G. ALDO ANTONELLI

PROTO-SEMANTICS FOR POSITIVE FREE LOGIC

Received 19 July 1999; revised 10 November 1999

ABSTRACT. This paper presents a bivalent extensional semantics for positive free logic without resorting to the philosophically questionable device of using models endowed with a separate domain of "non-existing" objects. The models here introduced have only one (possibly empty) domain, and a partial reference function for the singular terms (that might be undefined at some arguments). Such an approach provides a solution to an open problem put forward by Lambert, and can be viewed as supplying a version of parametrized truth non unlike the notion of "truth at world" found in modal logic. A model theory is developed, establishing compactness, interpolation (implying a strong form of Beth definability), and completeness (with respect to a particular axiomatization).

KEY WORDS: free logic, existence, denotation, semantics, Craig interpolation, Beth definability.

1. INTRODUCTION

Free logics, first introduced (in fact) by Leonard [18] and (in name) by Lambert [12] in the late 1950's, are characterized by the fact that singular terms do not have existential import. In other words, in a free logic one can have terms such as "Vulcan" or "Pegasus" or "the largest prime" that do not have any referent. This fact can be regarded as the culmination of a trend in logic that has been active throughout the history of the field (see Lambert [16]): since Aristotle, and certainly until the school of Port Royal, logic has regarded *general* terms as having existential import: as a consequence, the inference from "All P's are Q's" to "Some P's are Q's" was regarded as valid. On the other hand, modern symbolic logic, beginning with Frege, has done away with this assumption.

But a certain asymmetry remained, in that the inference from "Everything the same as t is P" to "Something the same as t is P" (where t is a singular term) was still regarded as valid: free logic can be viewed as an effort to remedy this situation, by allowing for the case in which t has no existential import. Going one step further, certain versions of free logic (known as "universally free logics") disallow the inference from "Everything is P" to "Something is P" (corresponding to the case where nothing altogether exists).

Journal of Philosophical Logic **29:** 277–294, 2000.

More recently, free logics have been applied to the representation of facts about computation, where it is natural to regard "the value of f at x" as a non-referential term whenever f is not defined at x. For instance, the classical "axiomatic semantics" for programming languages of Hoare [9] requires a free logical framework (see Gumb and Lambert [8] for a discussion of applications of free logic to computation and, more in general, for further references).

There are fundamentally three ways to develop free logic. In *negative* free logic one assumes that all atomic sentences involving non-referential terms are false. In *positive* free logics atomic sentences involving non-referential terms are sometimes true. This seems to be a natural requirement, at least to the extent that it is possible to make a case that sentences such as "Pegasus is a horse," "Vulcan is a planet," or "the present king of France = the present King of France," are not false. A third approach, perhaps not as prominent as the first two, is the "neutral" free logic of Smiley [21] and Lehman [17], in which atomic sentences containing at least one term with no existential import are truth-valueless. This paper is only concerned with *positive* free logic.

It is worth noting that both positive and negative free logic have been applied to the theory of computation: see the already cited Gumb and Lambert [8] for applications of positive free logic – including a claim that only the positive variety is adequate for the semantics of "lazy" programming languages – and Feferman [5] for applications of negative free logic.

Semantics for positive free logic come mainly in two main varieties. The first variety is the "outer-domain" approach first proposed in lectures by Lambert, and, independently, in lectures by Belnap in the late 1950's, and later utilized by Cocchiarella [4], Scott [20], and Leblanc and Thomason [11]. The second variety is the supervaluational approach of van Fraassen [6]. According to the first approach, models for free logic come equipped with two domains: an inner domain of "existing" objects, and an outer domain most naturally interpreted as a domain of objects providing references for terms with no existential import. This Meinongian interpretation of outer domains appears to some to be questionable on general philosophical grounds (alternative construals of outer domains that are not open to this sort of criticism are provided by the *nominal* semantics of Meyer and Lambert [19] and the *story* semantics of van Fraassen and Lambert [7]). The supervaluational approach, on the other hand, is not open to this kind of criticism, but at a price. In particular, the price consists in having to accept a framework that (i) is not bivalent; and (ii) is mathematically intractable (as Woodruff [22] has shown, supervaluational validity is neither compact nor recursively axiomatizable).

