidean geometry can be defined in terms of balls and the part-whole relation — and the result extends to n-dimensional Euclidean geometry for $n > -2$. As Szczcerba remarks (in [**?**]), the notion of a ball seems much more intuitive than the notion of a point — accordingly, this system of solid geometry is known as *Tarski's natural geometry*.

[129] In the original paper [Tarski, 1929a], Tarski used the word for 'sphere'. But in a footnote added later (fn. 1, p. 26), Tarski suggests that, since he is developing a geometry of solids and not a point geometry, it would be advisable to replace the word 'sphere' by 'ball'.

Tarski worked with open balls. S. Jaskowski — in the paper [Jaskowski, 1948] cited by Tarski in [Tarski, 1929a] (see fn. 2, p. 26) and by Szczerba in [Szczerba, 1986, p. 911] — simplified Tarski's geometry by modifying the definition of the relation 'A is concentric with B', and working with closed instead of open balls.

[130] Tarski observes (in [Tarski, 1929a, p. 27, fn. 1]) that balls are (first-order) individuals, while points are second-order objects. In point geometry, the reverse is true.

[131] Intuitively, the lines xy and xz are radii of the ball W.

7.3 The Banach-Tarski paradox

The famous Banach-Tarski paradox is a startling theorem proved by Tarski and Stefan Banach in their paper "Sur la decomposition des ensembles de points en parties respectivement congruentes" [Tarski, 1924d]. This was Tarski's second published paper in geometry, and like the first — [Tarski, 1924b] — it is concerned with the notion of *equivalence by finite decomposition.* According to the classical definition of this notion, two geometric figures are equivalent by finite decomposition if one can be decomposed (or "cut up") into finitely many figures that can be rearranged to form the other one. Tarski worked with a generalized set-theoretical version of this classical notion, which treats geometric figures such as polygons and polyhedra as sets of points. Tarski characterizes *equivalence by finite decomposition* along these lines:

> Two sets of points A and B (in particular, two polygons or polyhedra) are said to be *equivalent by finite decomposition* iff A can be decomposed into finitely many disjoint sets of points $A_1, A_2, \ldots A_n$, B can be decomposed into the same number of disjoint sets of points $B_1, B_2, \ldots B_n$, and A_i and B_i are congruent, for $1 \leq i \leq n$.[132]

In [Tarski, 1924b] Tarski considers two theorems that were known to hold for the classical notion of equivalence by finite decomposition:

Theorem 1 Two arbitrary polygons, where one is contained in the other, are never equivalent by finite decomposition.

Theorem 2 Two polygons are equivalent by finite composition if and only if they are equal in area.

Tarski shows in [Tarski, 1924b] that these theorems also hold for the set-theoretical version of equivalence by finite decomposition.

These results about polygons appear natural and intuitive, and one might expect them to hold for polyhedra too. But they do not. At the very end of [Tarski, 1924b], Tarski remarks that, if we take equivalence by finite decomposition in the set-theoretical sense, it can be shown that Theorems 1 and 2 do not extend to polyhedra. Tarski announces the following theorem, proved in the yet-to be-published paper [Tarski, 1924d]:

> Two arbitrary polyhedra (of the same volume or not) are equivalent by finite decomposition. [Tarski, 1924b, p. 64]

As Tarski cautiously puts it, the result "seems perhaps paradoxical" (p. 64). This theorem is the Banach-Tarski paradox.[133]

The theorem is more precisely stated in [Tarski, 1924d] as follows:

[132] See Tarski [1924b, p. 64].

[133] Previous paradoxical decompositions were constructed by Vitali [Vitali, 1905] and Hausdorff [Hausdorff, 1914]. Hausdorff's work was a major influence on Banach and Tarski's paper. For more historical background, see [Wagon, 1985, Chapters1–3].

> If A and B are any two sets of points in 3-dimensional Euclidean space, each being bounded and with a non-empty interior, then A and B are equivalent by finite decomposition.[134]

In particular, then, two spheres of different radii are equivalent by finite decomposition (in sharp contrast to the 2-dimensional case of circles). This encourages dramatic statements of the paradox:

> "It is possible to cut up a pea into finitely many pieces that can be rearranged to form a ball the size of the sun!"[135]

or

> "A ball, which has a definite volume, may be taken apart into finitely many pieces that may be rearranged via rotations of R^3 to form two, or even 1,000,000 balls, each identical to the original one."[136]

The theorem extends to n-dimensional Euclidean space, for $n \geq 3$.[137]

Jan Mycielski calls the Banach-Tarski result "the most surprising result of theoretical mathematics".[138] One kind of response to the paradox — the most common — has been to take the result at face value, and accept it as a demonstration that mathematics can sometimes fail in the most radical way to match up with physical reality.[139] A different kind of response — little seen nowadays–- has been to treat it as the conclusion of a genuine paradox, a conclusion that is so counterintuitive that we are forced us to review the argument that produced it. For some the culprit is the Axiom of Choice, which is indispensable to the proof.[140]

---

[134]See [Tarski, 1924d, p. 140, Theorem 24]. In the statement of the theorem, Tarski uses the term "ensembles-frontieres". This term is defined in fn 1, p. 133: it applies to exactly those sets of points that have no interior point.

[135][Wagon, 1985, p. xiii].

[136][Wagon, 1985, p. 28].

[137]See [Tarski, 1924d, p. 121], where the theorem is stated in its fully general form, and also p. 137, where Tarski indicates how to proceed with the proof for $n3$. Moore writes that Banach and Tarski independently discovered the Banach–Tarski paradox, and agreed to write a joint paper. In a personal communication to Moore, Tarski reported that the general result was due to him (see [Moore, 1982, p. 285, fn. 3].

[138]See Mycielski's Foreword to [Wagon, 1985].

[139]The mathematical lesson to learn is that the Banac-h-Tarski theorem

> "precludes the existence of finitely-additive, congruent-invariant measures over all bounded subsets in [three-dimensional space] R^3."

(Mycielski, in the Foreword to [Wagon, 1985])

[140]For a detailed discussion of the role of the Axiom of Choice in the proof of the Banach-Tarski theorem, see [Wagon, 1985, Chapter 13]. Wagon's book is a recent mathematical work motivated by the Banach–Tarski result. It contains a number of open problems, old and new. One of these Wagon calls "Tarski's Circle-Squaring Problem" (posed in 1925): Is a circle (with interior points) equivalent by finite decomposition to a square of the same area? (See [Wagon, 1985, p. 101ff] for a discussion of this and related problems.)

8 SET THEORY

Tarski's first published paper [Tarski, 1921] and his last published monograph [Tarski, 1987a] were both works in set theory. Tarski's early interest in set theory was inspired by Sierpinski, who, together with Janiszewski and Mazurkiewicz, started up the Warsaw School of Mathematics in 1919. In the 1920s and 1930s, Tarski wrote a number of papers on topics in general set theory. The subject of [Tarski, 1921] was the notion of a well-ordered set; subsequent papers dealt with such topics as the theory of finite sets, cardinal arithmetic, and the axiom of choice and its equivalents. As set theory developed and grew more sophisticated and specialized, so did Tarski's work. Tarski was a seminal figure in the development of the theory of large cardinals — he produced highly influential work on large cardinals from 1930 to the 1960s. Tarski also investigated set theory from the algebraic perspective, a perspective which - as we have seen - he so often brought to logic. In *Cardinal Algebras* [Tarski, 1949f] and *Ordinal Algebras* [Tarski, 1956c] Tarski investigated cardinal and ordinal addition within the framework of abstract algebraic systems. And in *A Formalization of Set Theory without variables* [Tarski, 1987a], Tarski and Steven Givant developed set theory within the framework of abstract relation algebras (see section 6.2 above for more on relation algebras).

8.1 General set theory

In his paper [Sierpinski, 1918], Sierpinski investigated in detail the role that the Axiom of Choice played in set theory and analysis, and threw out a challenge to mathematicians: determine the deductive relations between the Axiom of Choice and other propositions.[141] Tarski responded to this challenge in [Tarski, 1924A]. At the outset of this paper, Tarski lists seven propositions that are equivalent to the Axiom of Choice. Each of these propositions is a statement of cardinal arithmetic, where m, n, p, q are infinite cardinal numbers:

I. $m.n = m + n$

II. $m = m^2$

III. If $m^2 = n^2$ then $m = n$.

IV. If $m < n$ and $p < q$ then $m + p < n + q$.

V. If $m < n$ and $p < q$ then $m.p < n.q$.

VI. If $m + p < n + p$ then $m < n$.

VII. If $m.p < n.p$ then $m < n$.[142]

[141] See Moore [Moore, 1982, Chapter 4.1] for more on Sierpinski' work and influence.

