Towards an abstract theory of definitions

Edoardo Rivello

University of Torino, Torino, Italy
edoardo.rivello@unito.it

Contents

1 Introduction

In [8], Tarski[1] laid down a series of axioms aiming to characterise a primitive notion of *consequence* and showed that, by means of this notion only, other metalogical concepts could be defined; among them the concepts of *theory*, *logical equivalence*, *consistency*, *completeness*. Stemming from Tarski's seminal work, the study of abstract consequence relations — motivated by their logical interpretations — has developed into a mature and active field of research (see, for instance, Martin & Pollard's book [4]).

Later, in [9], Tarski initiated a study of the notion of *definition* showing important analogies with the abstract approach taken in [8]:

> In the methodology of the deductive sciences two groups of concepts occur which, although rather remote from one another in content, nevertheless show considerable analogies, if we consider their role in the construction of deductive theories, as well as the inner relations between concepts within each of the two groups themselves. To the first group belong such concepts as 'axiom', 'derivable sentence' (or 'theorem'), 'rule of inference', 'proof', to the second — 'primitive (undefined) concept' (or 'primitive term'), 'definable concept', 'rule of definition', 'definition'. A far-reaching parallelism can be established between the concepts of the two groups: the primitive concepts correspond to the axioms, the defined concepts to the derivable sentences, the process and rules of definition to the process and rules of proof. [10, p. 296].

According to Pogorzelski & Surma review of an English translation[2] of [9],

> "Paper X belongs to those papers of Tarski which have organized a certain branch of metalogic and established some of its fundamental notions. It deals with syntactic definability of terms, and together with some earlier results concerning the concept of semantic definability [...] it establishes the foundations of the theory of definability of terms. X is a natural extension of Tarski's papers [...] on the notion of consequence, since it establishes for terms a number of notions analogous to those

[1] An earlier exposition, without proofs, of the results collected in [8] appeared in [7].

[2] The English translation of [9] by J. H. Woodger is included in [10] and referred to as "paper X" of the collection.

which are fundamental for propositional expressions. We find in X the definitions of such notions as: equivalence of sets of terms, closure of a set of terms (an analogue of the concept of the system), system of primitive terms (an analogue of the concept of the set of axioms), the notion of independent and complete set of terms (the last notion relativized to an arbitrary but fixed set of sentences) [...] Results contained in X virtually exhausted syntactic problems of definability". [5, p. 104].

Even though the intimate connections of [9] with [8] are emphasised by Tarski himself, the study on definitions is developed in [9] within a framework which is considerably less 'abstract' than that assumed in [8] in order to study the notion of consequence: The former presupposes an internal structure of the sentences that distinguish variables and extra-logical terms, as the minimum setting for speaking about "the definability and the mutual independence of concepts" [10, p. 296]. By contrast, in the present paper my aim is that of establishing the fundamentals of an abstract theory of definitions in the same framework of Tarski's [8], by taking an arbitrary notion of consequence as the only primitive concept.

Reasons for undertaking the above project are mainly the same as those advanced by Tarski for an abstract study of the notion of consequence, namely, the wish of reaching the highest level of generality, by establishing the fundamental properties of concepts which are common to special meta-disciplines, and of applicability to specific deductive disciplines understood as instances of the abstract notion. In particular, an abstract theory of definitions might be applied to a realm of objects, for instance, propositions, which lack the internal structure of the sentences of a fully formalised language.

2 The classical theory of definitions

An abstract theory of definitions aims to define in terms of an abstract consequence relation some notions which intend to capture analogue concepts studied by theories of definitions within formalised languages and logic. Therefore we start by briefly recalling the fundamentals of the most developed and uncontroversial of such theories: The classical theory of definitions for first-order languages.

Let $\mathcal{L}$ be a first-order language with identity. We will use the same symbol $\mathcal{L}$ also to denote the set of all sentences (closed formulæ) of the language. The symbol $\vdash$ denotes the relation of (classical) logical consequence between sets of sentences of $\mathcal{L}$ and sentences of $\mathcal{L}$, equivalently defined, by the completeness theorem, either in terms of rules of inference or in terms of models.

We assume, for simplicity, that among the non-logical constants of $\mathcal{L}$ there is a unary predicate $\mathbf{P}$ we want to define in terms of the other non-logical constants of $\mathcal{L}$. We denote by $\mathcal{L}^-$ the set of sentences of the sublanguage of $\mathcal{L}$ built from the same non-logical constants of $\mathcal{L}$, except $\mathbf{P}$.

Let Σ be any set of sentences of $\mathcal{L}$. Let $\Sigma^- = \Sigma \cap \mathcal{L}^-$. We understand the sentences which are in Σ but not in Σ^- as axioms added to the base theory Σ^- in order to define the predicate $\mathbf{P}$. The classical theory of definitions[3] has that the set of sentences Σ is a correct definition of $\mathbf{P}$ (in terms of the base theory Σ^-) iff Σ has both the following properties:

- *Non-creativity*[4]: Every sentence of $\mathcal{L}^-$ which is provable from Σ is already provable from Σ^-.