Considerations such as these have lead Karel Lambert to put forward the following wish-list:

> It would be nice to have a gapless, bivalent semantic development for *positive* free logic in which the model structure is of the single domain variety, but the interpretation function applied to singular terms is *partial*, and doesn't appeal to senses or things of that kind. There is such a semantics for negative free logic, . . . but not one for positive free logic – at least there is not one in which the truthvalues of statements with irreferential singular terms approximate the ones we have clear intuitions about. So there is an open problem for you. (Lambert [15, p. 80])

This paper presents a semantic framework meeting the requirements put forward by Lambert above. In particular, a framework for *positive free logic* is given, in which terms are allowed to have no existential import, and sentences involving such terms are allowed to be true. Moreover, the framework is *bivalent*, which sets it apart from the usual supervaluational approach [6]. There is only one domain of objects, resisting the temptation to admit non-existent objects. The interpretation function (more specifically: the reference function for singular terms) is *partial*. The framework is completely *extensional*: there are no senses, intensions, concepts, or any other “creatures of the dark.” Finally, in this framework, the most fundamental principles of free logic are validated: first and foremost universal instantiation restricted to referential terms, but also Lambert's law (for any x: x is identical with the ϕ if and only if x and x only is ϕ).

Before getting down to precise details, consider the main idea informally. As is well known, there are numerous logical frameworks in the literature in which the notion of truth is *relativized* or, better, *parametrized*, i.e., made to depend upon one or more extra parameters that might not be apparent in the kind of linguistic phenomena for which those frameworks purport to supply an analysis. Perhaps the best known, and most successful, example of this strategy is *modal logic*, in which truth is made to depend upon a parameter – a possible world – that is hidden in the ordinary modal parlance of possibility and necessity.

The present proposal could be viewed as going in the same direction. Here, too, truth is parametrized to an index that does not appear explicitly in parlance involving terms with no existential import. But, in contrast to the modal case, the parameter rather than being extralinguistic is itself part of the language, so that one can speak, in analogy to modal logic, of *truth at* a term t.

For example, consider a language containing an individual constant c and a one-place predicate symbol P. A *proto-interpretation* for such a language is a function π that (among other things) assigns a *signed extension* to P, where a “signed extension” is a subset S (of some given domain D of objects) associated with a special marker $+$ or $-$. The idea is that the marker itself, $+$ or $-$, does not carry any existential commitments; the markers are not to be thought of as independent entities, and could in principle be identified with presence or absence of any given arbitrary property of S. If it were possible to assign, e.g., colors (blue and red) to subsets of D, then a set $S \subseteq D$ could be regarded as associated with a marker $+$ or $-$ according as S is blue or red.

Once a class of signed extensions has been identified (by whatever device should turn out to be convenient), and thereby also a class of proto-interpretations, an *interpretation* for the language is defined as an assignment of a proto-interpretation to each term. Truth then works as follows: where π is the proto-interpretation assigned to c, and S the extension π assigns to P, the atomic sentence $P(c)$ is true on the given interpretation if and only if: either c denotes some object d and $d \in S$; or c does not denote, and S has a positive sign $+$. In the general case where predicates can have more than one argument, proto-interpretations are assigned not only to terms but to n-tuples of terms.

Moreover, the following constraints will be laid down: (i) In order for the resulting semantics to be extensional, it is required that all terms be assigned proto-interpretations differing, if at all, only as to their *sign*, and (ii) in order for substitutivity of identicals to hold in general, whenever t and t' are assigned the same proto-interpretation, so must any two n-tuples of terms differing only because one has occurrences of t where the other has occurrences of t'. With this in mind, on to the details.

2. SEMANTICS

Consider a first-order language with definite descriptions. In particular, assume that the language has the ordinary connectives and quantifiers (though – in order to simplify definitions – $\neg$, $\wedge$ and $\forall$ are taken as basic); individual constants $c_0, c_1, \ldots$; predicates of any number of arguments; and identity (written $\equiv$). The usual recursive definition of well-formed formula is augmented by adding a clause according to which if A is a formula and x a variable, then $\imath x A$ is a term (intuitively, “the unique x such that A”). The basic intuition underlying free logic, of course, is that not all terms need have a referent. Quantifiers, however, are taken to range only over objects in some domain D of objects, so that the fact that a term t has existential import can be expressed by the sentence $\exists x(x \equiv t)$ (this is sometimes abbreviated as $E!$ (“E-shriek” or “E-bang”), as in Figure 1).