[142] Tarski's proof of the equivalences is based on Zermelo's system of axioms (omitting Choice, of course) together with two additional axioms which introduce the notion of cardinal number:
1. Every set has a cardinal number.
2. Two sets have the same cardinal number if and only there is a 1-1 correspondence between them.

no entry for Robinson 1951 or Grzegorczyk 1951 in bib.

In their joint paper [Tarski, 1926a], Tarski and Lindenbaum added to this list seven more propositions, all of them statements from the theory of cardinal numbers (see [Tarski, 1926a, pp. 185–6, Theorem 82]).[143] Thus, Tarski's early work in [**?**] and [Tarski, 1926a] was a major contribution to our understanding of the relation between cardinal arithmetic and the Axiom of Choice, and in particular to the project of finding so-called 'cardinal equivalents' of Choice.[144]

Another topic of Tarski's early work in set theory was the theory of finite sets. In [Tarski, 1924c] Tarski systematically constructed a theory of finite sets on the basis of Zermelo's first five axioms (Extensionality, Elementary sets, Separation, Power set, Union) — Zermelo's remaining two axioms, the Axiom of Choice and the Axiom of Infinity, were excluded. Tarski remarks that no previous attempt to construct such a theory had been completely successful; for example, Dedekind's treatment in [Dedekind, 1888] did not have a solid axiomatic foundation, and Russell and Whitehead's account (in [Russell and Whitehead, 1910–13]) is specific to their theory of types (see [Tarski, 1924c, pp. 67–8].)

Tarski's own starting point was a new definition of finite set:

> A set A is *finite* iff every non-empty set K of subsets of A has a $\subseteq$-least element B (that is, an element B such that no member of K is a proper subset of B). (See [Tarski, 1924c, p. 71, Definition 3].)

From this starting point, Tarski went on to prove a large number of theorems which established all the fundamental properties of finite sets with which he was familiar. Tarski's definition differs from the standard arithmetic definition, according to which a set is finite iff it is either empty or there exists a 1-1 correspondence between it and the set $\{1, 2, \ldots, n\}$ for some finite n. In contrast to this usual definition, Tarski's definition is independent of the concept of finite number. It also differs from Dedekind's, according to which a set is finite if there is no 1-1 correspondence between it and any of its proper subsets.

The question naturally arises as to the equivalence of these various definitions, and this question is also investigated in [Tarski, 1924c]. Tarski shows that his definition is equivalent to the usual arithmetical one (see [Tarski, 1924c, p. 80]), thereby showing that the theory of finite sets can be constructed from the usual definition without the need for the Axiom of Choice. He also shows if a set is finite in his sense, then it is finite in Dedekind's sense. For Tarski, the reverse direction was an open problem: without Choice, it was not known how to establish the equivalence of his definition and Dedekind's (see p. 95). In an appendix to [Tarski, 1924c], Tarski states a number of open problems of this kind, generated

[143] Tarski and Lindenbaum's [Tarski, 1926a] was an extensive summary of their recent results in set theory, presented without proofs.

[144] In [Rubin and Rubin, 1963], Herman and Jean Rubin present over one hundred equivalents of the Axiom of Choice — and one category of equivalents they call 'cardinal equivalents'. For more on Tarski's work on cardinal equivalents, see [Moore, 1982, pp. 213–219, Section 4.3]. Other work on cardinal arithmetic published by Tarski in the 1920s includes [Tarski, 1925; Tarski, 1926a; Tarski, 1929c]. Tarski later obtained further cardinal equivalents of the Axiom of Choice in [Tarski, 1938a], a paper that investigated inaccessible cardinals (see subsection 2 below).

by five alternative definitions of finite set that Tarksi lists (on p. 115). It turned out that none of these definitions — among them Tarski's and Dedekind's — were equivalent in the absence of the Axiom of Choice. So while Tarski's definition permitted the construction of the theory of finite sets without Choice, a number of alternative definitions (including Dedekind's) did not.[145]

8.2 Large cardinals

Tarski initiated the systematic study of large cardinals, and he and his school was largely responsible for its continuing development. An appropriate starting point is Tarski and Sierpinski's joint paper [Tarski, 1930a] in which they define the notion of an *inaccessible cardinal*. Their definition runs as follows:

> A cardinal number m is inaccessible iff it is not the product of a fewer number of cardinals of lesser power. (See Definition 1, p. 289).

The intuitive idea is that an inaccessible cardinal cannot be obtained from below. As Tarski and Sierpinski remark, $_0$ is clearly an inaccessible cardinal: it cannot be obtained from the product of finitely many finite cardinals. In current terminology, Tarski and Sierpinski's definition characterizes the class of *strongly inaccessible cardinals*.

In [Tarski, 1938a], Tarski provided alternative characterizations of the strongly inaccessible cardinals. He also characterized the weakly inaccessible cardinals, as follows:

> m is a *weakly inaccessible cardinal* iff $m = \aleph_\alpha$ where α is a regular limit ordinal.[146]

Tarski shows, with the help of the Axiom of Choice, that if a cardinal is strongly inaccessible, it is weakly inaccessible (pp. 360–361). He also proves the reverse direction, but, as he makes fully explicit, he does so only with the help of the Generalized Continuum Hypothesis (p. 362).

Tarski took inaccessible cardinals to be natueal objects of set theory:

> "...the belief in the existence of inaccessible cardinals $> \omega$ (and even of arbitrarily large cardinals of this kind) seems to be a natural consequence of basic intuitions underlying the "naive" set theory and referring to what can be called "Cantor's absolute"". [Tarski, 1962, p. 124]

[145] For more on Tarski on finite sets, see Moore [Moore, 1982, pp. 209–213, section 4.2].

[146] For the notion of a regular ordinal, we need some prior notions. We call two partially ordered sets X and Y *similar* (in symbols, $X \sim Y$) if there exists an order-preserving 1-1 correspondence between them. Next we need the notions of an initial ordinal and cofinality.

Definition α is an *initial ordinal* iff $\mathrm{Ord}\alpha \& \forall\beta < \alpha(\neg\alpha \sim \beta)$.

Definitions α is *cofinal* with β iff $\alpha \leq \beta \& \exists f(f : \alpha \to \beta) \& \beta = \mathrm{Urange}(f)$. The *cofinality* $\mathrm{cf}(\beta)$ of β is $\cap\{\alpha | \alpha$ is cofinal with $\beta\}$.

We are now in a position to define the notion of a regular ordinal:

Definition For an initial ordinal κ, κ is *regular* iff $\mathrm{cf}(\kappa) = \kappa$. (See [Tarski, 1938a, p. 360].)

And Zermelo's set theory cannot accommodate these intuitions:

> "on the basis of the usual axioms of Zermelo the existence of such numbers, apart from the two smallest of them, 2 and $\aleph_0$, cannot be established at all." [Tarski, 1939, p. 557]

Accordingly, Tarski introduced for the first time an axiom that guaranteed the existence of large cardinals.[147] Remarking that the axiomatization of increasingly large segments of "Cantor's absolute" is regarded by many as one of main aims of research in the foundations of set theory, Tarski continues:

> "Those who share this attitude are always ready to accept new "construction principles", new axioms securing the existence of new classes of "large" cardinals (provided they appear to be consistent with old axioms)." [Tarski, 1939, p. 557]

Tarski's *Axiom of Inaccessible Cardinals* was first presented in [Tarski, 1938a, p. 375]. Tarski's formulation was based on a technical characterization of the strongly inaccessible cardinals (drawn from [Tarski, 1938a, Theorem 21]). In [1939], Tarski replaced it by a more natural version:

For every set N there exists a set M with the following properties:

(i) N is equipollent to a subset of M.[148]

(ii) the set of subsets of M which are not equipollent to M (i.e. the set $\{x|x \subseteq M \& x < M\}$)[149] is equipollent to M.

(iii) there is no set P such that P's power set is equipollent to M. (See [Tarski, 1939, p. 558].)

Either version of the Axiom of Inaccessible Sets says, in effect, that for any set there is a larger set whose cardinal number is strongly inaccessible. The axiom "assures the existence of inaccessible numbers as large as we please" [Tarski, 1939, p. 558]. Moroever, the axiom has "great deductive power":

> "If it is included in Zermelo's or Zermelo-Fraenkel's axiom-system this axiom brings with it a great simplification and reduction of the system; and, be it noted, the axiom of choice then becomes a provable theorem." [Tarski, 1939, p. 557]

In a later series of papers [1943; 1957a; 1958a; 1958b; 1961a; 1962; 1964], Tarski investigated a family of problems concerning inaccessible cardinals. These problems share the same structure. The form of each problem is this: to determine

[147] Compare the introduction of the Axiom of Replacement to guarantee the existence of certain intuitive sets — for example, the denumerable set $\{Z_0, PZ_0, PPZ_0, \dots\}$, where Z_0 is the set of positive integers and P is the power set operation (see [Skolem, 1922], in [van Heijenoort, 1967, pp. 296–7]).

