Definitions in a Hyperintensional Free Logic*

Edward N. Zalta†
Philosophy Department
Stanford University

Abstract

In this paper we describe the theory of definitions developed for object theory (OT), i.e., the formal, deductive system developed in the monograph *Principia Logico-Metaphysica* (excerpted online). OT exhibits a number of features which call for great care when stating, and reasoning from, definitions. The features in question include: (a) identity is not a primitive, (b) there are complex terms (descriptions and λ-expressions) that may fail to denote, (c) all terms, including descriptions, are interpreted rigidly, (d) the axiomatization of an actuality operator includes a contingent axiom (and so the modal logic allows for reasoning from contingencies), and (e) the theory of relations in OT is hyperintensional and so one can't substitute necessary equivalents in all contexts. The theory of definitions addresses the issues that arise in connection with (a) – (e) and includes two kinds of definition, namely, definitions by equivalence and definitions by identity, each of which has a distinctive inferential role.

*This paper includes a number of (edited) excerpts from different sections of an unpublished manuscript (Zalta m.s.). These have been woven together here to present a more unified picture of the theory of definitions needed for any system with expressive power similar to that of the language formulated in the manuscript. I'd like to thank Daniel Kirchner and Daniel West for comments on the paper that led to improvements.

†I'm grateful to Otto Neumaier and Peter Simons for inviting me to contribute to this volume of papers dedicated to Edgar Morscher. I first met Edgar during the 1988–1989 academic year, when he came to Stanford to take up a one-year appointment to the Distinguished Visiting Austrian Chair. He contacted me, I think, because his colleague in Salzburg, Peter Simons, had reviewed my book of 1983 and both were aware that I had formalized some ideas of the Austrian philosopher Ernst Mally. We immediately discovered other topics of mutual interest and met on a regular basis during that year. I'm grateful to Edgar for inviting me to Salzburg to lecture in June 1990, since I've met many outstanding philosophers during my trips there. I'll miss his eager and earnest engagement with ideas and colleagues.

1 Introduction

The formal, deductive system of 'object theory' (OT) has been put forward and applied in a number of papers since 1983. The canonical formulation, however, appears in *Principia Logico-Metaphysica* (Zalta m.s.), where it is first expressed in a 2nd-order quantified modal language. Though OT is also formulated in relational type-theory, we focus here only on the theory of definitions formulated for 2nd-order OT. The language of 2nd-order OT has expressive power beyond that of classical 2nd-order QML; in addition to the modal operator, OT includes (a) a second mode of predication, i.e., OT includes both atomic exemplification formulas of the form $F^n x_1 \ldots x_n$ and *encoding* formulas of the form $x_1 \ldots x_n F^n$, (b) an actuality operator $\mathcal{A}$, (c) complex individual terms, namely, definite descriptions of the form $\imath x\varphi$ (interpreted rigidly) and (d) complex n-ary relation terms, namely, both (i) λ-expressions of the form $[\lambda x_1 \ldots x_n\, \varphi]$ (for $n \geq 0$) and (ii) formulas (for $n = 0$). The formulas denote propositions, where these are taken to be 0-ary relations. Both descriptions and n-ary relation terms ($n \geq 1$) may fail to denote, and so OT uses a negative free logic for reasoning with complex terms, though reasoning with primitive (individual and relation) constants and variables is still classical. Identity is not taken as a primitive in OT.

To systematize this additional expressive power in the language, OT adds, to the logic of 2nd-order quantified S5 modal logic, (a) definitions for existence and identity and (b) axioms governing encoding formulas, governing the actuality operator, and governing both kinds of complex terms. These definitions and axioms and will be introduced as needed in what follows. For now, it is important to know that 2nd-order OT has the following features:

- the modal logic assumes a fixed domain of individuals and a fixed domain of n-ary relations ($n \geq 0$),
- the definitions for identity yield n-ary *hyperintensional* relations, since $\Box\forall x_1 \ldots \forall x_n(F^n x_1 \ldots x_n \equiv G^n x_1 \ldots x_n)$ doesn't imply $F = G$, for any $n \geq 1$, and $\Box(p \equiv q)$ doesn't imply $p = q$, when $n = 0$.
- both kinds of complex terms (descriptions and λ-expressions) may fail to denote,
- all terms, including definited descriptions, are interpreted rigidly

(semantically, the denotation function isn't indexed to the primitive possible worlds used in the semantics, but simply assigns a value, in the relevant domain, to the terms),[1] and

- the axiomatization of the actuality operator includes a contingent axiom (and so, to avoid deriving necessities from contingencies, the Rule of Necessitation may not be applied to any line of a proof that depends on a contingency).

With definite descriptions and λ-expressions in the system, individual constants can be introduced by using a definite description as a definens and new n-ary relation constants can be introduced by using an n-ary λ-expression as a definiens.

We adopt the standard view that, in a definition, a *new* expression, the *definiendum*, is introduced by way of a *definiens* that contains only primitive expressions or previously defined expressions. In building a theory of definitions for this system, we shall not regard definitions as metalinguistic abbreviations of the object language.[2] Instead, we shall regard them as conventions for:

(a) extending the object language with new formulas and terms (often with the help of new syncategorematic expressions), and

(b) conservatively extending OT's deductive system with new and *safe* axioms.

As such, claims stated in terms of defined notions become genuine philosophical statements of the object language rather than statements of the metalanguage; indeed, some of the axioms and axiom schemas of OT are stated in terms of these defined notions.[3] We first focus our attention on how definitions achieve (a). Then we discuss (b) in Sections 4 and 5, where we carefully characterize the inferential role that such definitions play within OT's deductive system.

[1] Definite descriptions $\imath x\varphi$ are assigned an individual, if there is one, that uniquely satisfied the formula φ at the distinguished actual world.

[2] In developing the theory of definitions for this system, I found the following works especially helpful: Frege 1879, §24; Padoa 1900; Frege 1903a, §§55–67, §§139–144, and §§146–147; Frege 1903b, Part I; Frege 1914, 224–225; Suppes 1957; Mates 1972; Dudman 1973; Belnap 1993; Hodges 2008; Urbaniak & Hämäri 2012; and Gupta 2023. Hodges 2008 and Urbaniak & Hämäri 2012 provide insightful discussions of the contributions by Kotarbiński, Łukasiewicz, Leśniewski, Ajdukiewicz and Tarski to the elementary theory of definitions.

[3] Daniel Kirchner has observed, personal communication, that an additional reason for not thinking of definitions as abbreviations is that if they are abbreviations, both the extensions and the intensions of the terms would be identical. So if we want to allow for hyperintensionality, we shouldn't conceive of definitions as mere abbreviations.

Further, the inferential role of definitions described below will abide by two classical criteria for proper definitions: eliminability and non-creativity (conservativity). I shall not spend any time here discussing these criteria; they are well-known and are thoroughly discussed in the literature cited in footnote 2. But see, especially, Hodges 2008 (104), for noting that the eliminability criterion can be traced back to Pascal 1658 and even to Porphyry [OAC, 43]. As to non-creativity, Frege's discussion in 1903a (§§139–144 and and §§146–147) seems to be the starting point of discussion.

In what follows, I use Greek letters as metavariables: $\varphi, \psi, \ldots$ to range over formulas, τ and σ to range over terms generally, ν to range over individual variables, and α and β to range over both individual and relation variables indifferently. For simplicity, I forego the exact specification of the language via a metadefinition identifying the terms and formulas of the language. It is pretty much what you would expect. Though I think the reader should be able to infer the precise specification of the language of OT from the discussion, the manuscript referenced above (Zalta m.s., hereafter 'PLM') includes, in the chapter titled 'The Language', both a metadefinition with a simultaneous recursive definition of *term* and *formula*, and a separate BNF-style definition (in which the Greek metavariables abbreviate names of grammatical categories).

2 Two Kinds of Definition

The most important feature of OT's theory of definitions is that there are two kinds: *definitions-by-equivalence* and *definitions-by-identity*, or more simply, *definitions-by-*≡ and *definitions-by-*=, respectively. Here is an intuitive (but not quite accurate) characterization of the distinction: the former stipulate a necessary equivalence between formulas (= expressions having truth conditions), whereas the latter stipulate an identity between terms (= expressions having denotations). A definition-by-≡ has the form:

$$\varphi \equiv_{df} \psi$$

In the general case, where there are m distinct free variables $\alpha_1, \ldots, \alpha_m$

($m \geq 0$) in the definiens and definiendum, a definition-by-≡ has the form:[4]

$$\varphi(\alpha_1,\ldots,\alpha_m) \equiv_{df} \psi(\alpha_1,\ldots,\alpha_m)$$

By contrast a definition-by-= has the form:

$$\tau =_{df} \sigma$$

provided τ and σ are both terms *of the same type* (i.e., either both individual terms or both n-ary relation terms, for some n). In the general case, where there are m distinct free variables $\alpha_1,\ldots,\alpha_m$ ($m \geq 0$) in the definiens and definiendum, a definition-by-= has the form:

$$\tau(\alpha_1,\ldots,\alpha_m) =_{df} \sigma(\alpha_1,\ldots,\alpha_m)$$

provided τ and σ are terms of the same type.

Note that it would be incorrect to *distinguish* definitions-by-≡ and definitions-by-= by saying that the former introduce new formulas and the latter introduce new terms. This oversimplification is undermined by the fact, in OT, that all and only formulas are 0-ary relation terms, a fact which has the following consequences. First, the definiendum φ and definiens ψ in a definition-by-≡ are 0-ary relation terms as well as formulas—so in every case, these definitions introduce new 0-ary relation terms. Second, the definiendum τ and definiens σ in a definition-by-= may be formulas if τ and σ are 0-ary relation terms—so in some cases, such definitions introduce new formulas.

In light of these observations, we emphasize that the distinction between definitions-by-≡ and definitions-by-= concerns their *inferential role* in the deductive system. The inferential role of both kinds of definition has to be carefully formulated and this will be done in Sections 4 and 5 below. For now, we can describe their role pre-theoretically as follows: (a) a definition-by-≡ implicitly introduces necessary biconditionals as axioms,[5] and (b) a definition-by-= implicitly introduces axioms that intuitively assert: if the definiens has a denotation, then identity holds and if the definiens doesn't denote, then neither does the definiendum, i.e., formally, such axioms take the form $(\sigma\downarrow \rightarrow (\tau = \sigma))\ \&\ (\neg\sigma\downarrow \rightarrow \neg\tau\downarrow)$, where $\sigma\downarrow$ asserts in the object language that σ exists (and asserts semantically that σ has a denotation). Since necessary biconditionals and necessary conjunctions of conditionals (that aren't converses of each other) have very different inferential roles, the inferences one can draw from the two forms of definition will be very different. Note that if we are going to formulate the inferential role of definitions-by-= in terms as introducing identity statements, then OT's language without identity will need definitions-by-≡ for formulas of the form $\tau = \sigma$ (i.e., formulas such as $x = y$ and $F^n = G^n$, for $n \geq 0$). So OT bootstraps itself into a position where one may formulate the inferential role of definitions-by-=.