To specify a semantical interpretation for the language, fix a (possibly empty) domain D of objects: these are the *existing* objects – the only kind

Figure 1. The Winged Horse.

there is. In addition to this, suppose there is a *partial reference function* ρ, assigning objects in D to (some, all, or none of) the constants of the language. If a term t is in the domain of ρ then we write $\rho(t)\downarrow$, and we write $\rho(t)\uparrow$ otherwise.

A *proto-interpretation* is a function π assigning to each predicate P of the language a *signed extension* $\pi(P)$; a signed extension is a pair either of the form $(S, +)$ or of the form $(S, -)$, where $S \subseteq D^n$ (if P is a n-place predicate), and $+$ and $-$ are markers (these can be taken to be any two arbitrary distinct objects). When we want to leave unspecified the sign of a signed extension we write $\pi(P) = (S, \pm)$. Two proto-interpretations π and π' are said to be *equivalent* if for any P, $\pi(P)$ and $\pi'(P)$ differ at most on the sign of the extension of P.

Finally, an *interpretation* for the language is a function Π assigning to each n-tuple of terms, $\bar{t} = t_1 \ldots t_n$, a proto-interpretation $\Pi(\bar{t}) = \pi$, subject to the following two requirements:

1. If π and π' are both in the range of Π, then π is equivalent to π';
2. If $\Pi(t) = \Pi(t')$ then $\Pi(t_1 \ldots t \ldots t_n) = \Pi(t_1 \ldots t' \ldots t_n)$.

For readability write $\Pi_{\bar{t}}$ instead of $\Pi(\bar{t})$. The pair (Π, ρ) is called a *model*.

The following clauses provide a simultaneous recursive definition of truth on a model ($(\Pi, \rho) \models A$) and the "lifting" of the partial reference function ρ from assigning referents to just constants to arbitrary terms. In order to handle the case for the quantifier, suppose that the model (Π, ρ) is *saturated* in the sense that for any $d \in D$ there is a constant c_d in the language such that $\rho(c_d) = d$ (so that all members of D have names).

First the definition deals with atomic sentences:

$$(\Pi, \rho) \models P(t_1 \dots t_n) \text{ iff } \begin{cases} (\rho(t_1) \dots \rho(t_n)) \in S, \\ \quad \text{if } \rho(t_i) \downarrow \text{ (for } 1 \leqslant i \leqslant n) \\ \quad \text{and } \Pi_{\bar{t}}(P) = (S, \pm); \\ \Pi_{\bar{t}}(P) = (S, +), \\ \quad \text{if } \rho(t_i) \uparrow \text{ for some } i \in \{1, \dots, n\}. \end{cases}$$

The case for the connectives are as usual: $(\Pi, \rho) \models A \wedge B$ iff $(\Pi, \rho) \models A$ and $(\Pi, \rho) \models B$; and similarly $(\Pi, \rho) \models \neg A$ iff $(\Pi, \rho) \not\models A$. As mentioned, the case for the quantifier is handled substitutionally:

$$(\Pi, \rho) \models \forall x A \quad \text{iff for all } d \in D\text{: } (\Pi, \rho) \models A[c_d/x].$$

Turning to identity, recall that the principle of indiscernibility governs identity between terms with no existential import. Accordingly, we have:

$$(\Pi, \rho) \models t_1 \equiv t_2 \text{ iff } \begin{cases} \rho(t_1) = \rho(t_2), & \text{if } \rho(t_1) \downarrow \text{ or } \rho(t_2) \downarrow; \\ \Pi_{t_1} = \Pi_{t_2}, & \text{if both } \rho(t_1) \uparrow \text{ and } \rho(t_2) \uparrow. \end{cases}$$

Finally, the promised extension of ρ to descriptions:

$$\begin{cases} \rho(\imath x A) = d, & \text{if } (\exists! d \in D)(\Pi, \rho) \models A[c_d/x]; \\ \rho(\imath x A) \uparrow, & \text{otherwise.} \end{cases}$$

It is easy to see that according to this definition sentences involving terms with no existential import can be true on some interpretations and false on others, and not all atomic sentences containing such terms need have the same truth value. So for instance there are interpretations on which “Pegasus is a horse” is true, but “Vulcan is a horse” is false. It is worth nothing that all identities of the form “$t \equiv t$” will turn out true, and that whenever $t_1 \equiv t_2$ is true then the same sentences are true of them (i.e., t_1 and t_2 are indiscernible), regardless of whether they have referential import or not. In other words, the indiscernibility principle

$$t_1 \equiv t_2 \rightarrow (A(t_1) \leftrightarrow A(t_2))$$

is logically true, as is Lambert’s Law, i.e.:

$$\forall x(x \equiv \imath y A(y) \leftrightarrow \forall y(A(y) \leftrightarrow y \equiv x)).$$