[148] Two sets are equipollent iff there is a 1-1 correspondence between them.

[149] $A < B$ iff there is a set C such that A is equipollent to C and $C \subset B$.

the class of all the (infinite) cardinals that have a given property P, where it is known that $\aleph_0$does not have P, and that the accessible cardinals do. The problem reduces to this: which of the inaccessible cardinals, if any, have P?

For example, one problem examined in [Tarski, 1962] is the *compactness problem* for predicate logics with infinitely long formulas.[150] Let L_ω be ordinary predicate logic. For any regular cardinal α let L_α be the logical system that differs from L_ω in having α different variables, infinitary operations analogous to ordinary disjunction and conjunction, and new versions of universal and existential quantification adjusted to the infinitary setting.[151] The compactness problem for logics L_α is the problem of determining those cardinals for which the following compactness theorem holds:

> If S is any set of sentences in L_α, and if every subset of S with power $< \alpha$ has a model, then S also has a model.

That is, a cardinal α has the property P in the present case if the compactness theorem holds for α. If so, α is a *compact cardinal*; otherwise it is *incompact.* Now ω is compact, while all the accessible cardinals are incompact. What about the inaccessible cardinals? Here the problem about inaccessible cardinals has a metamathematical or metalogical setting. Tarski's student Hanf proved that a very large class of inaccessible cardinals are incompact.[152] Consider an arrangement of all the inaccessible cardinals in a transfinite increasing sequence $\theta_0, \theta_1, \ldots \theta_i$, $\ldots$ for ordinals $0, 1, \ldots i, \ldots$. Hanf proved that every cardinal α of the form $\alpha = \theta_i$ with $0 < i < \alpha$ is incompact.

In [Tarski, 1962] Tarski used this metamathematical result of Hanf's to establish certain other properties of this large class of inaccessible cardinals. Tarski still regarded the metamathematical approach as powerful and intuitive, even for results in pure set theory.[153] Referring back to [Tarski, 1962], Tarski and Keisler later wrote:

> "The results we have mentioned concerning large classes of inaccessible cardinals were originally obtained with the essential help of metamathematical (model-theoretic) methods. These methods still provide the intuitively and deductively simplest approach of the topic in its full generality. In our opinion this circumstance provides new and significant evidence of the power of metamathematics as a tool in purely mathematical research, and at the same time does not detract in the least from the value of the results obtained." [Tarski, 1964, p. 130]

Nevertheless in [Tarski, 1964] Tarski and Keisler undertake a purely mathematical (and very extensive) treatment of the topics discussed in [Tarski, 1962], motivated

[150] These logics are briefly discussed by Tarski in [1958a].
[151] For a full description see, for example, [Tarski, 1962, pp. 116–7].
[152] [Hanf, 1963–4].
[153] Work of Tarski's that adopts the metamathematical approach to the present family of problems are [Tarski, 1958a] (with D. Scott), [Tarski, 1961a; Tarski, 1961b]. See also [Keisler, 1962].

to some extent "by some (irrational) inclination toward puritanism in methods" [Tarski, 1964, p. 130].[154]

At the end of [1962], Tarski observes that the problems he has treated remain open only for the smallest inaccessible cardinal that is not incompact, and beyond. If there is no such cardinal, the problems are fully disposed of. And Tarski reports that

> "we do not know any example of a cardinal $\alpha > \omega$ which would possess a "constructive characterization" (in some very general and rather loose sense of this term) and of which we could not prove that it is incompact ..." [Tarski, 1962, p. 119]

So should we add an axiom to the effect that there are no such cardinals? Only if we are willing to compromise the study of "Cantor's absolute", which Tarski certainly was not. So we *should* be prepared to accept large cardinal axioms, but *not* prepared to accept

> "any axioms precluding the existence of [large] cardinals — unless this is done on a strictly temporary basis, for the restricted purpose of facilitating the metamathematical discussion of some axiomatic systems of set theory." [Tarski, 1962, p. 124]

8.3 The algebraization of set theory

We turn first to Tarski's book *Cardinal Algebras* [Tarski, 1949f], published in 1949 but conceived some twenty years earlier. It deals with cardinal arithmetic, in particular cardinal addition. Tarski distinguishes two types of results in cardinal arithmetic, those that rely on the Axiom of Choice, and those, more constructive in nature, that mostly do not. In [Tarski, 1949f], Tarski investigates results of the second type that concern cardinal addition. Tarski establishes that all of these results can be derived within the framework of so-called *cardinal algebras*, which satisfy just a small number of laws or postulates. Moreover, these laws apply not only to cardinal arithmetic, but also to various other mathematical systems. So cardinal algebras form a simple framework within which many results from many different systems are forthcoming.

A cardinal algebra is an ordered triple $\underline{A} = \langle A, +, \Sigma \rangle$ satisfying certain postulates. A is a set of arbitrary elements, $+$ is a 2-place operation, and Σ an operation on infinite sequences (intuitively, Σ is infinite addition). Postulates I and II are respectively the closure of A under $+$ and the closure of A under Σ. Postulate III is the associative postulate: $\sum_{i<\infty} a_i = a_0 + \sum_{i<\infty} a_{i+1}$. Postulate IV is the commutative-associative postulate: $\sum_{i<\infty}(a_i + b_i) = \sum_{i<\infty} a_i + \sum_{i<\infty} b_i$. Postulate V is the zero postulate: there is an element z in A such that $a + z = z + a = a$

[154] Tarski [1962] and Tarski and Keisler [1964] inspired a great deal of research on *measurable cardinals* and *ultraproducts* (see for example [Chang and Keisler, 1973, Ch. 4.2], and for some historical notes, pp. 520–1).

for every $a \in A$. Postulates VI and VII are the ones characteristic of cardinal algebras. Postulate VI is the refinement postulate: if $a + b = \sum_{i<\infty} c_i$ then there are a_i and b_i such that $a = \sum_{i<\infty} a_i, b = \sum_{i<\infty} b_i$, and $c_n = a_n + b_n$. Postulate VII is the remainder (or infinite chain) postulate: if $a_n = b_n + a_{n+1}$, for $n = 0, 1, 2, \ldots$ then there is an element $c \in A$ such that $a_n = c + b_{n+i}$ for $n = 0, 1, 2, \ldots$. (See [Tarski, 1949f, pp. 3–4].) It is straightforward to check that the cardinal numbers with the operations of binary and infinite addition form a cardinal algebra. The first part of [Tarski, 1949f] develops the arithmetic of cardinal algebras. (The second part extends these results to a wider class of algebraic systems, and examines general methods of constructing cardinal algebras; the third part investigates the connections between cardinal algebras and related algebraic systems.) Tarski remarks:

> "The idea of an algebraic treatment of [cardinal arithmetic] and certain aspects and implications of the algebraic development seem to be essentially new." (Preface to [Tarski, 1949f, p. xii])

Tarski's monograph *Ordinal Algebras* [Tarski, 1956c] does for ordinal addition what [Tarski, 1949a] did for cardinal addition. As Tarski puts it:

> "The method applied in this monograph in the development and presentation of the theory of ordinal addition is the abstract algebraic method which was applied for analogous purposes in an earlier work of the author, [Tarski, 1949f]." (introduction to [Tarski, 1956c, p. 2].)