Before we can state the inferential role of the two types of definitions, it is important to first discuss a number of issues that affect the formulation and interpretation of definitions. The first issue concerns a subtlety that arises in connection with encoding formulas. In an encoding formula of the form $x_1 \ldots x_n F^n$ ($n \geq 1$), the x_i are the *relata* of the relation F^n and we shall say that that they *occur in encoding position* in this formula. The presence of free variables in encoding position can give rise to the Clark/Boolos paradox; one may not assume that arbitrary conditions φ, in which there are free individual variables that occur in encoding position, can be used to define a property. For example, the expression $[\lambda x\, \exists F(xF\ \&\ \neg Fx)]$ (*being an x that fails to exemplify a property it encodes*), in which the variable bound by the λ occurs *in encoding position* in the matrix $\exists F(xF\ \&\ \neg Fx)$, provably fails to denote on pain of contradiction.[6]

[4]For simplicity, we shall consider only definitions in which all and only the variables that occur free in a definiens also occur free in the definiendum. Of course, for some purposes, it may be useful to relax this requirement by allowing the definiendum to contain free variables that aren't free in the definiens. Suppes (1957) explains why one can allow definitions in which variables occur free in the definiendum but not in the definiens; one can trivially get the variables to match by adding dummy clauses to the definiens. For example, Suppes notes (1957, 157) that the number-theoretic definition $Q(x,y) =_{df} x > 0$ can be turned into $Q(x,y) =_{df} x > 0\ \&\ y = y$. But Suppes also nicely explains why allowing the definiens to contain free variables that aren't free in the definiendum would be catastrophic. See his example (1957, 157) of how to derive a falsehood from a definition such as $R(x) \equiv_{df} x + y = 0$.

[5]Strictly speaking, the language of OT uses ¬ (not) and → (if-then) as the only primitive propositional connectives and so the propositional axioms of OT only govern these two connectives. So the logic of ≡ has to be derived. Since we state the inferential role of definitions-by-≡ before the logic of ≡ is derived, we will say, *in the first instance*, that the inferential role of $\varphi \equiv_{df} \psi$ is to implicitly introduce the *conditionals* $\varphi \rightarrow \psi$ and $\psi \rightarrow \varphi$ as non-contingent axioms; from this, the (modal) logic of OT will guarantee that such definitions yield $\Box(\varphi \equiv \psi)$ as theorems. We'll discuss this in more detail in Section 4.

[6]The Clark/Boolos paradox is this: if the expression $[\lambda x\, \exists F(xF\ \&\ \neg Fx)]$ were to denote a property, call it K, then OT's comprehension principle for abstract objects would assert the existence of an abstract object that encodes K. Any such object would exemplify K if and only if it doesn't. See Clark 1978 and Boolos 1987 for statements of the paradox

Consequently, it is provable that $\neg[\lambda x\,\exists F(xF\,\&\,\neg Fx)]\!\downarrow$; i.e., it is provable that *being an object that fails to exemplify a property it encodes* fails to exist. More generally, if any of the x_i bound by the λ in the expression $[\lambda x_1 \ldots x_n\, \varphi]$ occurs in encoding position in φ, the system doesn't assert that the λ-expression has a denotation.[7]

So, the subtlety introduced by encoding formulas is that if we introduce a definiendum via a definiens that has one or more free variables in encoding position, then we have to regard the definiendum as having those same free variables in encoding position. Here's a schematic example. Suppose one were to introduce a new condition on entities $\alpha_1,\ldots,\alpha_m$ by formulating a definition-by-equivalence with the following form ($m \geq 1$):

$$Notion(\alpha_1,\ldots,\alpha_m) \equiv_{df} \psi(\alpha_1,\ldots,\alpha_m)$$

Then, if one or more of the α_i is an individual variable occurring in encoding position somewhere in in the definiens $\psi(\alpha_1,\ldots,\alpha_m)$, we must regard the definiendum $Notion(\alpha_1,\ldots,\alpha_m)$ similarly, i.e., the α_i that occur in encoding position in ψ are to be regarded as occurring in encoding position in the definiendum $Notion(\alpha_1,\ldots,\alpha_m)$.[8]

Similarly, suppose one introduces a new function term τ by formulating a definition-by-identity such as the following:

$$\tau(\alpha_1,\ldots,\alpha_m) =_{df} \sigma(\alpha_1,\ldots,\alpha_m)$$

Then, if one or more of the α_i in τ is an individual variable that occurs in encoding position somewhere in σ, we must regard the definiendum

independent of each other and independent of the OT formalism.

In previous versions of OT, I've constructed the language so that λ-expressions such as $[\lambda x\,\exists F(xF\,\&\,\neg Fx)]$ are not well-formed. Now, however, OT allows such expressions to be well-formed and uses a free logic for complex terms so that such expressions aren't automatically assumed to denote relations.

[7] However, there is an axiom that asserts the conditions under which λ-expressions of this kind denote. It says, intuitively, that if the relation $[\lambda x_1 \ldots x_n\, \psi]$ exists and $\Box\forall x_1 \ldots \forall x_n(\psi \equiv \varphi)$, then $[\lambda x_1 \ldots x_n\, \varphi]$ exists. See the chapter on the axioms of OT in PLM.

[8] Here is a specific example of this convention. Suppose we define "x is a *concept of* y" as follows:

$$ConceptOf(x,y) \equiv_{df} \forall F(xF \equiv Fy),$$

Then we must regard the variable x in $ConceptOf(x,y)$ as occurring in encoding position. That means that the λ in the expression $[\lambda x\, ConceptOf(x,y)]$ binds a variable in encoding position and so the λ-expression isn't guaranteed to have a denotation.

τ similarly, i.e., the α_i that occur in encoding position in σ are to be regarded as occurring in encoding position in the definiendum τ.[9]

3 Object-Language Variables That Function as Metavariables

The next issue that arises for understanding the two types of definitions concerns the fact that, in a free logic, definitions must be formulated schematically, either by using object language variables and regarding them as functioning as metavariables or by formulating the definitions schematically with metavariables. This applies to both free variables and quantified variables in the definitions. Of course, we will chose to use object language variables and interpret them as metavariables, since that makes the definitions much easier to read. But here are the reasons why the variables in definitions in a free logic have to be formulated schematically. We consider free variables first and then bound variables.

3.1 Why Free Variables in Definitions Should Be, or Should Function as, Metavariables

To see why in systems like OT, free variables in definitions have to be (understood as) metavariables, let's first consider a definition-by-$\equiv$. Suppose one wanted to define the condition "object x *contingently exemplifies* property F" by stipulating that it holds just in case x exemplifies F but not necessarily. One might expect to see this definition formalized using object language variables as follows:

$$ContingentlyExemplifies(x,F) \equiv_{df} Fx\,\&\,\neg\Box Fx \quad (1)$$

If the variables x and F are interpreted as object language variables, then we run into the following problem.

[9] Here is a specific example of this convention. Suppose we define "the *sum of* x and y" ('$x \oplus y$') as "the abstract object ($A!x$) that encodes all and only those properties F such that either x encodes F or y encodes F", i.e., as follows:

$$x \oplus y =_{df} \imath z(A!z\,\&\,\forall F(zF \equiv xF \vee yF))$$

Then we must regard the free variables x and y as occurring in encoding position in $x \oplus y$. Thus, x and y occur in encoding position in the formula $Px{\oplus}y$ ("$x{\oplus}y$ exemplifies P"). So the λ in the expression $[\lambda xy\, Px \oplus y]$ binds variables in encoding position and isn't guaranteed to have a denotation.

On the traditional understanding, a definition such as (1) becomes available to the deductive system as a biconditional axiom, i.e., as an axiom where $\equiv$ replaces $\equiv_{df}$ in (1). This understanding of definitions suffices in systems of classical logic in which every term of the language has a denotation. The classical logic of quantification permits the instantiation of *any* individual or relation term τ of such a language into a universally quantified claim of the form $\forall\alpha\varphi$. In such systems, (1) not only extends the language with new formulas of the form

$$\mathit{ContingentlyExemplifies}(\kappa,\Pi)$$

(where κ is any individual term and Π any unary relation term), but also extends the deductive system with necessary axioms such as:

$$\mathit{ContingentlyExemplifies}(x,F) \equiv (Fx \,\&\, \neg\Box Fx) \tag{2}$$

$$\forall x\forall F(\mathit{ContingentlyExemplifies}(x,F) \equiv (Fx \,\&\, \neg\Box Fx)) \tag{3}$$

In classical systems, every object term κ and every property term Π can be instantiated, respectively, for $\forall x$ and $\forall F$ in (3) to yield biconditional theorems stating the necessary and sufficient conditions for the definiendum $\mathit{ContingentlyExemplifies}(\kappa,\Pi)$.

However, in OT, the complex individual terms (definite descriptions), complex n-ary ($n \geq 1$) relation terms (λ-expressions), and defined terms may fail to have a denotation.[10] It uses a *negative free logic* for such terms; the axiom for universal instantiation is formulated so that a term τ can be instantiated into a universal claim only on the condition that it has a denotation.[11] A complementary axioms asserts that primitive constants and variables do have denotations.

[10]We don't include 0-ary relation terms (and thus formulas) in this list, since it is a theorem of OT that all such terms have a denotation.

[11]OT uses an axiom that preserves the intent of the standard axiom used in most 2nd-order negative free logics. The standard axiom is:

$\forall\alpha\varphi \to (\exists\beta(\beta = \tau) \to \varphi^{\tau}_{\alpha})$, provided τ is substitutable for α in φ

This standard axiom assumes that identity is primitive or defined in the language. But in a classical 2nd-order quantified modal language (i.e., without encoding formulas) there isn't a good way to define $F = G$ except in terms of the condition $\Box\forall x(Fx \equiv Gx)$, which thereby collapses necessarily equivalent propositions. In some free logics (e.g., Feferman 1995) use a primitive $\downarrow$ (definedness) and revise the above axiom to:

(ϑ) $\forall\alpha\varphi \to (\tau\!\downarrow \to \varphi^{\tau}_{\alpha})$, provided τ is substitutable for α in φ

In OT, however, $\tau\!\downarrow$ is not primitive but rather defined: it is defined in terms of exemplification predication when τ is an individual term and in terms of encoding predication when τ is a relation term. OT then takes (ϑ) as axiomatic.