3. Model Theory

This section presents a few fundamental facts about the semantic framework, primarily compactness and interpolation. These are important properties because they set apart proto-semantics from other frameworks for positive free logic such as the supervaluational approach of [6]. In particular, it was shown by Woodruff [22] that supervaluational semantics is not compact and supervaluational validity is not recursively axiomatizable (it is, in fact, Π^1_1). In turn, interpolation yields the definability theorem, an important tool if free logics are to be applied in computational contexts, where a function can be defined by arbitrary expressions.

3.1. *Compactness*

The compactness theorem for proto-semantics states that if a set Γ of sentences is finitely satisfiable then Γ is satisfiable. The proof is modeled after the usual Henkin construction.

Step 1. Given a finitely satisfiable set Γ of sentences, we expand the language by adding countably many new constants $c_1, c_2, c_3, \ldots$; let φ_i (for $i \geqslant 0$) be an enumeration of the formulas of the expanded language having one free variable. Then extend Γ to a set Γ' by adjoining all sentences θ_n of the form:

$$\exists x \varphi_n \rightarrow (\exists y (y \equiv c_{n_i}) \wedge \varphi_n[c_{n_i}/x]),$$

where x is the free variable of φ_n, and c_{n_i} is the first new constant not occurring in any of the sentences θ_j for $j < n$.

LEMMA 3.1. *Γ' is finitely satisfiable.*

Proof. Let Γ_0 be a finite subset of Γ and consider and arbitrary finite subset of Γ':

$$\Gamma'_0 = \Gamma_0 \cup \{\theta_{i_1}, \ldots, \theta_{i_n}\}.$$

For simplicity assume $n = 1$. By hypothesis, there is a model (Π, ρ) such that $(\Pi, \rho) \vDash \Gamma_0$. We distinguish two cases.

First case: $(\Pi, \rho) \vDash \neg\exists\varphi_{i_j}$ then θ_{i_j} is true on (Π, ρ).

Second case: $(\Pi, \rho) \vDash \exists\varphi_{i_j}$; then for some $d \in D$, $(\Pi, \rho) \vDash \varphi_{i_j}$. Let c be the new constant occurring in θ_{i_j}, and let (Π', ρ') be the model just like (Π, ρ), except that $\rho'(c) = d$ and $\Pi'_c = \Pi_{c_d}$. Then $(\Pi', \rho') \vDash \theta_{i_j}$ and hence satisfies Γ'_0. □

LEMMA 3.2. *Let* (Π, ρ) *be a model. If both* $\rho(t) \uparrow$ *and* $\rho(t') \uparrow$ *and* $\Pi_t = \Pi_{t'}$ *then for every* φ

$$(\Pi, \rho) \models \varphi(t) \Leftrightarrow (\Pi, \rho) \models \varphi(t').$$

Proof. By induction of φ. □

Step 2. Now we extend Γ' to a maximal finitely satisfiable set Θ in the usual way. Let $\{\varphi_n : n \geqslant 0\}$ be an enumeration of all the sentences of the language. Put $\Theta_0 = \Gamma'$ and, assuming Θ_n already defined, put $\Theta_{n+1} = \Theta_n \cup \{\varphi_n\}$ if the latter is still finitely satisfiable, and let $\Theta_{n+1} = \Theta_n \cup \{\neg\varphi_n\}$, otherwise. Finally, put

$$\Theta = \bigcup_{n \geqslant 0} \Theta_n.$$

As in the case for first-order logic, the following are easily established:

1. for each n, Θ_n is finitely satisfiable (the connectives are classical);
2. for each φ, either $\varphi \in \Theta$ or $\neg\varphi \in \Theta$;
3. if $\exists x\varphi \in \Theta$, then for some term t, $\varphi(t) \in \Theta$.