Tarski remarks that the algebras themselves show some analogies — we find closure, associative, and remainder postulates. The main difference, according to Tarski, is that in ordinal algebras the operation $+$ is not commutative; also, the refinement postulate is replaced by a much stronger statement, the directed refinement postulate. (For a full specification of ordinal algebras, and more on these analogies and differences, see [**?**, Chapter 1]). As in the case of cardinal addition, Please clarify this reference. Tarski's novel algebraic approach to ordinal addition generated, from a very small base, a large number of results that are not limited to ordinal arithmetic.[155]

Tarski's final monograph *A Formalization of Set Theory without variables* (with Steven Givant) [Tarski, 1987a] is a wholesale algebraization of set theory. In the preface, Tarski and Givant announce:

> "In this work we shall show that set theory and number theory can be developed within the framework of a new, different, and very simple formalism, L^X." (p. xi)

The formalism L^X contains no variables, quantifiers, or sentential connectives. The vocabulary of L^X contains seven symbols: two atomic predicates, $\mathring{\imath}$ (denoting the

[155] Both [Jonsson, 1986] and [Levy, 1988] note that Tarski's work on cardinal and ordinal algebras has been somewhat neglected. Jonsson suggests that Tarski's thoroughness may in part explain this, "for he presented the mathematical community with two highly polished and elaborate creations, but with relatively few open problems" [Jonsson, 1986, p. 885].

identity relation between individuals) and E (denoting the membership relation); four operators, $+$ (Boolean addition), $^-$ (Boolean complement), $\breve{}$ (conversion — see 4.2 above), and $\odot$ (relative product — called ';' in 4.2 above); and a second identity predicate $=$ (denoting the identity relation between relations). Compound predicates are formed from the atomic predicates 1 and E using the operators. All mathematical statements in L^X are of the form '$A = B$', where A and B are arbitrary predicates.

The deductive apparatus of L^X is based upon ten logical axiom schemata (see [Tarski, 1987a, p. 46]):

(I) $A + B = B + A$

(II) $A + (B + C) = (A + B) + C$

(III) $(A^- + B)^- + (A^- + B^-)^- = A$

(IV) $A \odot (B \odot C) = (A \odot B) \odot C$

(V) $(A + B) \odot C = A \odot C + B \odot C$

(VI) $A \odot \mathring{\imath} = A$

(VII) $A^{\smile\smile} = A$

(VIII) $(A + B)^{\smile} = A^{\smile} + B^{\smile}$

(IX) $(A \odot B)^{\smile} = B^{\smile} \odot A^{\smile}$

(X) $A^{\smile} \odot (A \odot B)^- + B^- = B^-$.

These axioms are the analogues of the equational postulates for abstract relation algebras essentially given in Chin–Tarski [Tarski, 1951a, on p. 344]. There is just one rule of inference, the rule of replacing equals by equals, familiar from high school algebra. Tarski and Givant write that

> "L^X proves adequate for the formalization of practically all known systems of set theory, and hence for the development of all of classical mathematics." (preface to [Tarski, 1985, p. xii])

Thus set theory can be based entirely on a system which is free of variables and in which the only formulas are sentences which take the form of equations. (See [Tarski, 1987a, section 4.6] for the formalizability of systems of set theory in L^X.) The system L^X is a direct development of Tarski's work on the calculus of relations and relation algebras (see 4.2 above):

> "Roughly speaking, the formalism L^X that is the central focus of this work is obtained from Tarski's equational formalization of the calculus of relations by introducing the constant E and deleting all variables." (preface to [Tarski, 1985, p. xvii])

Givant reports that Tarski died shortly after the manuscript for [Tarski, 1987a] was completed.

9 CONCLUDING REMARKS

Tarski wrote twenty monographs and more than one hundred articles, and yet he published next to nothing about his philosophical views. Even in his work best-known to philosophers — his work on the concept of truth — Tarski's philosophical views are hard to pin down, as we have seen.

In a rare expression of his philosophical standpoint regarding the foundations of mathematics, Tarski once wrote:

> "...I may mention that my personal attitude towards this question agrees in principle with that which has found emphatic expression in the writings of S. Lesniewski and which I would call *intuitionistic formalism*." [Tarski, 1930d, p. 62]

But in a footnote added later to this paper (*ibid.*), Tarski characteristically drew back from such an unequivocal philosophical statement ("This last statement expresses the views of the author at the time when this article was originally published and does not adequately reflect his present attitude"). Mostowski (Tarski's first PhD student) reports that Tarski was influenced early on by Kotarbinski's "reism", a nominalist doctrine, and would express sympathy with nominalism in oral discussions. But as Mostowski points out, this seems to conflict with Tarski's constant use of abstract and general notions (consider, for example, Tarski's wholesale acceptance of set theory).[156]

We can find some brief philosophical remarks in two other pieces of Tarski's, both published posthumously. In a letter to Morton White, written in September 1944, Tarski followed what he took to be the Millian view, that logical and mathematical truths are, like empirical truths, the results of accumulated experience. Tarski goes on to say:

> "I think I am ready to reject certain logical premisses (axioms) of our science in exactly the same circumstances in which I am ready to reject empirical premisses (e.g. physical hypotheses)..." [Tarski, 1987b, p. 31]

This was a view that was later to find expression in [White, 1950] and [Quine, 1951], both of whom acknowledge their debt to Tarski. The second posthumous piece is the paper "What are logical notions?" [Tarski, 1986]. Tarski makes it clear that he is not after any normative answer to this question, nor any 'platonic' answer:

> "...people speak of catching the proper, true meaning of a notion, something independent of actual usage, and independent of any normative proposals, something like the platonic idea behind the notion. This last approach is so foreign and strange to me that I shall simply ignore it, for I cannot say anything intelligent on such matters." (p. 145)

[156] [Mostowski, 1967, p. 81].

Instead, Tarski aims to capture a possible use of the term 'logical notion', one that he thinks is in agreement with at least one actual usage of the term. As with the notions of definability, truth, and logical consequence, Tarski seeks to capture common usage as far as consistency and rigor allow.[157]

It is clear that Tarski's most significant philosophical contributions flow from his logical and metamathematical work, and not from any overtly philosophical writings. This is no accident: Tarski held firmly to the view that logical and mathematical investigations should proceed unhampered by any particular philosophical perspective. For example, as Mostowski emphasizes, Tarski's unrestricted use of set theory gave him a mathematical reach that was beyond the adherents to Hilbert's formalism or Brouwer's intuitionism. Tarski thought of himself first and foremost as a mathematician and a logician, and as a philosopher only in some secondary sense: "perhaps a philosopher of a sort" [Tarski, 1944, p. 693].

Tarski's philosophical contributions are to be found in his work on metamathematics, semantics and logic. In the course of more than sixty years of active research, Tarski articulated concepts and established theorems that have become standard in modern logic; he ushered in new fields; he helped to construct the very framework of modern logic. And through generation after generation of his students, Tarski's influence is still felt. In their review of [**?**], Pogorzelski and Surma write:

Which Tarski:24 is this?

> "...the thing that we find most striking is that there ...is hardly another scientist in the history of the exact sciences whose part in the construction of notions for a large domain of science was as powerful as the contribution of Tarski to the creation of conceptual apparatus for logic, metalogic, and even metamathematics. In fact, the conceptual structure of these disciplines is due to Tarski."[158]

Tarski offered an additional way to measure "the value of a man's work" [Tarski, 1944, p. 693]:

> "It seems to me that there is a special domain of very profound and strong human needs related to scientific research, which are similar in many ways to aesthetic and perhaps religious needs. And it also seems to me that the satisfaction of these needs should be considered an important task of research. Hence, I believe, the question of the value of any research cannot be adequately answered without taking into account the intellectual satisfaction which the results of that research bring to those who understand it and care for it. It may be unpopular and out-of-date to say - but I do not think that a scientific result which gives us a better understanding of the world and makes it more harmonious in our eyes should be held in lower esteem than, say, an

[157] Tarski's suggestion concerning logical notions is couched in geometric terms of invariance under transformations.

[158] This is the conclusion of [Pogorzelski and Surma, 1969].

invention which reduces the cost of paving roads, or improves household plumbing." [Tarski, 1944, p. 694]

ACKNOWLEDGEMENTS

I am very grateful to Michael Resnik for his valuable comments on an entire draft of this chapter. I have also benefited from some excellent secondary sources. In particular, I have been helped by the following: the survey/expository papers on Tarski's work published by the *Journal of Symbolic Logic*, comprising Vaught [1986] (on model theory), Jonsson [1986] (on general algebra), McNulty [1986] (on undecidable theories), Monk [1986] (on algebraic logic), Szczerba [1986] (on geometry), Levy [1988] (on set theory), Van den Dries [1988] (on Tarski's elimination theory for real closed fields), Doner and Hodges [1988] (on decidable theories), Blok and Pigozzi [1988] (on general metamathematics), Etchemendy [1988a] (on truth and logical consequence), and Suppes [1988] (on the philosophical implications of Tarski's work); Corcoran [1983a]; Moore [1982]; Mostowski [1967].

BIBLIOGRAPHY

WORKS BY TARSKI

[Tarski, 1921] 'Przyczynek do aksjomatyki zbioru dobrze uporzadkowanego' ('A contribution to the axiomatics of well-ordered sets'), Przeglad Filozoficzny, vol.24 , pp. 85–94, 1921.