Now consider a description like $\imath z(Pz \,\&\, \neg Pz)$, which provably fails to have a denotation. As noted previously, in OT, the metatheoretical notion of a term's having a denotation (i.e, a term's *being significant*) is represented in the object language by the theoretical notion of *existence*, which is defined and symbolized by $\downarrow$. (The definition need not concern us here; the proof that $\imath z(Pz \,\&\, \neg Pz)$ fails to be significant, i.e., the proof that $\neg\imath z(Pz \,\&\, \neg Pz)\!\downarrow$, goes by way of a reductio.) Let's abbreviate the formula $Pz \,\&\, \neg Pz$ as ψ_1, so that we know $\neg\imath z\psi_1\!\downarrow$. If definition (1) implicitly introduces (3) as an axiom, then although the classical logic of constants would allow us to instantiate the primitive relation constant P for the universal quantifier $\forall F$ in (3), the *negative free logic* of non-denoting terms would not allow us to instantiate the description $\imath z\psi_1$ for the universal quantifier $\forall x$ in (3). Thus, we wouldn't be able to derive from (3):

$$\mathit{ContingentlyExemplifies}(\imath z\psi_1, P) \equiv (P\imath z\psi_1 \,\&\, \neg\Box P\imath z\psi_1) \tag{4}$$

Given the classical understanding of definitions on which (1) implicitly introduces (2) and (3) as axioms, a negative free logic prevents us from deriving (4). Thus, in a logic where the complex terms may fail to denote, the notion *ContingentlyExemplifies* isn't completely defined: given that ψ_1 is $Pz \,\&\, \neg Pz$, no biconditional theorem states the necessary and sufficient conditions for the particular formula $\mathit{ContingentlyExemplifies}(\imath z\psi_1, \Pi)$, for any property term Π.

Moreover, if we can't derive (4), then we can't derive:

$$\neg\mathit{ContingentlyExemplifies}(\imath z\psi_1, P) \tag{5}$$

This would follow from (4) and the fact that in negative free logic, $\neg P\imath z\psi_1$ is a consequence of $\neg\imath z\psi_1\!\downarrow$.[12]

[12]If a 2nd-order negative free logic includes both individual and relation terms that might fail to denote, then it would assert, as an axiom, that for any relation term Π^n and individual terms $\kappa_1,\ldots,\kappa_n$ ($n \geq 1$):

$\Pi^n\kappa_1,\ldots,\kappa_n \to (\Pi^n\!\downarrow \,\&\, \kappa_1\!\downarrow \,\&\, \ldots \,\&\, \kappa_n\!\downarrow)$

So as an instance of the contrapositive, we have $\neg\imath z\psi_1\!\downarrow \to \neg P\imath z\psi_1$.

As an aside, it should be noted that in OT, an analogous axiom holds for atomic encoding formulas: for any relation term Π^n and individual terms $\kappa_1,\ldots,\kappa_n$ ($n \geq 1$):

$\kappa_1,\ldots,\kappa_n\Pi^n \to (\Pi^n\!\downarrow \,\&\, \kappa_1\!\downarrow \,\&\, \ldots \,\&\, \kappa_n\!\downarrow)$

is an axiom.

Similarly, consider a property term that provably fails to denote, such as the λ-expression $[\lambda x \exists F(xF \& \neg Fx)]$ that leads to the Clark/Boolos paradox. Let's abbreviate $\exists F(xF \& \neg Fx)$ as φ_1, so that the the λ-expression can be referenced as $[\lambda x\, \varphi_1]$. We know that this expression provably fails to denote, i.e., that $\neg[\lambda x\, \varphi_1]\downarrow$ is provable. Then, analogously, if (1) implicitly introduces (3) as an axiom, then although the classical logic of constants would allow us to instantiate the individual constant a for the quantifier $\forall x$ in (3), the *negative free logic* of non-denoting terms would not allow us to instantiate the $[\lambda x\, \varphi_1]$ for the quantifer $\forall F$ in (3). Thus, we wouldn't be able to obtain the following as a theorem:

$$ContingentlyExemplifies(a, [\lambda x\, \varphi_1]) \equiv ([\lambda x\, \varphi]a \;\&\; \neg\Box[\lambda x\, \varphi_1]a) \tag{6}$$

And without (6), we wouldn't be able to derive the desired fact that $\neg ContingentlyExemplifies(a, [\lambda x\, \varphi_1])$ from the fact that $\neg[\lambda x\, \varphi_1]\downarrow$.

We can avoid the general problem just described by using metavariables and formulating (1) as a schema. Let κ be a metavariable ranging over individual terms and Π be a metavariable ranging over unary relation terms. Then (7) avoids the problems (1) has:

$$ContingentlyExemplifies(\kappa, \Pi) \;\equiv_{df}\; \Pi\kappa \;\&\; \neg\Box\Pi\kappa \tag{7}$$

A definition schema such as (7) implicitly extends our language with the new syncategorematic expression *ContingentlyExemplifies* and new formulas of the form $ContingentlyExemplifies(\kappa, \Pi)$. But, just as importantly, (7) will implicitly introduce (the closures of) the instances of the following schema as new axioms:

$$ContingentlyExemplifies(\kappa, \Pi) \equiv (\Pi\kappa \;\&\; \neg\Box\Pi\kappa)$$

Given such an understanding of definition schemata, (7) yields (4), (5), and (6) as theorems, since these are all instances of the above biconditional. Consequently, the use of metavariables in (7) is required and (1), strictly speaking, doesn't suffice as a definition. However, since (7) is more complex and more difficult to read than (1), OT employs the convention: the free variables in (1) function as metavariables, so that (4), (5), and (6) become instances of the definition.

Now let's consider definitions-by-=. Though object-language variables that occur free in a definition-by-= should also function as metavariables, it is *not* for the reason just outlined. To see why, let's consider an unusual example that has some interesting probative features. In OT, it is a theorem $\imath x(x = y)\downarrow$. This asserts, for an arbitrary object y, that the x such that x is identical to y exists. Let's use $\imath x(x = y)$ as the definiens for ι_y ("the y") in the following definition-by-=:

$$\iota_y \;=_{df}\; \imath x(x = y) \tag{8}$$

For example, ι_a ('*the a*') is thereby defined as the individual x identical to a. (8) is a fine definition given that the definiens has a denotation for each value assigned to the free variable y. No matter what is assigned to y, ι_y denotes the individual that is identical to y, i.e., denotes y.

Traditionally, (8) would be understood as extending our language with a host of new complex terms. Though (8) uses the free object-language variable y, it is standard to assume that (8) would extend our language with terms of the form ι_κ, where κ is any term. So all of the following would be well-formed: ι_y, $\iota_{\imath z\psi_1}$, ι_{ι_y}, etc.

Also, traditionally, (8) would be understood as implicitly introducing the closures of the axiom $\iota_y = \imath x(x = y)$, so that the following is axiomatic:

$$\forall y\big(\iota_y = \imath x(x = y)\big) \tag{9}$$

In a classical logic, in which all terms have denotations, the quantifier $\forall y$ in (9) can be instantiated to any term other than x (since x would get captured by the variable-binding operator $\imath x$).

At first glance, this understanding of the inferential role of definitions-by-= would appear to be desirable, for in OT, we have individual terms that fail to denote; OT's negative free logic does not permit us to instantiate empty terms into (9). So where $\imath z\psi_1$ is the example of a non-denoting description introduced above, we may not instantiate (9) to infer:

$$\iota_{\imath z\psi_1} = \imath x(x = \imath z\psi_1) \tag{10}$$

Not only is (10) problematic on the grounds that both terms flanking the identity sign fail to have a denotation, but in OT it is both provably false[13] and leads to a contradiction.[14] Identity statements can't be true when one or both of the terms flanking them are empty, unless heroic measures are taken, something we'll forego here. So, in a negative free logic, the classical interpretation of (8) as introducing (9) blocks the introduction of identities like (10) with non-denoting descriptions.

Since (10) is problematic, and our negative free logic prevents us from inferring it from (9), one might conclude at this point that we *should* interpret (8) as introducing (9) and that we need not interpret the object-language variables in (8) as metavariables. For if we were to interpret (8) as introducing, for any individual term κ, the axiom schema:

$$\iota_\kappa = \imath x(x=\kappa)$$

then we would have (10) as an instance.

But the conclusion that we shouldn't interpret the free variables in (8) as metavariables is too hasty. To see why, consider the provably empty description $\imath z\psi_1$ introduced earlier. One might wish to allow that $\iota_{\imath z\psi_1}$ is a well-formed expression of the language but also have a mechanism for *proving* that the definiens $\imath x(x=\imath z\psi_1)$ and the definiendum $\iota_{\imath z\psi_1}$ both fail

[13]To see this, we have to cite definitions, axioms, and theorems not yet introduced, but here is an intuitive proof sketch. Assume $\neg\imath z\psi_1\!\downarrow$. Then choose a variable, say x, that isn't free in $\imath z\psi_1$. It follows by a theorem of OT that $\neg(x=\imath z\psi_1)$ (no value for x can satisfy the formula $x=\imath z\psi_1$). Since x isn't free in our assumption, it follows by GEN that this holds for any object x, i.e., that $\forall x\neg(x=\imath z\psi_1)$, i.e., $\neg\exists x(x=\imath z\psi_1)$.

From this last conclusion, we may infer, by the laws of definite descriptions and the definition of $\downarrow$ that $\neg\imath x(x=\imath z\psi_1)\!\downarrow$, as follows:

> Assume, for reductio, that $\imath x(x=\imath z\psi_1)\!\downarrow$. Then by definition of $\downarrow$, it follows that for some property, say P, that $P\imath x(x=\imath z\psi_1)$. But, then, by Russell's analysis of descriptions, it follows *a fortiori* that that $\exists x(x=\imath z\psi_1)$, which contradicts what we proved above.

From $\neg\imath x(x=\imath z\psi_1)\!\downarrow$, we can again conclude that $\neg(\iota_{\imath z\psi_1} = \imath x(x=\imath z\psi_1))$. Thus, we have a proof of the negation of (10).

[14]In OT, the definition of identity for individuals asserts that individuals x and y are identical just in case they are both ordinary objects ($O!x$) and necessarily exemplify the same properties (i.e., $O!x \,\&\, O!y \,\&\, \Box\forall F(Fx \equiv Fy)$), or they are both abstract objects that (necessarily) encode the same properties (i.e., $A!x \,\&\, A!y \,\&\, \Box\forall F(xF \equiv yF)$). (We put 'necessarily' in parenthesis because the modal logic of encoding is $xF \rightarrow \Box xF$, and so if one can show that two abstract objects in fact encode the same properties, i.e., that $\forall F(xF \equiv yF)$, then it will follow that they necessarily do so, i.e., that $\Box\forall F(xF \equiv yF)$.) So, given the definition of identity, (10) will imply, for example, either $O!\imath x(x=\imath z\psi_1)$ or $A!\imath x(x=\imath z\psi_1)$. In either case, an axiom for free logic of the kind discussed in footnote 12 would let us conclude that the description is significant, i.e., that $\imath x(x=\imath z\psi_1)\!\downarrow$. But this would contradict something we established in footnote 13, namely that $\neg\imath x(x=\imath z\psi_1)\!\downarrow$.

to have a denotation. Of course, one expedient adopted in other systems is to disallow instances of (8) unless one can prove that the definiens of that instance is provably significant. And some systems allow for 'conditional' definitions. But these could be considered heroic measures: they would force us to establish certain existence claims before introducing definitions and they could potentially leave us with expressions in the language (such as $\iota_{\imath z\psi_1}$) that appear to be perfectly well-formed but are simply undefined. In Section 5, we'll focus on the issues that arise for definitions-by-identity when the definiens fails to denote.