Now we proceed to define a model (Π, ρ) for Θ. First let

$$U = \{t : \exists x(x \equiv t) \in \Theta\},$$

and as usual define an equivalence relation $\simeq$ on U by letting $t \simeq t'$ if and only if $t \equiv t' \in \Theta$. If we denote by $[t]$ the equivalence class of t, we obtain the domain of the interpretation:

$$D = \{[t] : t \in U\}.$$

Next, define the reference function ρ as follows:

$$\rho(t) = \begin{cases} [t] & \text{if } t \in U; \\ \uparrow & \text{if } t \notin U. \end{cases}$$

The next item is the assignment of extensions. For each n-place predicate P, let

$$\text{ext}(P) = \{[t_1] \ldots [t_n] : Pt_1 \ldots t_n \in \Theta\}.$$

Having done this, $\Pi_{\bar{t}}$ can be directly defined: for each n-place predicate

$$\Pi_{\bar{t}}(P) = (\text{ext}(P), +) \Leftrightarrow P\bar{t} \in \Theta,$$

and

$$\Pi_{\bar{t}}(P) = (\text{ext}(P), -),$$

otherwise. This completely specifies a model (Π, ρ). There is no need explicitly to give the clauses for $(\Pi, \rho) \vDash t \equiv t'$, for identity is part of the logical vocabulary. In particular, it will be the case that if either $\rho(t)\downarrow$ or $\rho(t')\downarrow$ then $(\Pi, \rho) \vDash t \equiv t'$ if and only if $\rho(t) = \rho(t')$; and otherwise $\Pi, \rho \vDash t \equiv t'$ if and only if $\Pi_t = \Pi_{t'}$.

LEMMA 3.3. *In the model* (Π, ρ), $\rho(\imath x A)\downarrow$ *if and only if* $\rho(\imath x A)$ *is the unique* $d \in D$ *such that* $(\Pi, \rho) \vDash A(c_d)$.

Proof. For the non-trivial direction: assume $\rho(\imath x A)\downarrow$. Then $\exists x(x \equiv \imath x A) \in \Theta$. It follows that $\exists x \forall y(A(y) \leftrightarrow x \equiv y) \in \Theta$, because otherwise

$$\{\exists x \forall y(A(y) \leftrightarrow x \equiv y), \exists x(x \equiv \imath x A)\}$$

would be an unsatisfiable finite subset of Θ. Hence, if $A(t) \in \Theta$, then $t \equiv \imath x A \in \Theta$ and $[t] = [\imath x A]$. □

LEMMA 3.4. *Suppose that both* $\rho(t_1)\uparrow$ *and* $\rho(t_2)\uparrow$. *Then* $(\Pi, \rho) \vDash t_1 \equiv t_2$ *if and only if for every formula* $\varphi(x)$,

$$(\Pi, \rho) \vDash \varphi(t_1) \leftrightarrow \varphi(t_2).$$

Proof. By induction on φ. □

From the previous lemma we immediately obtain that if t_1 and t_2 are non-denoting terms (in the model (Π, ρ) obtained above), then $(\Pi, \rho) \vDash t_1 \equiv t_2$ if and only if $\Pi_{t_1} = \Pi_{t_2}$.

Having come so far, it is the easy to prove that the usual properties of maximal finitely satisfiable sets hold, and hence to show that the obtained model (Π, ρ) verifies all sentences in Θ and hence also all sentences in Γ. This concludes the proof of the compactness theorem.

3.2. *Interpolation*

Consider now Craig's Interpolation theorem. The reader might consult the proof of Chang and Keisler [3, pp. 87–90], to see exactly what added complications are necessary for the proof to go through in the framework of free logic. For simplicity, assume a language $\mathcal{L}$ with no function symbols. Observe that any argument for classical first-order logic that only requires appeal to either universal introduction (generalization on constants) or existential elimination can be reproduced *verbatim* in the present framework. On the other hand, one needs to develop workarounds for any arguments depending on universal elimination or existential introduction. In what follows, a *theory* is just a set of sentences.

DEFINITION 3.5. Consider languages $\mathcal{L}_1$ and $\mathcal{L}_2$, and let $\mathcal{L}_0 = \mathcal{L}_1 \cap \mathcal{L}_2$. If C is a set of new constants not already in $\mathcal{L}_1 \cup \mathcal{L}_2$, let $\mathcal{L}'_i = \mathcal{L}_i \cup C$ (for $i = 0, 1, 2$).