[Tarski, 1923] 'O wyrazie pierwotnym logistyki', Doctoral dissertation, published in Przeglad Filozoficzny, vol.26, pp. 68–89, 1923. An English translation appears in [Tarski, 1983a, pp. 1–23], under the title 'On the primitive term of logistic'.

[Tarski, 1924A] 'Sur quelques theormes qui equivalent a l'axiome du choix', Fundamenta Matematyczno-fizyczny, vol.2, pp. 47–60, 1924.

[Tarski, 1924b] 'O rownowaznosci wielokatow' ('On the equivalence of polygons'), Przeglad Matematyczno-fizyczny, vol. 2, pp. 47–60, 1924.

[Tarski, 1924c] 'Sur les ensembles finis', Fundamenta Mathematicae, vol. 6, pp. 45–95, 1924.

[Tarski, 1924d] with S. Banach. 'Sur la decomposition des ensembles de points en parties respectivement congruentes', Fundamenta Mathematicae, vol. 6, pp. 45–95, 1924.

[Tarski, 1925] 'Quelques theormes sur les alephs', Fundamenta Mathematicae, vol. 7, pp. 1–14, 1925.

[Tarski, 1926a] with A. Lindenbaum. 'Communication sur les rcherches de la théorie des ensembles', Comptes Rendus des Seances de la Société des Sciences et des Lettres de Varsovie, Classe III, vol. 19, pp. 299–330, 1926.

[Tarski, 1927a] 'Remarque concernant l'arithmetique des nombres cardinaux' (abstract), Annales de la Socit Polonaise de Mathematique, vol. 5, p. 101, 1927.

[Tarski, 1927b] with A. Lindenbaum. 'Sur l'independance des notions primitives dans les systemes mathématiques', Abstract, Annales de la Socit Polonaise de Mathématique, vol. 5, pp. 111–113, 1927.

[Tarski, 1929a] 'Les fondements de la geometrie des corps', Ksiega Pamiatkowa Pierwszego Polskiego Zjazdu Matematycznego (Supplement to Rocznik Polskiego Towarzystwa Matematycznego), Krakow, pp. 29–33, 1929.

[Tarski, 1929b] 'Remarques sur les notions fondamentales de la méthodologie des mathématiques', Annales de la Société Polonaise de Mathématique, vol. 7, pp. 270–2, 1929.

[Tarski, 1929c] 'Geschichtliche Entwicklung und gegenwartiger Zustand der Gleichmachtigkeitstheorie und der Kardinalzhlarithmetik', Ksiega Pamiatkowa Pierwszego Polskiego Zjazdu Matematycznego (Supplement to Rocznik Polskiego Towarzystwa Matematycznego), Krakow, pp. 48–54, 1929.

[Tarski, 1930a] with W. Sierpinski. 'Sur une propriete caracteristique des nombres inaccessibles', Fundamenta Mathematicae vol. 15, pp. 292, 300, 1930.

[Tarski, 1930b] 'Uber einige fundamentale Begriffe der Metamathematik', Comptes Rendus des sances de la Socit des Sciences et des Lettres de Varsovie, vol.23, cl.iii, pp. 22–29, 1930. An English translation appears in [Tarski, 1983a, pp. 30–37], under the title 'On some fundamental concepts of metamathematics'. (Page references are to [Tarski, 1983a].)

[Tarski, 1930c] with J. Łukasiewicz. 'Untersuchungen uber den Aussagenkalkul', Comptes Rendus des sances de la Société des Sciences et des Lettres de Varsovie, vol.23, c.iii, pp. 30–50, 1930. An English translation appears in [Tarski, 1983a, pp. 38–59], under the title 'Investigations into the sentential calculus'. (Page references are to [Tarski, 1983a].)

[Tarski, 1930d] 'Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften. I', Monatshefte fur Mathematik und Physik, vol. 37, pp. 361–404, 1930. An English translation appears in [Tarski, 1983a, pp. 60–109], under the title 'Fundamental concepts of the methodology of the deductive sciences'. (Page references are to [Tarski, 1983a].)

[Tarski, 1931a] 'Sur les ensembles definissables de nombres reels. I.', Fundamenta Mathematicae, vol. 17, pp. 210–239, 1931. An English translation appears in [Tarski, 1983a, pp. 110–142], under the title 'On definable sets of real numbers'. (Page references are to [Tarski, 1983a].)

[Tarski, 1931b] with C. Kuratowski. 'Les operations logiques et les ensembles projectifs' (with C. Kuratowski), Fundamenta Mathematicae, vol.17, pp. 240–8, 1931. An Eglish translation appears in [Tarski, 1983a, pp. 143–151], under the title 'Logical operations and projective sets'. (Page references are to [Tarski, 1983a].)

[Tarski, 1933a] 'Pojecie prawdy w jezykach nauk dedukcyjnych' ('The concept of truth in the languages of deductive sciences'), Travaux de la Socit des Sciences et des Lettres de Varsovie, Classe III Sciences Mathematiques et Physiques, no.34, Warsaw, 1933. A German translation of this was published under the title 'Der Wahrheitsbegriff in den formalisierten Sprachen', in Studia Philosophica, vol. 1, pp. 261–405, 1935. An English translation of the German appears under the title 'The concept of truth in formalized languages' in [Tarski, 1983a, pp. 152–278]. (Page references are to [Tarski, 1983a].)

[Tarski, 1933b] 'Einige Betrachtungen uber die Begriffe ?- Widerspruchsfreiheit und der ?-Vollstandigkeit', Monatshefte fur Mathematik und Physik, vol. 40, pp. 97–112, 1933. An English translation appears in [Tarski, 1983a, pp. 179–295], under the title 'Some observations on the concepts of ?-consistency and ? -completeness'. (Page references are to [Tarski, 1983a].)

[Tarski, 1934] 'Z badan metodologicznych nad definiowalnosca terminow' in Przeglad Filozoficzny, vol. 37, pp. 438–460, 1934. A German translation appears under the title 'Einige methodologische Untersuchungen uber die Definierbarkeit der Begriffe', in Erkenntnis, vol. 5, pp. 80–100, 1935. An English translation of the German, supplemented by some passages translated from the Polish original, appears under the title 'Some methodological investigations on the definability of concepts' in [Tarski, 1983a, pp. 296–319]. (Page references are to [Tarski, 1983a].)

[Tarski, 1935] 'Zur Grundlegung der Booleschen Algebra I', in Fundamenta Mathematica, vol. 24, pp. 177–98, 1935. An English translation appears under the title 'On the foundations of Boolean algebra' in [Tarski, 1983a, pp. 320–341].

[Tarski, 1935+1936] 'Grundzuge des Systemenkalkul, Erster Teil', Fundamenta Mathematicae, vol. 25, pp. 503–26, 1935, and 'Grundzuge des Systemenkalkul, Zweiter Teil', ibid., vol. 26, pp. 283–301, 1936. The two parts are translated as one paper in [Tarski, 1983a, pp. 342–383], under the title 'Foundations of the Calculus of Systems'. (Page references are to [Tarski, 1983a].)

[Tarski, 1936a] 'O ugruntowaniu naukowej semantyki' ('The establishment of scientific semantics'), Przeglad Filozoficzny, vol. 39, pp. 50–57, 1936. A German translation was published under the title 'Grundlegung der wissenschaftlichen Semantik', in Actes du Congres International de Philosophie Scientifique, vol. 3, pp. 1–8. Actualites Scientifiques et Industrielle, vol.390, Hermann & Cie, Paris, 1936. An English translation of the German appears under the title 'The establishment of scientific semantics' in [Tarski, 1983a, pp. 401–408]. (Page references are to [Tarski, 1983a].)

[Tarski, 1936b] 'O pojciu wynikania logicznego', Przeglad Filozoficzny, vol. 39, pp. 58–68, 1936. A German translation of this was published under the title 'Uber den Begriff der logischen Folgerung', in Actes du Congres International de Philosophie Scientifique, vol. 7, pp. 1–11, Actualites Scientifiques et Industrielle, vol.394, Hermann & Cie, Paris, 1936. An English translation of the German appears under the title 'On the concept of logical consequence' in [Tarski, 1983a, pp. 409–420]. (Page references are to [Tarski, 1983a].)

[Tarski, 1936c] with A. Lindenbaum. 'Uber die Beschranktheit der Ausdrucksmittel deduktiver Theorien' (with A. Lindenbaum), Ergebnisse eines Mathematischen Kolloquiums, vol. 7, pp. 15–22, 1936. An English translation appears under the title 'On the limitations of the means of expression of deductive theories' in [Tarski, 1983a, pp. 384–392]. (Page references are to [Tarski, 1983a].)