But for now, note that in footnote 13, we saw that $\neg\imath x(x=\imath z\psi_1)\!\downarrow$ is a consequence of $\neg\imath z\psi_1\!\downarrow$, but (8) doesn't allow us to derive $\neg\iota_{\imath z\psi_1}\!\downarrow$ from $\neg\imath x(x=\imath z\psi_1)\!\downarrow$. Intuitively, if the definiens $\imath x(x=\imath z\psi_1)$ fails to be significant, then we should be able to derive that the definiendum $\iota_{\imath z\psi_1}$ fails to be significant. So the problem is that (8), under the standard interpretation of its object-language variables, doesn't give us a means to conclude $\neg\iota_{\imath z\psi_1}\!\downarrow$ from the theorem $\neg\imath x(x=\imath z\psi_1)\!\downarrow$. So we are in a situation where the term $\iota_{\imath z\psi_1}$ appears to be *well-formed* and we know the claim $\iota_{\imath z\psi_1}\!\downarrow$ is false (because $\imath x(x=\imath z\psi_1)$ is empty) but we can't prove it.

Our solution will be to let the object-language variables in definitions-by-= function as metavariables but *revise* our understanding of the inferential role of these definitions. We'll allow any terms to be substituted for the free object-language variables so that we have instances of the definition for every individual term of the language. But the inferential role of the definition, formulated in Section 5 below, will be introduced by a metarule stipulating that a certain conjunction of conditionals is axiomatic. For the particular instance of (8) we're now considering, namely, $\iota_{\imath z\psi_1} =_{df} \imath x(x=\imath z\psi_1)$, the metarule will stipulate that the following is a necessary axiom schema: if $\imath x(x=\imath z\psi_1)$ exists, then $\iota_{\imath z\psi_1} = \imath x(x=\imath z\psi_1)$, and if $\imath x(x=\imath z\psi_1)$ fails to exist, then $\iota_{\imath z\psi_1}$ fails to exist.

3.2 Why Bound Variables in Definientia Should Be, or Should Function As, Metavariables

Consider definitions-by-≡ first. Suppose one were to define: x and y are *indiscernible* just in case x and y exemplify the same properties. We might introduce this definition formally as:[15]

[15]The following example is illustrative only, for in OT, the proper way to introduce this notion of *indiscernibility* is to define-by-identity a relation $\doteqdot$ as follows:

$$Indiscernible(x,y) \equiv_{df} \forall F(Fx \equiv Fy) \tag{11}$$

In traditional logics, this converts to an axiom asserting:

$$Indiscernible(x,y) \equiv \forall F(Fx \equiv Fy)$$

from which it is then derivable, by a theorem schema that asserts the (necessary) equivalence of alphabetic variants, that:

$$Indiscernible(x,y) \equiv \forall G(Gx \equiv Gy) \tag{12}$$

So one doesn't need the definition:

$$Indiscernible(x,y) \equiv_{df} \forall G(Gx \equiv Gy) \tag{13}$$

since we can derive (12) from the original definition (11).

Though OT does have the resources to prove that alphabetically-variant formulas such as $\forall F(Fx \equiv Fy)$ and $\forall G(Gx \equiv Gy)$ are (necessarily) equivalent, quite a number of definitions are developed immediately after the language of OT is specified and *before* the proof theory (i.e., axioms and rules of inference) of OT is specified. So to ensure the understanding that any alphabetic variant of a definiens will yield the same definition with the same inferential role, it is useful (and possibly prudent) to take the *bound* variable $\forall F$ in (11) to be a metavariable.[16] Thus (13) becomes a perfectly good version of the definition.

The foregoing discussion can be easily adapted to explain why we similarly interpret the bound object language variables in definitions-by-identity as metavariables. But we leave this for the reader and instead summarize the situation thus far. Our discussion suggests that,

$$\dot{\equiv} =_{df} [\lambda xy\, \forall F(Fx \equiv Fy)]$$

Then by adopting infix notation, the formula $x \dot{\equiv} y$ becomes a well-formed exemplification formula. (In the λ-expression used as definiens, none of the variables bound by the λ are in encoding position in $\forall F(Fx \equiv Fy)$, and so it denotes a relation.) While this is the correct way to define a relation, it doesn't give us an example of a definition-by-equivalence.

[16] It should be noted that in OT, one has to prove not only the (necessary) equivalence of arbitrary, alphabetically-variant formulas, but also that alphabetically-variant individual terms (e.g., $\imath xFx$ and $\imath yFy$) and alphabetically-variant relation terms (e.g., $[\lambda x\, \neg Fx]$ and $[\lambda y\, \neg Fy]$) have the same denotation. The proof of these facts does take place once a body of theorems has been established, and though a proof for specific instances may be developed on a case-by-case basis, it is clear that we can avoid circularity issues (e.g., requiring facts about alphabetic-variants to prove facts about alphabetic-variants) by formulating definitions either with metavariables or with object language variable under the assumption that they are to be interpreted as metavariables.

strictly speaking, we should formulate (11) as follows. Where Ω is a metavariable ranging over primitive unary relation variables, κ_1 and κ_2 are metavariables representing any distinct individual terms, and κ_1 and κ_2 have no free occurrences of Ω:[17]

$$Indiscernible(\kappa_1,\kappa_2) \equiv_{df} \forall \Omega(\Omega\kappa_1 \equiv \Omega\kappa_2) \tag{14}$$

Since (14) is much more difficult to read and process than (11), it should now be clear why OT adopts the convention of treating both the free and bound object language variables in definitions as metavariables. So this issue, about how to understand the free and bound variables in definitions in a free logic with both non-denoting, complex individual and relation terms, has now been addressed.

4 The Inferential Role of Definitions-by-≡

Though the general case of a definition-by-≡ has the form $\varphi(\alpha_1,\ldots,\alpha_n) \equiv_{df} \psi(\alpha_1,\ldots,\alpha_n)$, let us abbreviate this more simply as $\varphi \equiv_{df} \psi$ and suppose that this represents any valid instance of the definition, i.e., any instance having the form $\varphi(\tau_1,\ldots,\tau_n) \equiv_{df} \psi(\tau_1,\ldots,\tau_n)$, where $\tau_1,\ldots,\tau_n$ are substitutable for $\alpha_1,\ldots,\alpha_n$, respectively, in ψ.

The first bootstrapping device employed by OT concerns the fact that the equivalence symbol (≡) is not a primitive. Instead $\varphi \equiv \psi$ is defined in the usual way as the conjunction $\varphi \rightarrow \psi \,\&\, \psi \rightarrow \varphi$, where the conjunction $\varphi \,\&\, \varphi$ has been previously defined as $\neg(\varphi \rightarrow \neg\psi)$. So it isn't useful to formulate the inferential role of definitions-by-equivalence in terms of inferences involving the biconditional until we have proven the tautologies (governing ≡ and &) that allow us to reason from those biconditionals. Consequently, we first specify the inferential role of definitions-by-equivalent in terms of conditionals, as captured by the following primitive metarule of inference, in which the *closures* of a formula χ to be any formula in which χ is immediately prefaced by 0 or more strings of quantifiers $\forall\alpha$, modal operators $\Box$ (and actuality operators $\mathcal{A}$):

Rule of Definition by Equivalence
A definition-by-≡ of the form $\varphi \equiv_{df} \psi$ introduces the closures of $\varphi \rightarrow \psi$ and $\psi \rightarrow \varphi$ as necessary axioms.

[17] The restriction that κ_1 and κ_2 have no free occurrences of Ω precludes instances of the definition in which free occurrences of Ω in the κ_i get *captured* by the quantifier $\forall\Omega$ in the definiens.

So, to take an example, reconsider definition (1):

$$\mathit{ContingentlyExemplifies}(x,F) \equiv_{df} Fx \,\&\, \neg\Box Fx$$

By the Rule of Definition by Equivalence, (1) introduces all of the following converse pairs of conditionals (plus others) as axioms:

- $\mathit{ContingentlyExemplifies}(x,F) \to Fx \,\&\, \neg\Box Fx$
  $Fx \,\&\, \neg\Box Fx \to \mathit{ContingentlyExemplifies}(x,F)$
- $\forall F\forall x(\mathit{ContingentlyExemplifies}(x,F) \to Fx \,\&\, \neg\Box Fx)$
  $\forall F\forall x\big(Fx \,\&\, \neg\Box Fx \to \mathit{ContingentlyExemplifies}(x,F)\big)$
- $\Box(\mathit{ContingentlyExemplifies}(x,F) \to Fx \,\&\, \neg\Box Fx)$
  $\Box\big(Fx \,\&\, \neg\Box Fx \to \mathit{ContingentlyExemplifies}(x,F)\big)$
- $\mathcal{A}(\mathit{ContingentlyExemplifies}(x,F) \to Fx \,\&\, \neg\Box Fx)$
  $\mathcal{A}\big(Fx \,\&\, \neg\Box Fx \to \mathit{ContingentlyExemplifies}(x,F)\big)$
- $\Box\forall F\forall x(\mathit{ContingentlyExemplifies}(x,F) \to Fx \,\&\, \neg\Box Fx)$
  $\Box\forall F\forall x\big(Fx \,\&\, \neg\Box Fx \to \mathit{ContingentlyExemplifies}(x,F)\big)$
- etc.

Once the tautologies governing & and ≡ are derived in OT, one can then derive, from the first of the above pairs of conditionals, the biconditional claims intuitively implied by a definition-by-equivalence. And using the logic of $\forall$, $\Box$, and $\mathcal{A}$, one can derive quantified, modal, and actualized biconditionals from the remaining pairs of conditionals listed above.

The Rule of Definition by Equivalence thus preserves the traditional understanding of this type of definition. And, in doing so, it preserves hyperintensionality in OT's theory of relations. OT's theory of relations includes the following definitions for relation identity, which are given by cases. For the case of unary relations F and G:

$$F = G \equiv_{df} \Box\forall x(xF \equiv xG) \tag{15}$$

For the case of n-ary relations generally, where $n \geq 2$:

$$\begin{aligned} F = G \equiv_{df} \forall y_1 \ldots \forall y_{n-1}\big(&[\lambda x\, Fxy_1 \ldots y_{n-1}] = [\lambda x\, Gxy_1 \ldots y_{n-1}] \,\&\\ &[\lambda x\, Fy_1xy_2 \ldots y_{n-1}] = [\lambda x\, Gy_1xy_2 \ldots y_{n-1}] \,\&\, \ldots \,\&\\ &[\lambda x\, Fy_1 \ldots y_{n-1}x] = [\lambda x\, Gy_1 \ldots y_{n-1}x]\big) \end{aligned} \tag{16}$$

Intuitively, this tells us that if each of the n ways of plugging any given $n-1$ objects into the same argument positions of F and G results in identical properties, then F and G are identical. For the case of 0-ary relations, i.e., propositions:

$$p = q \equiv_{df} [\lambda x\, p] = [\lambda x\, q] \tag{17}$$

Definitions (16) and (17) reduce relation and proposition identity to definition (15) of property identity.