Let S and T be theories in $\mathcal{L}_1$ and $\mathcal{L}_2$ respectively; let φ, ψ be sentences in $\mathcal{L}_1$ and $\mathcal{L}_2$ respectively, and θ a sentence in $\mathcal{L}'_0$. Then

(1) θ *separates* theories S and T if and only if $S \vDash \theta$ and $T \vDash \neg\theta$.
(2) θ *separates* φ and ψ if and only if its separates $\{\varphi\}$ and $\{\psi\}$.
(3) S and T are *inseparable* if and only if there is no sentence θ that separates them (similarly for φ and ψ being inseparable).
(4) θ is a *Craig interpolant* of φ and ψ if and only if $\varphi \vDash \theta$ and $\theta \vDash \psi$, and moreover every relation and constant symbols (other than identity) occurring in θ, also occurs in both φ and ψ.

The following definition provides an infinite supply of syntactically distinct non-denoting terms.

DEFINITION 3.6. For each $n > 0$, let ι_n be the term

$$\imath x\, \underbrace{(x \neq x) \wedge \cdots \wedge (x \neq x)}_{n \text{ times}}.$$

As we noted, the terms ι_n are syntactically distinct; moreover, they are non-denoting in every model. It follows that, although the matrices of the terms ι_n are logically equivalent to one another, they can be assigned arbitrary proto-interpretations, a fact that will be made use of in the proof of Craig interpolation.

LEMMA 3.7. *Let $\theta(x_1, \ldots, x_n)$ be an $\mathcal{L}$-formula, with the free variables shown, and let φ, ψ be $\mathcal{L}$-sentences such that all non-logical symbols of θ already occur in both φ and ψ.*

If $c_1, \ldots, c_n$ are new individual constants not occurring in any of φ, ψ or θ, and $\theta(c_1, \ldots, c_n)$ separates φ and ψ, then there is a Craig interpolant of φ and $\neg\psi$.

Proof. In classical first-order logic, it suffices to observe that the universal closure of θ is the desired interpolant. Such an easy solution is not available in free logic, as some of the c_i's might be non-denoting.

Given θ, φ and ψ, pick n terms of the form ι_i not already occurring in θ, φ, or ψ. With no loss in generality, assume these are $\iota_1, \ldots, \iota_n$.

Introduce the following notations: let $I = \{\iota_1, \ldots, \iota_n\}$, and let 2^I be the power set of I. Given $J \in 2^I$, let $\theta[J]$ denote the result of simultaneously substituting each $\iota_i \in J$ for x_i in θ, and let $\forall\theta[J]$ be the universal closure of $\theta[J]$. Put

$$\theta^* = \bigvee_{J \in 2^I} \forall\theta[J].$$

The sentence θ^* is an interpolant of φ and $\neg\psi$. To see this, it suffices to establish the two claims $\varphi \vDash \theta^*$ and $\theta^* \vDash \psi$, as the conditions on non-logical symbols is clearly met.

Part (I): ad $\varphi \vDash \theta^*$. Let $(\Pi, \rho) \vDash \varphi$; suffices to show that (Π, ρ) is a model of some disjunct in θ^*. In fact, (Π, ρ) is a model of *every* disjunct of θ^*.

Let $(\Pi, \rho) \vDash \varphi$ be a $\mathcal{L}$-model, and $\forall\theta[J]$ any disjunct in θ^*, to show $(\Pi, \rho) \vDash \forall\theta[J]$.

Note that since $c_1, \ldots, c_n$ do not occur in φ, any expansion (Π', ρ') of (Π, ρ) to the language $\mathcal{L}' = \mathcal{L} \cup \{c_1, \ldots, c_n\}$ still is a model of φ; and since $\varphi \vDash \theta(c_1, \ldots, c_n)$, any such expansion is a model of $\theta(c_1, \ldots, c_n)$.

Now let $x_{i_1}, \ldots, x_{i_m}$ be the free variables of $\theta[J]$, and pick arbitrary elements $a_1, \ldots, a_m$ of the domain of Π. Let $(\Pi'\rho')$ be the expansion of (Π, ρ) such that:

(1) $\rho'(c_{i_k}) = a_k$, for $k \in \{1, \ldots, m\}$ (with $\Pi'(c_{i_k})$ defined in any arbitrary manner);
(2) $\rho'(c_j) \uparrow$ and $\Pi(c_j) = \Pi(\iota_j)$, for $c_j \notin \{c_{i_1}, \ldots, c_{i_m}\}$.