[Tarski, 1936d] O logice matematycznej i metodzie dedukcyjnej (On mathematical logic and the deductive method), Bibljoteczka Matematyczna, vol. 3–5, 167pp. Ksiaznica-Atlas, Lwow and Warsaw, 1936.

[Tarski, 1938a] 'Uber unerreichbare Kardinalzahlen', Fundamenta Mathematicae, vol. 30, pp. 68–89, 1938.

[Tarski, 1938b] 'Der Aussagenkalkul und die Topologie', Fundamenta Mathematicae, vol. 31, pp. 103–34, 1938. An English translation appears in [Tarski, 1983a, pp. 421–454], under the title 'Sentential Calculus and Topology'. (Page references are to [Tarski, 1983a].)

[Tarski, 1939] 'On well-ordered subsets of any set', Fundamenta Mathematicae, vol. 32, pp. 176–183, 1939.

[Tarski, 1941a] Introduction to logic and to the methodology of deductive sciences, Oxford University Press, Oxford and New York, 1941.

[Tarski, 1941b] 'On the calculus of relations', Journal of Symbolic Logic, vol. 6, pp. 73–89, 1941.

[Tarski, 1943] with P. Erdos. 'On families of mutually exclusive sets', Annals of Mathematics, ser. 2, vol. 44, pp. 315–329, 1943.

[Tarski, 1944] 'The semantic conception of truth and the foundations of semantics', Philosophy and Phenomenological Research, vol. 4, pp. 341–375, 1944.

[Tarski, 1948a] A Decision Method for Elementary Algebra and Geometry (prepared for publication by J. C. C. McKinsey), US Air Force Project RAND, R-109, the RAND Corporation, Santa Monica, California, 1948. (A second, revised edition was published by the University of California Press, Berkeley and Los Angeles, California 1951.)

[Tarski, 1948b] 'A problem concerning the notion of definability', Journal of Symbolic Logic, 13, 79–104, 1948.

[Tarski, 1949a] 'Arithmetical classes and types of Boolean algebras' (abstract), Bulletin of the American Mathematical Society, vol. 55, pp. 64, 1192, 1949.

[Tarski, 1949b] 'On essential undecidability', Journal of Symbolic Logic, vol. 14, pp. 75–76, 1949.

[Tarski, 1949c] 'Undecidability in the arithmetic of integers and in the theory of rings', Journal of Symbolic Logic, vol. 14, p. 76, 1949.

[Tarski, 1949d] 'Undecidability of group theory' (abstract), Journal of Symbolic Logic, vol. 14, pp. 76–77, 1949.

[Tarski, 1949e] 'Undecidability of the theories of lattices and projective geometries' (abstract), Journal of Symbolic Logic, vol. 14, pp. 77–78, 1949.

[Tarski, 1949f] Cardinal Algebras, Oxford University Press, Oxford and New York, 1949.

[Tarski, 1951a] with L. H. Chin. 'Distributive and modular laws in the arithmetic of relation algebras', University of California Publications in Mathematics, new series, vol. 1, no. 9, pp. 341–384, 1951.

[Tarski, 1951b] with B. Jonsson. 'Boolean algebras with operators. Part I', American Journal of Mathematics, vol. 73, pp. 891–939, 1951.

[Tarski, 1952] with B. Jonsson. 'Boolean algebras with operators. Part II', American Journal of Mathematics, vol. 74, pp. 127–162, 1952.

[Tarski, 1953] with A. Mostowski and R. M. Robinson. Undecidable theories, North-Holland Publishing Company, Amsterdam, 1953.

[Tarski, 1956a] with E. W. Beth. 'Equilaterality as the only primitive notion of Euclidean geometry', Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings, Series A, vol. 59 (=Indagationes Mathematicae, vol. 18), pp. 462–467, 1956.

[Tarski, 1956b] 'A general theorem concerning primitive notions of Euclidean geometry', Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings, Series A, vol. 59 (= Indagationes Mathematicae, vol. 18), pp. 468–474, 1956.

[Tarski, 1956c] Ordinal Algebras, North-Holland Publishing Company, Amsterdam 1956.

[Tarski, 195d] Logic, Semantics, Metamathematics. Papers from 1923 to 1938, Clarendon Press, Oxford, 1956. (Translated by J. H. Woodger.)

[Tarski, 1957a] with E. C. Smith. 'Higher degrees of distributivity and completeness in Boolean algebras', Transactions of the American Mathematical Society, vol. 84, pp. 230–257, 1957.

[Tarski, 1957b] with R. L. Vaught. 'Arithmetical extensions of relational systems', Compositio Mathematica, vol. 13, pp. 81–102, 1957.

[Tarski, 1958a] with D. S. Scott. 'The sentential calculus with infinitely long expressions, Colloquium Mathematicum, vol. 6, pp. 165–170, 1958.

[Tarski, 1958b] 'Remarks on predicate logic with infinitely long expressions' Colloquium Mathematicum, Vol. 6, pp. 171–176, 1958.

[Tarski, 1959] 'What is elementary geometry?', The axiomatic method, with special reference to geometry and physics, L. Henkin *et al.*, editors, pp. 15–29. North-Holland Publishing Company, Amsterdam, 1959.

[Tarski, 1961a] with P. Erdos. 'On some problems involving inaccessible cardinals', Essays on the foundations of mathematics, Y. Bar-Hillel *et al.*, editors, pp. 50–82. Magnus Press, Jerusalem, 1961.

[Tarski, 1961b] 'Representable Boolean algebras and infinitary logics' (abstract), Notices of the American Mathematical Society, vol. 8, pp. 154–155, 1961.

[Tarski, 1961c] 'A model-theoretic result concerning infinitary logics' (abstract), Notices of the American Mathematical Society, vol. 8, p. 260, 1961.

[Tarski, 1962] 'Some problems and results relevant to the foundations of set theory', Logic, methodology, and philosophy of science. Proceedings of the 1960 International Congress, E. Nagel *et al.*, editors, pp. 125–135. Stanford University Press, Stanford, California, 1962.

[Tarski, 1964] wtih H. J. Keisler. 'From accessible to inaccessible cardinals. Results holding for all accessible cardinal numbers and the problem of their extension to inaccessible ones', Fundamenta Mathematicae, vol. 53, pp. 225–308, 1964.

[Tarski, 1965a] with L. W. Szczerba. 'Metamathematical properties of some affine geometries', Logic, methodology, and philosophy of science. Proceedings of the 1964 International Congress, Y. Bar-Hillel, editor, pp. 166–178. North-Holland Publishing Company, Amsterdam, 1965.

[Tarski, 1965b] with J. W. Addison and L. Henkin, eds. The theory of models, North-Holland, Amsterdam, 1965.

[Tarski, 1967] The completeness of elementary algebra and geometry, Institut Blaise Pascal, Paris, 1967.

[Tarski, 1969] 'Truth and Proof', Scientific American, vol. 220, no. 6, pp. 63–77, 1969.

[Tarski, 1971] with L. Henkin and J. D. Monk. Cylindric Algebras, Part I, North-Holland Publishing Company, Amsterdam, 1971.

[Tarski, 1978] with J. Doner and A. Mostowski. 'The elementary theory of well-ordering - A metamathematical study', Logic Colloquium '77, A. Macintyre *et al.*, editors, pp. 1–54. North-Holland Publishing Company, Amsterdam.

[Tarski, 1979] with L. W. Szczerba. 'Metamathematical discussion of some affine geometries', Fundamenta Mathematicae, vol. 104, pp. 155–92, 1979.

[Tarski, 1981] with H. Andreka, L. Henkin, J. D. Monk and I. Nemeti. Cylindric set algebras, Lecture Notes in Mathematics, vol. 883, Springer-Verlag, Berlin and New York 1981.

[Tarski, 1983a] Logic, semantics, metamathematics. Papers from 1923 to 1938 (translations by J. H. Woodger). Second, revised edition of [Tarski, 1956d], with editor's introduction and an analytic index (J. Corcoran, editor), Hackett Publishing Company, Indianapolis, Indiana, 1983.

[Tarski, 1983b] with W. Schwabhauser and W. Szmielew. Metamathematische Methoden in der Geometrie, Hochschultext, Springer- Verlag, Berlin and New York, 1983.

[Tarski, 1985] with L. Henkin and J. D. Monk. Cylindric Algebras, Part II, North-Holland Publishing Company, Amsterdam 1985.

[Tarski, 1986] 'What are logical notions?' John Corcoran, editor, History and Philosophy of Logic, 7, pp. 143–154, 1986.