Intuitively, relations as have two extensions in OT. One extension, the *exemplification* extension, may vary from world to world, while a second extension, the *encoding* extension, does not. (15) tells us that properties (= unary relations) that necessarily have the same the same encoding extension are identical. But in OT, properties that necessarily have the same exemplification extension, i.e., properties such that $\Box\forall x(Fx \equiv Gx)$, need not have the same encoding extension and so need not be identical. Consider the property *being a barber who shaves all and only those who don't shave themselves*, which we may represent as $[\lambda x\, Bx \,\&\, \forall y(Sxy \equiv \neg Syy)]$. This property is necessarily empty. And so is the property *being a colored and colorless dog*, i.e., $[\lambda x\, Cx \,\&\, \neg Cx \,\&\, Dx]$. Though one can prove:

$$\Box\forall z([\lambda x\, Bx \,\&\, \forall y(Sxy \equiv \neg Syy)]z \equiv [\lambda x\, Cx \,\&\, \neg Cx \,\&\, Dx]z)$$

it doesn't follow in OT that:

$$[\lambda x\, Bx \,\&\, \forall y(Sxy \equiv \neg Syy)] = [\lambda x\, Cx \,\&\, \neg Cx \,\&\, Dx]$$

This hyperintensionality extends, by definitions (16) and (17), to n-ary relations for $n \geq 2$ and for $n = 0$. For example, in the latter case, the easily-provable fact that $\Box(p \equiv \neg\neg p)$ doesn't entail the fact that $p = \neg\neg p$. One may consistently assert, for some or all propositions q that $q \neq \neg\neg q$.

The Rule of Definition by Equivalence preserves these facts. Given the inferential role it stipulates for definitions-by-equivalence, one may not validly substitute necessarily equivalent formulas within λ-expressions. For example, given $\Box\forall x(\varphi \equiv \psi)$, one may not infer $[\lambda x\, \psi]a$ from $[\lambda x\, \varphi]a$.

Before we turn to the inferential role of definitions-by-identity, it is worth pointing out how one can formulate definitions-by-equivalence so as to make sure that if an empty term fills one of the argument places in the definiendum, the resulting definiendum is provably false. As a purely illustrative example, suppose one stipulated that a property *is conditionally necessary for* an object just in case the object necessarily exemplifies the property whenever it exemplifies the property. Then we would formalize this definition as:

$$CondNecFor(F,x) \equiv_{df} Fx \rightarrow \Box Fx \tag{18}$$

Given that the variables function as metavariables, then the following would be a perfectly good instance of the definition, where P is a property constant and $\imath z\psi_1$ is again the definite description introduced above that is provably empty:

$$CondNecFor(P,\imath z\psi_1) \equiv_{df} P\imath z\psi_1 \rightarrow \Box P\imath z\psi_1 \tag{19}$$

So the Rule of Definition by Equivalence tells us that, given (19), the following claim (among others) is axiomatic:

$$CondNecFor(P,\imath z\psi_1) \equiv P\imath z\psi_1 \rightarrow \Box P\imath z\psi_1 \tag{20}$$

Now since we can prove $\neg\imath z\psi_1\!\downarrow$, it follows by negative free logic that $\neg P\imath z\psi_1$. This in turns lets us prove the right-side condition of (20), by failure of the antecedent. Hence, by logic alone, one can use (20) to establish $CondNecFor(P,\imath z\psi_1)$. But intuitively, the *garbage in*, *garbage out* principle suggests that, in cases like this, our definition should allow us to infer that the definiendum fails to hold, i.e., that $\neg CondNecFor(P,\imath z\psi_1)$.

To achieve this, one simply has to reformulate (18) as follows:

$$CondNecFor(F,x) \equiv_{df} x\!\downarrow \& \,(Fx \rightarrow \Box Fx) \tag{18$'$}$$

Thus, (19) is no longer an instance, but instead we have:

$$CondNecFor(P,\imath z\psi_1) \equiv_{df} \imath z\psi_1\!\downarrow \& \,(P\imath z\psi_1 \rightarrow \Box P\imath z\psi_1)$$

We can not then invoke the Rule of Definition by Equivalence to establish $CondNecFor(P,\imath z\psi_1)$ since we can prove the negation of the first conjunct of the definiens. Instead, given the Rule, it becomes provable that $\neg CondNecFor(P,\imath z\psi_1)$

However, we're not quite done, for consider a (provably) non-denoting λ-expression such as the one leading to the Clark/Boolos paradox, namely $[\lambda x\,\exists F(xF \,\&\, \neg Fx)]$. Again, abbreviate this as $[\lambda x\,\varphi_1]$, so that it is provable that $\neg[\lambda x\,\varphi_1]\!\downarrow$. Now consider the following instance of (18′), where a is an individual constant:

$$CondNecFor([\lambda x\,\varphi_1],a) \equiv_{df} a\!\downarrow \& \,([\lambda x\,\varphi_1]a \rightarrow \Box[\lambda x\,\varphi_1]a)$$

Then the Rule of Definition by Equivalence tells us that the following is axiomatic:

$$CondNecFor([\lambda x\,\varphi_1],a) \equiv a\!\downarrow \& \,([\lambda x\,\varphi_1]a \rightarrow \Box[\lambda x\,\varphi_1]a) \tag{21}$$

But one can still prove that $CondNecFor([\lambda x\,\varphi_1],a)$, contrary to intuition: the first conjunct of right-side condition of (21), namely, $a\!\downarrow$, is an axiom (given a is a primitive constant), and the second conjunct of the right-side condition of (21), namely, $[\lambda x\,\varphi_1]a \rightarrow \Box[\lambda x\,\varphi_1]a$ is still provable by failure of the antecedent (given that $\neg[\lambda x\,\varphi_1]\!\downarrow$, our negative free logic yields $\neg[\lambda x\,\varphi_1]a$).

So, the way to formulate definitions-by-equivalence when the definiens doesn't imply the significance of the terms substitutable for the free variables is to make sure that, for every such free variable, there is a conjunct asserting existence. Thus, (18) and (18′) are both properly formulated as:

$$CondNecFor(F,x) \equiv_{df} F\!\downarrow \& \,x\!\downarrow \& \,(Fx \rightarrow \Box Fx) \tag{18$''$}$$

This rules out the problematic instances, and lets us preserve the idea that if you put garbage in (i.e., instantiate to empty terms), then you can prove there is garbage out (i.e., prove that the resulting definiendum is false).

To prepare for the next section, note that the definitions of $x = y$ and $F = G$ in OT are definitions-by-equivalence. Now that we have a good grasp of the inferential role of such definitions, we may start to prove facts about identity. For example, it should be easy to see that from the definition of property identity (15), one can prove that identity for properties is reflexive. For by logic alone we know $xF \equiv xF$. So by GEN, $\forall x(xF \equiv xF)$. And by the Rule of Necessitation, $\Box\forall x(xF \equiv xF)$. But the Rule of Definition by Equivalence tells us that, if we replace G by F in (15), then it is axiomatic that $F = F \equiv \Box\forall x(xF \equiv xF)$. Hence $F = F$, and so by GEN, this holds generally for all properties. This theorem, plus the axiom of OT asserting that in any context, one may substitute identicals, yields that identity for properties is symmetric and transitive. Analogous results can be obtained from the definition of identity for individuals, and the definitions of identity for n-ary relations, for $n \geq 2$ and $n = 0$. With these facts in hand, we can now formulate the inferential role of definitions-by-identity.

5 The Inferential Role of Definitions-by-=

The presence of modal operators, rigid definite descriptions, and empty terms presents challenges for any theory of definition-by-identity. In subsection 5.1, we lay out the problems, and in subsection 5.2, formulate an inferential role for definitions-by-identity.

5.1 Problems for the Classical Theory

5.1.1 The Problem of Modality

In a classical, non-modal, first-order predicate calculus with identity, but *without* definite descriptions, the classical theory of definitions provides a method for introducing a new individual constant, say δ, into the system. The method is to use a definition-by-equivalence, as follows. When $\exists!x\varphi$ (i.e., there is a unique x such that φ) is provable for some formula φ in which x is the only free variable, one may introduce, by definition, a new individual constant, say δ, to designate the object satisfying φ, as follows:

(a) $\delta = x \ \equiv_{df} \ \varphi$

(cf. Suppes 1957, 159–60; Gupta 2023, Section 2.4). If $\delta\downarrow$ is defined as $\exists y(\delta = y)$, one can prove that $\delta\downarrow$ from (a) and the fact that $\exists!x\varphi$.[18] Moreover, the definition (also correctly) allows one to eliminate δ from any formula in which it subsequently occurs.[19]

But this method of introducing new individual constants would be disastrous for OT, given its modal logic and theory of identity implies identity for individuals is necessary, i.e., $x = y \rightarrow \Box(x = y)$. To see the potential disaster, consider the following example, in which we may suppose that the following two claims are provable as theorems:

(b) $\exists!x\varphi$

(c) $\Diamond\neg\exists x\varphi$

Since (b) is a theorem, the classical theory would allow one to stipulate definition (a). Definition (a) would, in OT, allow one to take the (closures of) the biconditional $\delta = x \equiv \varphi$ as axioms. So, for example, one would be able to take the following as axioms:

(d) $\forall x(\delta = x \equiv \varphi)$

(e) $\forall x\Box(\delta = x \equiv \varphi)$

But these would allow one to derive $\Box\exists x\varphi$,[20] which contradicts (c). Clearly, if well-formed definitions introduce contradictions, then something has gone wrong.

One might suggest here that in order to use (a) as a definition in OT, we have to require more than just $\vdash \exists!x\varphi$. Instead, the suggestion goes, the condition $\vdash \Box\exists!x\varphi$ is required. Such a condition would block the example we've been discussing since (c) couldn't be a theorem if $\vdash \Box\exists!x\varphi$.

---

[18] The definition licenses the axiom:

(ϑ) $\forall x(\delta = x \equiv \varphi)$,

since this is a universal closure of the equivalence $\delta = x \equiv \varphi$. Since $\exists!x\varphi$ is a theorem, let b be such an object, so that we know both φ^b_x and $\forall y(\varphi^y_x \rightarrow y = b)$. Instantiating ($\vartheta$) to b, it follows that $\delta = b \equiv \varphi^b_x$. Since we know φ^b_x, it follows that $\delta = b$ and, by symmetry, But then generalizing on b, it follows that $\exists y(\delta = y)$. Hence $\delta\downarrow$.