Then $(\Pi'\rho') \vDash \theta(c_1, \ldots, c_n)$, whence also $(\Pi', \rho') \vDash \theta[J][c_{i_1}/x_{i_1}, \ldots, c_{i_m}/x_{i_m}]$, and by the arbitrariness of $a_1, \ldots, a_m$, also $(\Pi', \rho') \vDash \forall\theta[J]$. But now the constants $c_{i_1}, \ldots, c_{i_m}$ no longer occur in $\forall\theta[J]$, so $(\Pi, \rho) \vDash \forall\theta[J]$ as desired.

Part (II): ad $\theta^* \vDash \psi$. By hypothesis, $\neg\psi \vDash \neg\theta(c_1, \ldots, c_n)$. Let $(\Pi, \rho) \vDash \theta^*$: then there is a disjunct $\forall\theta[J]$ of θ^* such that $(\Pi, \rho) \vDash \forall\theta[J]$. Assume $J = \{\iota_{i_1}, \ldots, \iota_{i_m}\}$, and expand (Π, ρ) to a model (Π', ρ') satisfying:

(1) $\rho'(c_k) \uparrow$, and $\Pi'(c_{i_k}) = \Pi(\iota_{i_k})$ for $k \in \{1, \ldots, m\}$;
(2) for $c_j \notin \{c_{i_1}, \ldots, c_{i_m}\}$, pick an arbitrary a_j in the domain of Π and put $\rho'(c_j) = a_j$ (with $\Pi(c_j)$ arbitrarily defined).

Then $(\Pi'\rho') \vDash \theta(c_1, \ldots, c_n)$, whence $(\Pi'\rho') \vDash \psi$. Since ψ does not contain any of $c_1, \ldots, c_n$, also $(\Pi, \rho) \vDash \psi$, as desired. □

THEOREM 3.8. *Let φ, ψ be $\mathcal{L}$-sentences such that $\varphi \vDash \psi$. Then there is a Craig interpolant θ of φ and ψ.*

Proof. We assume that there is no interpolant θ for φ and ψ and show that $\varphi \nvDash \psi$ by exhibiting a model of φ falsifying ψ. By Lemma 3.7, φ and $\neg\psi$ are inseparable. Let

$$\varphi_0, \varphi_1, \varphi_2, \ldots; \psi_0, \psi_1, \psi_2, \ldots$$

be enumerations of $\mathcal{L}'_1$ and $\mathcal{L}'_2$ (where these are defined as in 3.5). We construct two increasing sequences of theories $T_0 \subset T_1 \subset \cdots$ and $U_0 \subset U_1 \subset \cdots$; proceeding by induction, we put $T_0 = \{\varphi\}$ and $U_0 = \{\neg\psi\}$.

For the inductive step, we obtain T_{m+1} and U_{m+1} from T_m and U_m by the following procedure:

1. if $T_m \cup \{\varphi_m\}$ and U_m are inseparable, then put φ_m in T_{m+1};
2. if $U_m \cup \{\psi_m\}$ and the result of step (1) are inseparable, then put ψ_m in U_{m+1};
3. if φ_m is an existential sentence, say $\exists x\xi(x)$, and φ_m has been listed in T_{m+1} in step 1, then pick a new constant c not already in T_m, U_m, φ_m or ψ_m, and put $\xi(c)$ in T_{m+1};
4. if ψ_m is an existential sentence, say $\exists x\xi(x)$, and ψ_m has been listed in U_{m+1} in step 2, then pick a new constant c not already in T_m, U_m, φ_m or ψ_m, and put $\xi(c)$ in U_{m+1}.

This completes the definition of the two sequences. For each m, T_m and U_m are inseparable. For T_0 and U_0 are inseparable, and the inductive step preserves inseparability. Clearly inseparability is preserved by steps (1) and (2) in the definition of T_{m+1} and U_{m+1}. And the classical argument that adjoining $\xi(c)$ when $\exists x\xi(x)$ is already present only depends on existential elimination, which, as observed, is valid in free logic.

Now let $T_\infty = U_{n\geqslant 0}T_n$ and similarly $U_\infty = U_{n\geqslant 0}U_n$. By compactness, T_∞ and U_∞ are inseparable, and, in particular, both consistent (for the purposes of this proof, take 'consistent' to mean 'satisfiable'). Of course, the theorem requires the stronger claim that $T_\infty \cup U_\infty$ is consistent.