[Tarski, 1987a] with S. Givant. A formalization of set theory without variables, Colloquium Publications, vol. 41, American Mathematical Society, Providence, Rhode Island, 1987.

[Tarski, 1987b] 'A philosophical letter of Alfred Tarski', Journal of Philosophy, pp. 28–32, 1987.

WORKS BY OTHER AUTHORS

[Ajdukiewicz, 1921] K. Ajdukiewicz. Z metodologji nauk dedukcyjnych (From the methodology of the deductive sciences), Lwow, 1921.

[Aristotle, 1941] Aristotle, Posterior Analytics, pp. 108–186 of [McKeon, 1941].

[Barwise and Etchemendy, 1987] J. Barwise and J. Etchemendy. The Liar, Oxford University Press, 1987.

[Beth, 1953] E. W. Beth. 'On Padoa's method in the theory of definition', Indagationes Mathematicae, xv, 330–9, 1953.

[Birkhoff, 1948] G. Birkhoff. Lattice Theory, American Mathematical Society Colloquium publications, vol. 25, revised edition, New York, 1948.

[Blackburn and Simmons, 1999] S. Blackburn and K. Simmons. Truth, in the series Oxford Readings in Philosophy, Oxford University Press, 1999.

[Blok and Pigozzi, 1988] W. J. Blok and D. Pigozzi. 'Alfred Tarski's work on general metamathematics', Journal of Symbolic Logic, vol. 53, no. 1, pp. 36–50, 1988.

[Bolzano, 1837] B. Bolzano. Wissenschaftslehre (Theory of Science), Sulzbach: Seidel, 4 vols., 1837.

[Burge, 1979] T. Burge. 'Semantical Paradox', Journal of Philosophy 76, pp. 169–98, 1979; reprinted in [Martin, 1984, pp. 83–117]. Page references are to [Martin, 1984].

[Carnap, 1934] R. Carnap. Logische Syntax der Sprache, Wien, 1934.

[Carnap, 1935] R. Carnap. 'Ein Gultigkeitskriterium fur die Satze der klassischen Mathematik', Monatshefte fur Mathematik und Physik, xlii, 163–90, 1935.

[Chang and Keisler, 1973] C. C. Chang and H. J. Keisler. Model Theory, North-Holland Publishing Company, Amsterdam, 1973.

[David, 1994] M. David. Correspondence and Disquotation, Oxford University Press, 1994.

[Davidson, 1967] D. Davidson. 'Truth and Meaning', Synthese 17, 304–23, 1967.

[Davidson, 1996] D. Davidson. 'On the Folly of Trying to Define Truth', Journal of Philosophy 93, 263–278, 1996.

[Dedekind, 1888] R. Dedekind. Was sind und was sollen die Zahlen? Braunschweig: Vieweg, 1888.

[Doner and Hodges, 1988] J. Doner and W. Hodges. 'Alfred Tarski and decidable theories', Journal of Symbolic Logic, vol. 53, no. 1, 20–35, 1988.

[Edwards, 1967] P. Edwards, ed. The Encyclopedia of Philosophy, Macmillan Publishing Co., New York, 1967.

[Etchemendy, 1983] J. Etchemendy. 'The doctrine of logic as form', Linguistics and Philosophy, vol. 6, 319–334, 1983.

[Etchemendy, 1988a] J. Etchemendy. 'Tarski on truth and logical consequence', Journal of Symbolic Logic, vol. 53, no. 1, 51–79, 1988.

[Etchemendy, 1988b] J. Etchemendy. 'Models, semantics and logical truth', Linguistics and Philosophy, vol. 11, 91–106, 1988.

[Etchemendy, 1990] J. Etchemendy. The Concept of Logical Consequence, Harvard University Press, Cambridge 1990.

[Field, 1986] H. Field. 'The Deflationary Conception of Truth', in Fact, Science, and Morality: Essays on A. J. Ayer's 'Language, Truth and Logic', edited by G. MacDonald and C. Wright, pp. 55–117. Oxford: Basil Blackwell, 1986.

[Field, 1994] H. Field. 'Deflationist views of meaning and content', Mind 103, 249–84, 1994.

[Gaifman, 1992] H. Gaifman. 'Pointers to Truth', Journal of Philsophy 89, 223–61, 1992.

[Givant, 1986] S. Givant. 'Bibliography of Alfred Tarski', Journal of Symbolic Logic, vol. 51, no. 4, pp. 913–941, 1986.

[Glanzberg, 2001] M. Glanzberg. 'The Liar in context', Philosophical Studies 103, 217–251, 2001.

[Gödel, 1930a] K. Gödel. 'Die Vollstandigkeit der Axiome des logischen Funktionenkalkuls', Monatshefte fur Mathematik und Physik, 37, 349–360; translated in [van Heijenoort, 1967, pp. 582–591].

[Gödel, 1930b] K. Gödel. 'Einige metamathematische Resultate uber Entscheidungsdefinitheit und Widerspruchsfreiheit', Akademie der Wissenschaften in Wien, Mathematisch- naturwissenschaftliche Klasse, Akademischer Anzeiger, lxvii, 214–215, 1930.

[Gödel, 1931] K. Gödel. 'Uber formal unentscheidbare Satze der Principia Mathematica und verwandter Systeme I', Monatshefte fur Mathematik und Physik, xxxviiim 173–98, 1931.

[Gómez-Torrente, 1996] M. Gómez-Torrente. 'Tarski on logical consequence', Notre Dame Journal of Formal Logic, vol. 37, no. 1, 125–151, 1996.

[Hanf, 1963–4] W. P. Hanf. 'Incompactness in languages with infinitely long expressions', Fundamenta Mathematicae, vol. 53, 309–324, 1963–4.

[Hausdorff, 1914] F. Hausdorff. 'Bemerkung uber den Inhalt von Punktmengen' Mathematische Annalen 75, 428–433, 1914.

[Herbrand, 1928] J. Herbrand. 'Sur la theorie de la demonstration', Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences, vol. 186, 1274–1276, 1928.

[Herzberger, 1981] H. Herzberger. 'New Paradoxes for Old', Proceedings of the Aristotelian Society 81, pp. 109–23, 1981.

[Hilbert, 1922] D. Hilbert. Grundlagen der Geometrie, 3rd ed., Teubner, Leipzig, 1922.

[Hilbert and Bernays, 1939] D, Hilbert and P. Bernays. Grundlagen der Arithmetik, vol. 1, Berlin, 1934; vol.2, Berlin, 1939.

[Hodges, 1986a] W. Hodges. 'Alfred Tarski', Journal of Symbolic Logic, vol. 51, number 4, p. 866, 1986.

[Hook, 1950] S. Hook, ed. John Dewey: Philosopher of Science and Freedom, New York, Dial Press, 1950.

[Huntington, 1904] E. V. Huntington. 'Sets of independent pRostulates for the algebra of logic', Transactions of the American Mathematical Society, v, 288–309, 1904.

[Jaskowski, 1948] S. Jaskowski. 'Une modification des definitions fondamentales de la geometrie des corps de M. A.Tarski', Annales de la Societe Polonaise de Mathematique, xxi, 298–301, 1948.

[Jonsson, 1986] B. Jonsson. 'The contributions of Alfred Tarski to general algebra', The Journal of Symbolic Logic, vol. 51, no. 4, 883–889, 1986.

[Keisler, 1962] H. J. Keisler. ''Some applications of the theory of models to set theory', Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress, pp. 80–86. Stanford, 1962.

[Koons, 1992] R. C. Koons. Paradoxes of Belief and Strategic Rationality, Cambridge University Press, Cambridge, 1992.

[König, 1905] J. König. 'On the foundations of set theory and the continuum problem', in [van Heijenoort, 1967, pp. 145–149].

[Langford, 1927] C. H. Langford. 'Some theorems on deducibility', Annals of Mathematics, ser. 2, vol. 28, 16–40, 1927.

[Langford, 1928] C. H. Langford. 'Theorems on deducibility (second paper)', Annals of Mathematics, ser. 2, vol. 28, 459–471, 1928.

[Leeds, 1978] S. Leeds. 'Theories of Reference and Truth', Erkenntnis 13, 111–129, 1978.

[Lesniewski, 1916] S. Lesniewski. 'Podstawy ogolnej teorii mnogosci I' (Foundations of general set theory I) Moscow, 1916.

[Lesniewski, 1929] S. Lesniewski. 'Grundzuge eines neuen Systems der Grundlagen der Mathematik', Secs 1–11, Fundamenta Mathematicae, xiv, 1–81, 1929.