[19] To see this, we show that ψ^δ_x (i.e., any formula ψ in which δ has been substituted for all the free occurrences of x) is equivalent to $\exists!x\varphi \,\&\, \exists x(\varphi \,\&\, \psi)$. ($\rightarrow$) Assume ψ^δ_x. Since we've just seen that the definition of δ implies $\delta\downarrow$, we can conclude $\delta = \delta \equiv \varphi^\delta_x$ by instantiating δ into $\forall x(\delta = x \equiv \varphi)$ (a universal claim licensed by definition (a) in the text). And we can independently infer $\delta = \delta$ by instantiating δ into $\forall x(x = x)$, which is obtained by GEN from the fact that identity is reflexive. Hence, φ^δ_x. Conjoining what we know, we have $\varphi^\delta_x \,\&\, \psi^\delta_x$. Hence, by $\exists$I, $\exists x(\varphi \,\&\, \psi)$. But, by hypothesis, $\exists!x\varphi$ is a theorem. Hence, $\exists!x\varphi \,\&\, \exists x(\varphi \,\&\, \psi)$. ($\leftarrow$) Assume $\exists!x\varphi \,\&\, \exists x(\varphi \,\&\, \psi)$, to show ψ^δ_x. Let a be a witness to the first conjunct and let b be a witness to the second conjunct, so that we know both:

(ζ) $\varphi^a_x \,\&\, \forall y(\varphi^y_x \rightarrow y = a)$

(ξ) $\varphi^b_x \,\&\, \psi^b_x$

Then by the first conjunct of (ζ) and the definition of δ, it follows that $\delta = a$. It follows from this and the second conjunct of (ζ) that $\forall y(\varphi^y_x \rightarrow y = \delta)$. But this and the first conjunct of (ξ) imply $b = \delta$. So we may substitute δ for b in the second conjunct of (ξ) to conclude ψ^δ_x.

[20] To see why, suppose b is some witness to (b), so that we know φ^b_x and $\forall y(\varphi^y_x \rightarrow y = b)$, by the definition of the uniqueness quantifier $\exists!x\varphi$. Then by (d) it follows that:

(ϑ) $\delta = b \equiv \varphi^b_x$

and by (e) it follows that:

(ξ) $\Box(\delta = b \equiv \varphi^b_x)$

Now by the reasoning in footnote 18, definition (a) implies $\delta\downarrow$. Hence we can instantiate the necessity of identity (a theorem mentioned above) to infer:

(ζ) $\delta = b \rightarrow \Box\delta = b$

So we can establish $\varphi^b_x \rightarrow \Box\varphi^b_x$ by a hypothetical syllogism chain, as follows: $\varphi^b_x \rightarrow \delta = b$, by ($\vartheta$); $\delta = b \rightarrow \Box\delta = b$, by ($\zeta$); and $\Box\delta = b \rightarrow \Box\varphi^b_x$, by ($\xi$) and the modal theorem $\Box(\psi \equiv \chi) \rightarrow (\Box\psi \equiv \Box\chi)$. Having thus established that $\varphi^b_x \rightarrow \Box\varphi^b_x$, then since we know φ^b_x, it follows that $\Box\varphi^b_x$. Hence, by $\exists$I, $\exists x\Box\varphi$. So by the Buridan formula, $\Box\exists x\varphi$, which contradicts (c).

But this suggestion for preserving the classical theory doesn't work. We can see why, at least intuitively, if we temporarily speak in the familiar idiom of semantically-primitive possible worlds. Suppose it were a theorem that $\Box\exists!x\varphi$ and suppose there were just two possible worlds, $\boldsymbol{w}_\alpha$ (the actual world) and $\boldsymbol{w}_1$, and two distinct objects a and b such that a is uniquely φ at $\boldsymbol{w}_\alpha$ and b is uniquely φ at $\boldsymbol{w}_1$. In this modal situation, an equivalence licensed by definition (a), namely $\delta = a \equiv \varphi^a_x$, would fail to be necessary. Since the terms of OT are rigid, the definition would introduce δ as a rigid designator of a, since a is uniquely φ at $\boldsymbol{w}_\alpha$. So $\delta = a$ would be true at $\boldsymbol{w}_1$, since δ rigidly denotes a. But φ^a_x would be false at $\boldsymbol{w}_1$ since, by hypothesis, b is uniquely φ at $\boldsymbol{w}_1$. Hence, the equivalence $\delta = a \equiv \varphi^a_x$ would fail to be true at $\boldsymbol{w}_1$. So the universalized modal equivalence (e), which is licensed by definition (a), can't be true, since it has false instances.

But another suggestion along these lines presents itself, namely, that definition (a) becomes legitimate if we require that $\exists!x\Box\varphi$, instead of $\Box\exists!x\varphi$, be a theorem. Unfortunately, this suggestion fails as well. That's because $\exists!x\Box\varphi$ can be true while $\exists!x\varphi$ is not. Intuitively, from the fact that there is exactly one thing which is φ at every possible world, it doesn't follow that there is exactly one thing which is in fact φ. Suppose there were just two things a and b, and just two worlds $\boldsymbol{w}_\alpha$ and $\boldsymbol{w}_1$, and that a exemplifies P at both $\boldsymbol{w}_\alpha$ and $\boldsymbol{w}_1$, and that b exemplifies P only at $\boldsymbol{w}_\alpha$. Then, in that modal situation, at $\boldsymbol{w}_\alpha$, there is exactly one object (namely a) that exemplifies P at every world, i.e., $\exists!x\Box Px$. But it is not the case, at $\boldsymbol{w}_\alpha$, that there is exactly one thing that exemplifies P, since both a and b exemplify P there. Thus, given only that $\vdash \exists!x\Box\varphi$, we can't *define* δ by saying $\delta = x \equiv_{df} \varphi$, since $\vdash \exists!x\Box\varphi$ doesn't guarantee $\vdash \exists!x\varphi$.

By now, it may be apparent that if one wants to stipulate $\delta = x \equiv_{df} \varphi$ in a modal context, the conditions $\vdash \exists!x\Box\varphi$ and $\vdash \exists!x\varphi$ are both required. But we shall not adapt the classical theory of definitions by introducing new individual constants in this way. Instead, we shall take advantage of the fact that definite descriptions are part of the language of OT, interpreted rigidly. We may then introduce a new individual constant when we know that $\vdash \imath x\varphi\downarrow$ and x is the sole variable that occurs free in φ. For then, we may use the following definition-by-identity:

(a′) $\delta =_{df} \imath x\varphi$

This introduces the rigidly-designating constant δ by way of a significant, rigidly-designating description. It blocks the examples that were problematic for the classical theory because (a′) doesn't license the equivalence $\delta = x \equiv \varphi$ as axiomatic. Instead, as we'll see in the next subsection, (a′) will imply the identity $\delta = \imath x\varphi$ when it is known, by proof or by hypothesis, that $\imath x\varphi\downarrow$. And we'll see that one need not require that there be a proof of $\imath x\varphi\downarrow$ to introduce (a′)—the inferential role of definitions by identity will implicitly introduce axioms that assert both (a) if $\imath x\varphi\downarrow$, then $\delta = \imath x\varphi$, and (b) if $\neg\imath x\varphi\downarrow$, then $\neg\delta\downarrow$. So we shall defer further discussion of the inferential role of (a′) until Section 5.2.

However, two further observations are in order. The first is that OT has an analogous procedure for introducing a new *relation* constant, except that relation constants are to be introduced in by λ-expressions instead of by definite descriptions. A definition-by-identity would take the following form, where Π is a new n-ary relation constant ($n \geq 1$) and, for the moment, we assume that $[\lambda x_1 \ldots x_n\, \varphi]$ is a λ-expression with no free variables and for which $\vdash [\lambda x_1 \ldots x_n\, \varphi]\downarrow$:

$$\Pi =_{df} [\lambda x_1 \ldots x_n\, \varphi]$$

For example, where $E!$ is the property *being concrete*, we might introduce an *ordinary* object ($O!$) as one that is possibly concrete, and introduce an *abstract* object ($A!$) as one that couldn't possibly be concrete:

$$O! =_{df} [\lambda x\, \Diamond E!x]$$

$$A! =_{df} [\lambda x\, \neg\Diamond E!x]$$

These definitions introduce new unary relation *constants*. In OT, both of the definiens denote properties. Both properties have an exemplification extension that varies from world to world. And both have an encoding extension among the abstract objects that encode them, as given by OT's comprehension principle for abstract objects.

The second observation is that the foregoing remarks about the definition of new individual and relation constants have to be generalized in two ways: (a) we need to consider the case in which the definiens may fail to denote, and (b) we need to consider the case where there are free variables in the definiens and definiendum. These cases give rise to new and interesting issues, to which we now turn.

5.1.2 The Problem of Empty Terms in Defined Operations

Should one allow definitions-by-identity for which the definiens fails to denote? If so, what have we defined and what is the inferential role of such a definition? To keep these questions simple, let's temporarily suppose there are no free variables in the definiens and definiendum. So consider two definitions of the form $\tau =_{df} \sigma$, where ψ_1 is again used to abbreviate $Pz \,\&\, \neg Pz$ and φ_1 is again used to abbreviate the formula $\exists F(xF \,\&\, \neg Fx)$:

$$e =_{df} \imath z\psi_1 \tag{22}$$

$$T =_{df} [\lambda x\, \varphi_1] \tag{23}$$

Neither definiens has a denotation and, as we've previously noted, in OT one can prove both $\neg \imath z\psi_1\!\downarrow$ and $\neg[\lambda x\, \varphi_1]\!\downarrow$. So, if the inferential role of a definition by identity were to implicitly introduce identity statements as axioms, then we would surely want to avoid the above definitions, since a negative free logic should also imply that $e = \imath z\psi_1$ and $T = [\lambda y\, \varphi_1]$ are false if even one term flanking the identity sign fails to denote.

But suppose we formulate the inferential role of a definition-by-identity differently, so that such definitions implicitly introduce axioms that assert both (a) that the identity holds when the definiens denotes and (b) that the definiendum fails to denote when the definiens fails to denote. Then definitions like the (22) and (23) are just harmless. When we formulate the Rule of Definition by Identity in Section 5.2, it will assert that (22) and (23) implicitly introduce the following axioms, respectively:

$$(\imath z\psi_1\!\downarrow \to e = \imath z\psi_1) \,\&\, (\neg \imath z\psi_1\!\downarrow \to \neg e\!\downarrow) \tag{24}$$

$$([\lambda x\, \varphi_1]\!\downarrow \to T = [\lambda x\, \varphi_1]) \,\&\, (\neg[\lambda x\, \varphi_1]\!\downarrow \to \neg T\!\downarrow) \tag{25}$$

Clearly, the theorems that $\neg \imath z\psi_1\!\downarrow$ and $\neg[\lambda x\, \varphi_1]\!\downarrow$ respectively trigger the second conjuncts of (24) and (25) and so the only effect of allowing definitions such as (22) and (23) is to let us additionally prove, as theorems, that $\neg e\!\downarrow$ and $\neg T\!\downarrow$. In a negative free logic, such claims would seem desirable.