[3] The two notions, described below, of "non-creativity" and "eliminability" are first explicitly introduced (under a different terminology) as criteria for a correct definition in [9, fn. 3]. According to Hodges [3, p. 105], Suppes' [6] is probably the first place where the two criteria are "paired as the conditions for a sound definition".

[4] An alternative name for the non-creativity property is *(syntactic) conservativeness.*

- *Eliminability*[5]: Every formula ϕ of $\mathcal{L}$ is provably equivalent in Σ to a formula ϕ^- of $\mathcal{L}^-$.

Moreover, the classical theory of definitions, via Beth's theorem, establishes that the two conditions of non-creativity and eliminability are jointly equivalent to the semantic condition of *determinability*: Every model of Σ^- has one and only one expansion to a model of Σ[6].

3 An abstract theory of definitions

We now turn to Tarski's [8] abstract setting. We work with an arbitrary non-empty set A and with a primitive notion of consequence between elements of A. In the primarily intended interpretation — the above-sketched classical theory of definitions — the set A is replaced by the set of all sentences of $\mathcal{L}$, however, in the abstract setting no properties of A are assumed and its elements can be taken to be sentences as well as any other kind of "unstructured" entities such as, for instance, propositions. For the notion of consequence — primarily interpreted by the relation of classical first-order logical consequence — it is customary to start with the properties characterising a generic notion of "closure", to which further axioms can be added to model more specific intended situations.

Officially, a *consequence relation* on A is a relation $\models$ between subsets Φ and elements ϕ of A satisfying the following properties:

- $\{\phi\} \models \phi$ (*reflexivity*).
- $\Phi \subseteq \Phi' \Rightarrow \forall\phi\,(\Phi \models \phi \Rightarrow \Phi' \models \phi)$ (*monotonicity*).
- $\Phi \models \Psi \wedge \Psi \models \phi \Rightarrow \Phi \models \phi$ (*transitivity*).

Given a consequence relation $\models$ on A we define:

- $\mathsf{Thm}_{\models}(\Phi) = \{\phi \in A \mid \Phi \models \phi\}$.
- $\mathsf{C}_{\models} = \{\Phi \subseteq A \mid \mathsf{Thm}_{\models}(\Phi) = \Phi\}$.

We omit the index $\models$ in $\mathsf{Thm}_{\models}$ and $\mathsf{C}_{\models}$ (and in subsequent similarly defined objects) when it is clear from the context. The members of $\mathsf{Thm}(\Phi)$ are called the *theorems* of Φ (under the consequence relation $\models$). The members of C are called the *theories* of $\models$.

The map $\Phi \mapsto \mathsf{Thm}(\Phi)$ is a *closure operator on* A, i.e., is a function from $\mathcal{P}(A)$ to $\mathcal{P}(A)$ which is monotone, progressive and idempotent. The family C of subsets of A is a *closure system on* A, i.e., $A \in \mathsf{C}$ and for every non-empty family $\mathcal{F} \subseteq \mathsf{C}$ the intersection $\bigcap \mathcal{F}$ belongs to C.

Following Tarski, we say that a subset Φ of A is *consistent* iff there exists $\phi \in A$ such that $\Phi \not\models \phi$. We say that Φ is *maximal consistent* iff Φ is maximal with respect to inclusion in the family of all consistent subsets of A. We denote by U the family (possibly empty) of all maximal consistent subsets of A.

Definitions in first-order logic assume that the full object language $\mathcal{L}$ is split into two subsets: The set of the sentences of $\mathcal{L}$ in which the distinguished predicate $\mathbf{P}$ occurs and the set of the sentences in which $\mathbf{P}$ does not occur, the latter denoted by $\mathcal{L}^-$. Analogously, we assume that the abstract setting is endowed with a distinguished subset A^- of A. We denote by $\models^-$ the

[5]An alternative name for the eliminability property is *(logical) definability*.

[6]By removing from determinability its existence claim, we obtain the uniqueness condition on Σ which in literature is frequently called *implicit definability*: Every model of Σ^- has at most one expansion to a model of Σ.

consequence relation $\models$ restricted to subsets and elements of A^-, which turns out to be itself a closure relation.

Since A^- has to play the role of a "sub-language" of A, it is reasonable to assume A^- to have some degree of "closure". We assume A^- to be *closed under classical negation*, a technical condition which corresponds to the intuitive requirement for a language of being closed under negation and which implies its inconsistency. This is enough to prove that the theories of $\models^-$ are exactly the intersections with $\mathcal{L}^-$ of theories of $\mathcal{L}$.

We can give the following abstract counterpart of corresponding notions involved in the classical theory of definitions. Let X be any subset of A, and let $X^- = X \cap A^-$. We say that

- W is a *syntactic definition* iff W has the properties (with respect to A^-) of non-creativity and *abstract eliminability*, namely, every element $\phi \in A$ is equivalent in W to an element ϕ^- of A^-.
- W is a *relative definition* iff (a) for every consistent subset X of A^- such that $W^- \subseteq X$, the set $X \cup W$ is consistent, and (b) for every maximal (in A^-) consistent subset X of A^- such that $W^- \subseteq X$, the set $X \cup W$ is maximal consistent.
- W is a *semantic definition* iff for every maximal (in A^-) consistent subset X of A^- such that $W^- \subseteq X$, there exists one and only one maximal consistent set U such that $X \cup W \subseteq U$.