[Lesniewski, 1927–31] S. Lesniewski. 'O podstawach matematyki' (On the foundations of mathematics), Przeglad Filozoficzny, xxx, 164–206, 1927; ibid., ibid., xxxi, 261–91, 1928; ibid., xxxii, 60–105, 1929; ibid., xxxiii, 77–105, 1930; ibid., xxxiv, 142–70, 1931.

[Levy, 1988] A. Levy. 'Alfred Tarski's work in set theory', Journal of Symbolic Logic, vol. 53, no. 1, 2–6, 1988.

[Lowenheim, 1915] L. Lowenheim. 'Uber Moglichkeiten im Relativkalkul', Mathematische Annalen, vol. 76, 447–470, 1915.

[Lyndon, 1950] R. Lyndon. 'The representation of relational algebras', Annals of Mathematics, ser. 2, vol. 51, 707–729, 1950.

[Lyndon, 1956] R. Lyndon. 'The representation of relational algebras. II', Annals of Mathematics, ser. 2, vol. 63, 294–307, 1956.

[Mal'cev, 1936] A. Mal'cev. 'Untersuchungen aus dem Gebiete der mathematischen Logik', Matematiceskii Sbornik, vol. 1 (43), 323–336, 1936.

[Martin, 1984] R. L. Martin, ed. Recent Essays on Truth and the Liar Paradox, Oxford University Press, 1984.

[Martin, 1976] R. L. Martin. 'Are Natural Languages Universal?', Synthese 32, 271–91, 1976.

[McKeon, 1941] R. McKeon, ed. The Basic Works of Aristotle, Random House, New York, 1941.

[McNulty, 1986] G. McNulty. 'Alfred Tarski and undecidable theories', The Journal of Symbolic Logic, vol. 51, no. 4, 890–898, 1986.

[Moore, 1982] G. Moore, Zermelo's Axiom of Choice: its origins, development and influence, Springer Verlag, New York, 1982.

[Monk, 1986] J. D. Monk. 'The Contributions of Alfred Tarski to Algebraic Logic', The Journal of Symbolic Logic, vol. 51, no. 4, 899–906, 1986.

[Mostowski, 1967] A. Mostowski. 'Alfred Tarski', in [Edwards, 1967, Vol. 8, pp. 77–81].

[Padoa, 1900] A. Padoa. 'Logical introduction to any deductive theory', 1900. Translated in [van Heijenoort, 1967, pp. 118–123].

[Parsons, 1974] C. Parsons. 'The Liar Paradox', Journal of Philosophical Logic 3, 381–412, 1974. Reprinted in [Martin, 1984]. Page references are to Martin 1984.

[Pasch, 1982] M. Pasch. Vorlesungen uber neuere Geometrie (Leipzig: Teubner) 1882.

[Peano, 1895–1908] G. Peano. Formulaire de mathematiques, 5 vols. Turin: Bocca, 1895-1908.

[Pieri, 1908] M. Pieri. 'La geometria elementare istituita sulle nozioni di 'punto' e 'sfera'. Memorie di Matematica e di Fisica della Societa Italiana delle Scienze, ser. 3, 15, 345–450, 1908.

[Pogorzelski and Surma, 1969] W. Pogorzelski and S. Surma. Review of Tarski 1956, Journal of Symbolic Logic 34, 99–106, 1969.

[Presburger, 1930] M. Presburger. 'Uber die Vollstandigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt', Comptesrendus du I congres des mathematiciens des pays slaves, Warszawa, 1929, Ksiaznica Atlas T.N.S.W. Warsaw, pp, 92–101, 395, 1930.

[Quine, 1951] W. V. O. Quine. 'Two Dogmas of Empiricism' Philosophical Review, LX, 1, 20–43, 1951.

[Quine, 1970] W. V. O. Quine. Philosophy of Logic, Englewood Cliffs, N.J.: Prentice-Hall, 1970.

[Resnik, 1990] M. Resnik. 'Immanent Truth', Mind, vol. 99, no. 395, 405–424, 1990.

[Richard, 1905] J. Richard. 'Les principes des mathmatiques et le problme des ensembles', in Revue gnrale des sciences pures et appliquées 16, 541, 1905. Also in Acta Mathematica 30, 295–6, 1906. An English translation appears in [van Heijenoort, 1967, pp. 143–44].

[Rubin and Rubin, 1963] H. Rubin and J. Rubin. Equivalents of the Axiom of Choice, Amsterdam, North-Holland, 1963.

[Russell, 1908] B. Russell. 'Mathematical logic as based on the theory of types', American Journal of Mathematics 30m 222–262, 1908. Reprinted in [van Heijenoort, 1967, pp. 150–182].

[Russell and Whitehead, 1910–13] B. Russell and A. N. Whitehead. Principia Mathematica, Cambridge, Cambridge University Press: vol.1, 1910; vol. 2, 1912; vol. 3, 1913.

[Schröder, 1890–1905] E. Schröder. Vorlesungen uber die Algebra der Logik (exakte Logik) (Leipzig; i, 1890; ii, part 1, 1891; ii, part 2, 1905; iii, part 1, 1895).

[Schröder, 1895] E. Schröder. Algebra und Logik der Relative, 1895, vol.iii, part 1 of [Schroder, 1890–1905].

[Sierpinski, 1918] W. Sierpinski. 'L'axiome de M. Zermelo et son role dans la theorie des ensembles et l'analyse', Bulletin de l'Academie des Sciences de Cracovie, Classe des Sciences Math., Serie A, 97–152, 1918.

[Simmons, 1993] K. Simmons. Universality and the Liar, Cambridge University Press, Cambridge and New York, 1993.

[Skolem, 1919] T. Skolem. 'Untersuchungen uber die Axiome des Klassenkalkuls und uber Produktations - und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen', Skrifter utgit av Videnskapsselskapet i Kristiania, I. klasse 1919, no. 3, Oslo, 1919

[Skolem, 1922] T. Skolem. 'Some remarks on axiomatized set theory', 1922. Reprinted in [van Heijenoort, 1967, pp. 290–301].

[Soames, 1999] S. Soames. Understanding Truth, Oxford University Press, Oxford and New York, 1999.

[Stone, 1936] M. H. Stone. 'The theory of representation for Boolean algebras', Transactions of the American Mathematical Society, xi, 37–111, 1936.

[Stone, 1937] M. H. Stone. 'Applications of the theory of Boolean rings to general topology', Fundamenta Mathematicae, xxix, 223–303, 1937.

[Stone, 1937–38] M. H. Stone. 'Topological representations of distributive lattices and Brouwerian logics', Casopis pro Pestovani Matematiky a Fysiky, lxvii, 1–25, 1937–8.

[Suppes, 1988] P. Suppes. 'Philosophical implications of Tarski's work', Journal of Symbolic Logic, vol. 53, no. 1, 80–91, 1988.

[Szczerba, 1986] L. W. Szczerba. 'Tarski and geometry', The Journal of Symbolic Logic, vol. 51, no. 4, 907–912, 1986.

[van den Dries, 1988] L. van den Dries. 'Alfred Tarski's elimination theory for real closed fields', Journal of Symbolic Logic, vol. 53, no. 1, 7–19, 1988.

[van Heijenoort, 1967] J. van Heijenoort, ed. From Frege to Gödel: a source book in mathematical logic, Harvard University Press, 1967.

[Vaught, 1965] R. L. Vaught. 'A Lowenheim-Skolem theorem for cardinals far apart', Proceedings of the 1963 Berkeley symposium on the theory of models, J. Addison *et al.*, editors, pp. 390–401. North-Holland, Amsterdam, 1965.

[Vaught, 1974] R. L. Vaught. 'Model theory before 1945', Proceedings of the Tarski symposium, Proceedings of Symposia in Pure Mathematics, vol. 25, pp. 153–172. American Mathematical Society, Providence, Rhode Island, 1974.

[Vaught, 1986] R. L. Vaught. 'Alfred Tarski's work in model theory', The Journal of Symbolic Logic, vol. 51, no. 4, 869–882, 1986.

[Veblen and Young, 1910] O. Veblen and J. W. Young. Projective Geometry, Ginn, Boston, Massachusetts (2 vols., first, 1910; second 1918); reprint, Blaisdell, New York, 1938.

[Vitali, 1905] G. Vitali. Sul problema della mesura dei gruppi di punti di una retta, Bologna, Tip. Gamberini e Parmeggiani, 1905.

[Wagon, 1985] S. Wagon. The Banach-Tarski Paradox, Cambridge University Press, 1985.

[White, 1950] M. White. 'The Analytic and the Synthetic: An Untenable Dualism'. In [Hook, 1950, pp. 316–30].