With this in mind, we can now turn to the more interesting (and classically puzzling) case of definitions-by-identity in which there are corresponding free variables in the definiendum and definiens. Given the general form of definitions-by-identity, as described in Section 2, the ones with free variables take the following form, where $\alpha_1, \ldots, \alpha_m$ $(m \geq 1)$ occur free:

$$\tau(\alpha_1, \ldots, \alpha_m) =_{df} \sigma(\alpha_1, \ldots, \alpha_m)$$

In a system that allows empty terms, one may not suppose that the inferential role of such definitions is to introduce (the closures of) axioms asserting the identity of the definiendum and definiens. The simple problem that arises is most easily seen if we consider a 'classical' theory such as the theory of real numbers. Let's suppose, only for illustrative purposes, that this theory has been formulated in the language and logic of OT.[21] In real number theory, mathematicians want to define division, i.e., x/y, in a way that ignores the case where y is 0, so as to avoid or ignore terms like $3/0$, $3/(\pi - \pi)$, $3/(3/0)$, etc. So, they might offer a conditional definition (cf. Suppes 1957, 165–169):

If (it is a theorem that) $y \neq 0$, then $x/y \;=_{df}\; \imath z(x = y \cdot z)$

But if this definition is supposed to conservatively extend real number theory with new expressions and axioms, this it is a somewhat awkward way of doing so, for the question of eliminability arises. What is the status of terms like '3/0'? Are they part of the language or not? How could they be part of the language of real number if the logic isn't free? If they are not part of the language, then how does one specify the expanded language of real number theory (i.e., the one represented by the definition) so as to include terms of the form x/y only when $y \neq 0$? (Any general specification of the language would typically allow terms such as $3/0$, $3/(\pi - \pi)$, $3/(3/0)$, etc., to be well-formed.) It is not so easy to specify the language to include those terms only if there is a proof that their denominator is not identical to 0, since one typically specifies the proof system after specifying the language. One could perhaps, as a heroic way out, specify a sequence of language and proof system pairs, so that at each pair in the sequence, the proof system of that pair is used to

[21] OT doesn't include the primitives needed to formulate real number theory and, indeed, doesn't include any mathematical primitives. But Nodelman & Zalta (forthcoming) show that one can derive 2nd-order Peano Arithmetic in OT. And Simpson 1999 [2009] describes a known way of using 2nd-order Peano Arithmetic to reconstruct real number theory, though in his reconstruction, the real numbers are not individuals. In any case, OT doesn't officially include the primitive notions and axioms of real number theory or any other mathematical theory.

specify the language of the next pair. But this hardly seems like a good solution to the problem of defining division.

What is needed is a way to state the definition in a completely general way so that (a) it doesn't yield identities in which one of the terms involves a division by 0, and (b) it gives one the ability to prove that such facts as $\neg(3/0)\downarrow$, $\neg(3/(\pi-\pi))\downarrow$, $\neg(3/(3/0))\downarrow$, etc. We'll see in Section 5.2 how the Rule of Definition by Identity solves this problem. This solution starts by admitting that real number theory is most naturally expressed in a (free) logic that allows for complex terms that fail to have a denotation, such as $3/0$, $3/(\pi-\pi)$, $3/(3/0)$, etc.

5.2 The Metarule for Definitions-by-Identity

As with definitions-by-equivalence, the inferential role in OT of definitions-by-identity is specified in the form of a metarule. To simplify the presentation, consider a definition-by-identity in which the definiens and definiendum have two free variables, α_1 and α_2. So the definition has the form $\tau(\alpha_1,\alpha_2) =_{df} \sigma(\alpha_1,\alpha_2)$. Now consider any terms τ_1 and τ_2 that are of the same type as, and are substitutable for, α_1 and α_2, respectively, in $\sigma(\alpha_1,\alpha_2)$. Then where $\tau(\tau_1,\tau_2)$ and $\sigma(\tau_1,\tau_2)$ are the result of substituting the τ_i for all the free occurrences of the α_i in $\sigma(\alpha_1,\alpha_2)$ and $\tau(\alpha_1,\alpha_2)$, respectively, we can formulate the Rule of Definition by Identity as follows:

> **Rule of Definition by Identity** (Two-Free Variables)
> Whenever τ_1 and τ_2 are any terms substitutable, respectively, for α_1 and α_2 in $\sigma(\alpha_1,\alpha_2)$, then a definition of the form $\tau(\alpha_1,\alpha_2) =_{df} \sigma(\alpha_1,\alpha_2)$ introduces (the closures of) the following axiom schema:
>
> $$\big(\sigma(\tau_1,\tau_2)\downarrow \rightarrow \tau(\tau_1,\tau_2) = \sigma(\tau_1,\tau_2)\big) \,\&\, \big(\neg\sigma(\tau_1,\tau_2)\downarrow \rightarrow \neg\tau(\tau_1,\tau_2)\downarrow\big)$$

To see this rule in action, consider the definition of division in real number theory. If we let τ_1 be x and τ_2 be y, so that $\sigma(\tau_1,\tau_2)$ is $\imath z(x = y\cdot z)$ and $\tau(\tau_1,\tau_2)$ is x/y, then the Rule of Definition by Identity specifies that the definition $x/y =_{df} \imath z(x = y\cdot z)$ would introduce the (closures of the) axiom:

$$\big(\imath z(x=y\cdot z)\downarrow \rightarrow x/y = \imath z(x=y\cdot z)\big) \,\&\, \big(\neg\imath z(x=y\cdot z)\downarrow \rightarrow \neg(x/y)\downarrow\big)$$

So, in the case where τ_1 is x and τ_2 is 0, the axiom would be:

$$\big(\imath z(x=0\cdot z)\downarrow \rightarrow x/0 = \imath z(x=0\cdot z)\big) \,\&\, \big(\neg\imath z(x=0\cdot z)\downarrow \rightarrow \neg(x/0)\downarrow\big)$$

Since the antecedent of the second conjunct is a theorem of real number theory (for arbitrary x, there is no unique object z such that $x = 0\cdot z$),[22] it follows that $\neg(x/0)\downarrow$, and so by GEN, $\forall x\neg(x/0)\downarrow$ is a theorem. And similarly when τ_2 is $\pi-\pi$. Moreover, in the case where τ_2 is $3/0$, one can show that $\neg(x/(3/0))\downarrow$, for any x, since in this case, the rule asserts that the following is axiomatic:

$$\big(\imath z(x=(3/0)\cdot z)\downarrow \rightarrow x/(3/0) = \imath z(x=(3/0)\cdot z)\big) \,\&\, \big(\neg\imath z(x=(3/0)\cdot z)\downarrow \rightarrow \neg(x/(3/0))\downarrow\big)$$

Since $\neg(3/0)\downarrow$, $(3/0)\cdot z$ is provably empty, for any z.[23] So $x = (3/0)\cdot z$ is always false, for any x, implying thereby that $\neg\imath z(x = (3/0)\cdot z)\downarrow$. Thus, the rule yields $\neg(x/(3/0))\downarrow$, for any x.

This shows that the Rule of Definition by Identity handles the definition of division in real number theory in a general way—it extends the language with new terms of the form κ/κ', for arbitrary individual terms κ and κ'; it asserts that the definition yields identities when the definiens is significant; and it allows us to prove that the definiendum is empty when the definiens is empty.

We can also verify that the Rule of Definition by Identity correctly handles the case of a non-denoting description ($\imath z\psi_1$) in the example definition-by-= (8) introduced earlier. The Rule then asserts that the inferential role of the definition $\iota_{\imath z\psi_1} =_{df} \imath x(x = \imath z\psi_1)$ is to implicitly introduce the closures of the following axioms:

$$\big(\imath x(x=\imath z\psi_1)\downarrow \rightarrow \iota_{\imath z\psi_1} = \imath x(x=\imath z\psi_1)\big) \,\&\, \big(\neg\imath x(x=\imath z\psi_1)\downarrow \rightarrow \neg\iota_{\imath z\psi_1}\downarrow\big)$$

Since $\neg\imath z\psi_1\downarrow$ is provable, so is $\neg\imath x(x=\imath z\psi_1)\downarrow$, which triggers the second conjunct, thereby allowing one to infer $\neg\iota_{\imath z\psi_1}\downarrow$. So we can prove that the definiendum fails to denote when the definiens fails to denote. In OT,

[22] Even when $x=0$, there is no unique real number z such that $0=0\cdot z$. Every real number satisfies this formula.

[23] For reductio, suppose $((3/0)\cdot z))\downarrow$. Then by the axioms for multiplication, $(3/0)\cdot z = 3z/0$. But by the previous case we examined in the text, we saw that $\neg(x/0)\downarrow$, for any x, and so $\neg(3z/0)\downarrow$. Hence, $\neg((3/0)\cdot z = 3z/0)$, by the contrapositive of a theorem that asserts that if an identity is true, then the terms flanking the identity denote. Contradiction. Thus, $\neg((3/0)\cdot z)\downarrow$. Alternatively, just assume that the formulation of real number theory under consideration allows for empty complex terms, so that multiplication is axiomatized in such a way that the term $\kappa\cdot\kappa'$ is empty if either κ or κ' is empty.

then, one can reason secure in the knowledge that no true exemplification formula, encoding formula, or identity formula will have $\imath z\psi_1$ as one of the relata in the formula.

Since we have now sufficiently motivated the Rule of Definition by Identity for two free variables, we leave it to the reader to formulate the general form of the Rule for m free variables, for $m \geq 0$. When $m = 0$, the Rule of Definition by Identity stipulates that definitions (22) and (23) implicitly introduce the axioms (24) and (25), respectively. And when $m \neq 0$, the general form of the Rule for m free variables preserves a classical understanding of the inferential role of term-forming operators (i.e., function terms). We can rest assured that term-forming operators introduced into our language by definition are logically well-behaved if their definientia are significant when applied to arguments that match the type of the free variables.