The above-mentioned abstract notions of definitions are motivated as follows. The property of non creativity *verbatim* translates from the first-order to the abstract setting. The property of abstract eliminability is a straightforward weakening of the property of first-order eliminability, which we can call, in the first-order context, *sentential eliminability*. Sentential eliminability is the property we obtain from first-order eliminability by replacing the existence of a correspondent equivalent *formula* in the base language for every *formula*, with the existence of a correspondent equivalent *sentence* in the base language for every *sentence*. The property of being a relative definition can be stated *verbatim* in the first-order context and turns out to be equivalent to the conjunction of non-creativity and (first-order) eliminability. Finally, the notion of semantic definition is an abstract counterpart of the first-order notion of determinability: The talk about models is replaced by talk about maximal consistent sets by observing that, in the first-order context, a set of sentences is maximal consistent if and only if is the set of all sentences which are true in a model.

The virtue of the three notions of definition above introduced is that they can be formulated in terms of just an arbitrary consequence relation $\models$ on A and a subset A^- of A, and that they looks as natural counterparts of well-known first-order notions. However, we can say little about the mutual relationships between the three notions in the general case. Even worse, without further assumptions on the consequence relation $\models$ the existence of maximal consistent set is not granted, hence the notions of relative and semantic definition can trivialise.

For these reasons, we need to specify further the relation of consequence and the sublanguage we are dealing with in order to study how the corresponding notions of definitions behave. As a matter of example we can consider the notion of *Henkin consequence*. Recall that a non-empty family $\mathcal{S}$ of subsets of A is a *closure base* for $\models$ iff for every subset Φ and element ϕ of A,

$$\Phi \models \phi \Leftrightarrow \forall Z \in \mathcal{S}\, (\Phi \subseteq Z \Rightarrow \phi \in Z).$$

We say that a closure base $\mathcal{S}$ for $\models$ *has exclusion negation* iff for every $\phi \in A$ there exists $\phi' \in A$ such that

$$\forall Z \in \mathcal{S}\, (\phi \in Z \Leftrightarrow \phi' \notin Z).$$

Finally, we say that $\models$ is *Henkin* iff there exists a closure base for $\models$ having exclusion negation.

Some useful consequences of $\models$ being Henkin:

1. The family U of all maximal consistent sets is not empty and is the unique closure base for $\models$ having exclusion negation.
2. The family U^- of all maximal (in A^-) consistent sets is formed by the intersections with $\mathcal{L}^-$ of the members of U.

The above-mentioned properties allow us to prove the following

Thm 3.1. *Let $\models$ be a Henkin consequence relation on the non-empty set A and let A^- be a non-empty subset of A closed under classical negation. Then, a non-empty subset W of A is a relative definition iff is a semantic definition.*

Moreover, in the first-order context, Theorem 3.1 leads to the following "sentential" version of Beth's theorem, which equates sentential eliminability with a natural weakening of first-order implicit definability:

Thm 3.2. *For a first-order theory Σ, the* sentential *eliminability property is equivalent to the following model-theoretic property: Any two expansions of elementarily equivalent models of Σ^- to models of Σ are elementarily equivalent.*

Finally, I conjecture that, by exploiting the abstract version of Craig's interpolation lemma given in [2], we can prove in the abstract setting that, under the hypotheses of Theorem 3.1, if the Henkin consequence relation $\models$ on A satisfies the further conditions of compactness and weak conjunction[7], then a non-empty subset W of A is a syntactic definition iff is a semantic definition.

References

[1] Cleave, J. P., *A study of logics*, Oxford University Press (1991).

[2] Kosta Došen and Peter Schroeder-Heister, *Uniqueness, definability and interpolation*, The Journal of Symbolic Logic **53** (1988), 554–570.

[3] W. Hodges, *Tarski's theory of definition.* In D. Patterson (ed.), "New Essays on Tarski and Philosophy", Oxford University Press (2008), 94–132.

[4] N. M. Martin & S. Pollard, *Closure Spaces and Logic*, Kluwer Academic Publishers, 1996.

[5] W. A. Pogorzelski & S. J. Surma, *Review of A. Tarski, "Logic, Semantics, Metamathematics. Papers from 1923 to 1938"*, The Journal of Symbolic Logic **34** (1969), 99–106.

[6] P. Suppes, *Introduction to Logic*, D. Van Nostrand Company, 1957.

[7] A. Tarski, *Ueber einige fundamentalen Begriffe der Metamathematik*, Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie **23** cl. III (1930), 22–29.

[8] A. Tarski, *Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften. I*, Monatshefte für Mathematik und Physik **37** (1930), 361–404.

[9] A. Tarski, *Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe*, Erkenntnis **5** (1935), 80–100.

[10] A. Tarski, *Logic, Semantics, Matamathematics. Papers from 1923 to 1938*, Translated from German to English by J. H. Woodger, Clarendon Press, 1969.

[7] Cfr. [1] for the weak conjunction property.