Before we turn to our concluding section, there is one subtlety to discuss; it concerns the question of refining the Rule further should one believe that the *garbage in*, *garbage out* principle is an absolute value. Some might see the following case as a problem. Suppose we introduce the *negation of* an n-ary relation F ($n \geq 0$), written $\overline{F}$, as *being objects* $z_1,\ldots,z_n$ that fail to exemplify F, i.e.,

$$\overline{F} =_{df} [\lambda z_1 \ldots z_n\, \neg Fz_1 \ldots z_n] \tag{26}$$

Now consider an instance of this definition when we substitute the unary property term $[\lambda x\, \exists F(xF \,\&\, \neg Fx)]$ for F (which is functioning as a metavariable). To keep the notation simple, let's again abbreviate the formula $\exists F(xF \,\&\, \neg Fx)$ as φ_1. Then as an instance of (26) we have:

$$\overline{[\lambda x\, \varphi_1]} =_{df} [\lambda z\, \neg[\lambda x\, \varphi_1]z] \tag{27}$$

But recall that the expression $[\lambda x\, \varphi_1]$ provably fails to denote (since otherwise the Clark/Boolos paradox results). So by negative free logic, one can infer $\neg[\lambda x\, \varphi_1]\kappa$, for any individual term κ you please, since the formula $[\lambda x\, \varphi_1]\kappa$ is false when one of its terms fails to denote. But in the latest developments of OT, the definiens $[\lambda z\, \neg[\lambda x\, \varphi_1]z]$ denotes a property even though the embedded λ-expression $[\lambda x\, \varphi_1]$ is empty. Intuitively, the property $[\lambda z\, \neg[\lambda x\, \varphi_1]z]$ denotes (some arbitrarily chosen, hyperintensional) property in the domain of properties that has exemplification conditions specified by the definition of truth for the formula $\neg[\lambda x\, \varphi_1]z$. The definition of truth for a free logic will imply that everything in the domain of individuals satisfies the open formula $\neg[\lambda x\, \varphi_1]z$, since it is provable that nothing in the domain of individuals satisfies the open formula $[\lambda x\, \varphi_1]z$, i.e., it is provable that no z is such that $[\lambda x\, \exists F(xF \,\&\, \neg Fx)]z$, since the λ-expression fails to denote. Hence, the definiens of (27), i.e., $[\lambda z\, \neg[\lambda x\, \varphi_1]z]$, denotes a *universal* property, and so the Rule of Definition by Identity will imply that it is axiomatic that $\overline{[\lambda x\, \varphi_1]}$ is identical to that property.

So, $\overline{[\lambda x\, \varphi_1]}$ is a case of what we might call an *impractical* definiendum. It has a denotation but its definiens, as specified by definition (27), is a term that denotes even though its embedded term $[\lambda x\, \varphi_1]$, which is its argument given definition (26), fails to denote. It should be clear, from the foregoing discussion, that the presence of terms like $[\lambda z\, \neg[\lambda x\, \varphi_1]z]$ don't cause problems. Since it has a denotation, β-Conversion implies:

$$[\lambda z\, \neg[\lambda x\, \varphi_1]z]y \equiv \neg[\lambda x\, \varphi_1]y$$

But for the term $[\lambda x\, \varphi_1]$, β-Conversion implies only:

$$[\lambda x\, \varphi_1]\downarrow \rightarrow \big([\lambda x\, \varphi_1]y \equiv \varphi_1{}^{y}_{x}\big)$$

i.e.,

$$[\lambda x\, \varphi_1]\downarrow \rightarrow \big([\lambda x\, \varphi_1]y \equiv \exists F(yF \,\&\, \neg Fy)\big)$$

The antecedent to this claim is never triggered, since $[\lambda x\, \varphi_1]$ provably fails to denote. So we can't conclude from the fact that there are objects y such that $\exists F(yF \,\&\, \neg Fy)$ that $[\lambda x\, \varphi_1]y$. The latter would, contrary to fact, imply that $[\lambda x\, \varphi_1]$ denotes.

Since we already have terms, such as $[\lambda z\, \neg[\lambda x\, \varphi_1]z]$, that denote even though they have a non-denoting subterm, we may as well allow them to be definientia so as to introduce impractical definienda. So I've chosen to forego further refinement of the Rule of Definition by Identity in the attempt to have it yield an axiom from which one can derive $\neg\overline{[\lambda x\, \varphi_1]}\downarrow$ as a theorem. I don't regard the preservation of the *garbage in*, *garbage out* principle as an absolute value. OT is a system that lives with impractical terms and lives with such facts as that $[\lambda z\, \neg[\lambda x\, \neg\varphi_1]z]$ denotes a universal property. This term, and other impractical relation terms, have perfectly well-defined exemplification conditions and are logically well-behaved in derivations.[24]

[24] There are also impractical individual terms in OT as well. For example, consider the

6 Some Final Considerations

There are other subtleties in connection with definitions in a hyperintensional free logic. These are:

- One may introduce a definition-by-identity, say $\tau =_{df} \sigma$, for which the proof that the definiens denotes (i.e., the proof of $\sigma\downarrow$) rests on a contingent axiom or a contingent theorem. But in that case, any derivation (i.e., from the axiom implicitly introduced by the definition as prescribed in the Rule of Definition by Identity) that makes use of $\sigma\downarrow$ will similarly depend on the contingent axiom or theorem needed for the proof of $\sigma\downarrow$. Thus, the Rule of Necessitation can't be applied to the result of such derivations.

- New 0-ary relation terms can be introduced either by a definition-by-equivalence or by a definition-by-identity. Which definition one chooses to use depends on the inferential role that one wants the definiendum to have. For example, one might stipulate $q_0 =_{df} \forall p(p \to p)$ and thereby appeal to the identity $q_0 = \forall p(p \to p)$. This would allow one to substitute $q_0 = \forall p(p \to p)$ in any context. But if one stipulated instead $q_0 \equiv_{df} \forall p(p \to p)$, then given the hyperintensionality of relations, one may not infer the identity $q_0 = \forall p(p \to p)$ from the necessarily equivalence $\Box(q_0 \equiv \forall p(p \to p))$.

- The axiom of OT's negative free logic asserting that constants and variables have denotations, has to be formulated carefully. If empty definientia are used to introduce empty definienda in a definition-by-identity, then one must make sure that the axiom in question asserts only that *primitive* constants (and variables) denote. That way, the defined constant a can be introduced by the definition $a =_{df} \imath x(Px \& \neg Px)$ without worrying that the relevant axiom of negative free logic will then assert that it has a denotation.

definition:

$$\boldsymbol{a}_y \ =_{df} \ \imath x(A!x \,\&\, \forall F(xF \equiv Fy))$$

and consider the instance of the definition in which $\imath z\psi_1$ has been substituted for y:

$$\boldsymbol{a}_{\imath z\psi_1} \ =_{df} \ \imath x(A!x \,\&\, \forall F(xF \equiv F\imath z\psi_1))$$

Here, the definiens has a denotation even though $\imath z\psi_1$ fails to denote: since $F\imath z\psi_1$ is false for every F, the definiens denotes the abstract object that encodes no properties! So $\boldsymbol{a}_{\imath z\psi_1}$ is well-defined, but impractical.

Although these subtleties raise interesting issues, we need not discuss them further here. It is sufficient to have seen how to define the inferential roles of definitions-by-equivalence and definitions-by-identity for a system of hyperintensional, second-order, modal, negative free logic (without identity) with an actuality operator governed by a contingent axiom and which includes a second kind of atomic formula and the added expressive power of definite descriptions and λ-expressions.

Bibliography

Belnap, N., 1993, "On Rigorous Definitions", *Philosophical Studies*, 72(2–3): 115–146.

Boolos, G., 1987, "The Consistency of Frege's Foundations of Arithmetic", in J. Thomson (ed.), *On Being and Saying*, Cambridge, MA: MIT Press, 1987, pp. 3–20.

Clark, R., 1978, "Not Every Object of Thought has Being: A Paradox in Naive Predication Theory", *Noûs*, 12: 181–188.

Dudman, V. H., 1973, "Frege on Definition", *Mind*, 83(328): 609–610.

Feferman, S., 1995, "Definedness", *Erkenntnis*, 43: 295–320.

Frege, G., 1879, *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle a. S.: Louis Nebert; translated as *Concept Script, a formal language of pure thought modelled upon that of arithmetic*, by S. Bauer-Mengelberg, in J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Cambridge, MA: Harvard University Press, 1967, 1–82.

Frege, G., 1903a [2013], *Grundgesetze der Arithmetik*, Volume II, Jena: Verlag Hermann Pohle; reprinted, Hildesheim: Georg Olms Verlagsbuchhandlung, 1962; translated and edited by Philip Ebert and Marcus Rossberg, with Crispin Wright, *Basic Laws of Arithmetic*, Oxford: Oxford University Press, 2013. [All quotes in translation are from Ebert and Rossberg 2013.]

Frege, G., 1903b, "Über die Grundlagen der Geometrie" (First Series), *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 12 (1903):

319–324 (Part I), 368–375 (Part II); translated as "The Foundations of Geometry", by M. E. Szabo, *The Philosophical Review*, 69(1) (1960): 3–17; translation by E.-H. W. Kluge, in *On the Foundations of Geometry and Formal Theories of Arithmetic*, New Haven: Yale University Press, 1971, 22–37, and reprinted in *Collected Papers on Mathematics, Logic and Philosophy*, B. F. McGuinness (ed.), Oxford: Blackwell Publsishers, 1984, 273–284.

Frege, G., 1914, "Über Logik in der Mathematik", unpublished lecture note, in H. Hermes, F. Kambartel, and F. Kaulbach (eds.), *Gottlob Frege: Nachgelassene Schriften* (Band 1), Hamburg: Meiner, 1969; translated as "Logic in mathematics", in H. Hermes, F. Kambartel, and F. Kaulbach (eds.), *Gottlob Frege: Posthumous Writings*, Oxford: Basil Blackwell, 1979, 203–250.

Gupta, A., 2023, "Definitions", *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/fall2023/entries/definitions/.

Hodges, W., 2008, "Tarski's Theory of Definitions", in D. Patterson (ed.), *New Essays on Tarski and Philosophy*, Oxford: Oxford University Press, 94–132.

Mates, B., 1972, *Elementary Logic*, 2nd edition, Oxford: Oxford University Press.

Nodelman, U., and E. Zalta, forthcoming, "Number Theory and Infinity Without Mathematics", *Journal of Philosophical Logic*.

Padoa, A., 1900, "Logical introduction to any deductive theory", in J. van Heijenoort (ed.), *From Frege to Goedel: A Source Book in Mathematical Logic, 1879–1931*, Cambridge, MA: Harvard University Press, 119–123.

Pascal, B., 1658, "De l'Esprit Géométrique et de l'Art de Persuader", in *Oeuvres Complètes*, Paris: Seuil, 1963, pp. 348–59.

Porphyry, [OAC], *In Aristotelis Categorias*, in *Porphyrii Isagoge et in Aristotelis Categorias Commentarium*, Adolfus Busse (ed.), Berlin: Reimer, 1887; translated *On Aristotle's Categories*, in Steven K. Strange (trans.), *Porphyry: On Aristotle's Categories*, London: Bloomsbury, 1992; page reference is to the translation.

Simpson, S., 1999 [2009], *Subsystems of Second Order Arithmetic*, second edition, Cambridge: Cambridge University Press, 2009.

Suppes, P., 1957, *Introduction to Logic*, Princeton: Van Nostrand.

Urbaniak, R., and S. Hämäri, 2012, "Busting a Myth about Leśniewski and Definitions", *History and Philosophy of Logic*, 33(2): 159–189.

Zalta, E., m.s., *Principia Logico-Metaphysica*, URL = https://mally.stanford.edu/principia.pdf