Next, observe that T_∞ is a maximal consistent theory in $\mathcal{L}'_1$ and U_∞ is a maximal consistent theory in $\mathcal{L}'_2$. Suppose for instance that $\varphi_m, \neg\varphi_m \notin T_\infty$. Then if $\varphi_m \notin T_\infty$, also $\varphi_m \notin T_{m+1}$, which implies that $T_m \cup \{\varphi_m\}$ and U_m are separable, and therefore that there is a sentence θ of $\mathcal{L}'_0$ such that

$$T_\infty \vDash \varphi_m \to \theta \quad \text{and} \quad U_\infty \vDash \neg\theta.$$

Similarly, for n such that $\varphi_n = \neg\varphi_m$, if $\neg\varphi_m \notin T_{n+1}$ then there is a sentence θ' of $\mathcal{L}'_0$ such that

$$T_\infty \vDash \neg\varphi_m \to \theta' \quad \text{and} \quad U_\infty \vDash \neg\theta'.$$

By propositional logic,

$$T_\infty \vDash \theta \vee \theta' \quad \text{and} \quad U_\infty \vDash \neg(\theta \vee \theta'),$$

against the fact that T_∞ and U_∞ are inseparable. This shows that T_∞ is a maximal consistent set; the maximality of U_∞ is similar.

The last fact needed before we can construct a model is that $T_\infty \cap U_\infty$ is maximal consistent in $\mathcal{L}_0'$. Let θ be any $\mathcal{L}_0'$-sentence. Since we know that $T_\infty \cap U_\infty$ is consistent, we need to show that either $\theta \in T_\infty \cap U_\infty$ or $\neg\theta \in T_\infty \cap U_\infty$. Since each of T_∞ and U_∞ is maximally consistent in its own language, either $\theta \in T_\infty$ or $\neg\theta \in T_\infty$; and similarly either $\theta \in U_\infty$ or $\neg\theta \in U_\infty$. Of the four resulting cases, by inseparability, we cannot have $\theta \in T_\infty$ and $\neg\theta \in U_\infty$ or vice-versa. So there are two cases left: either θ is in both T_∞ and U_∞, or $\neg\theta$ is, as required.

Finally, we construct a model for $T_\infty \cup U_\infty$ (and hence for $\{\varphi, \neg\psi\}$). This is where the argument again deviates somewhat from the classical proof. Since T_∞ is consistent, let (Π_1, ρ_1) be a model of T_∞ with domain D_1. Clearly D_1 contains denotations $\rho_1(c)$ for all constants c such that $T_\infty \vDash \exists x(x \equiv c)$. Now consider the model (Π_1', ρ_1') having domain

$$D_1' = \{\rho_1(c) : T_\infty \vDash \exists x(x \equiv c)\},$$

and where (Π_1', ρ_1') are otherwise unchanged. This new model is elementarily equivalent to (Π, ρ) (by induction on sentences, using the fact that T_∞ has witnesses to all the existential claims), and therefore still a model of T_∞. Similarly, U_∞ has a model (Π_2', ρ_2') having domain D_2' comprising all and only the denotations of the denoting constants.

Now consider (Π_1'', ρ_1'') and (Π_2'', ρ_2''), the reducts of (Π_1', ρ_1') and (Π_2', ρ_2'), to the language $\mathcal{L}_0'$ (with the same domains D_1' and D_2'). The mapping h: $D_1' \to D_2'$ such that $h(\rho_1''(c)) = \rho_2''(c)$ is a "classical" isomorphism between the two models, in the sense that h is a bijection of the respective domains with the property that for each n-tuple $\rho_1''(c_1), \ldots, \rho_1''(c_n)$ and predicate P, if $(\Pi_1''(c_1 \ldots c_n))(P) = (S_1, \pm)$, and $(\Pi_2''(c_1 \ldots c_n))(P) = (S_2, \pm)$ then

$$(\rho_1''(c_1), \ldots, \rho_1''(c_n)) \in S_1 \Leftrightarrow (h(\rho_1''(c_1)), \ldots, h(\rho_1''(c_n))) \in S_2.$$

That h is such a "classical" isomorphism follows from the fact that $T_\infty \cap U_\infty$ is maximal consistent in $\mathcal{L}_0'$. By the same token, we see that the same terms must be denoting in (Π_1'', ρ_1'') and (Π_2'', ρ_2''), and moreover that if $\bar{t}$ is an n-tuple containing a non-denoting term (in either model), then $(\Pi_1''(\bar{t}))(P)$ has a $+$ sign if and only if $(\Pi_2''(\bar{t}))(P)$ has $+$ sign.

Thus, for each n-tuple $\bar{t}$ of terms, $\Pi_1''(\bar{t})$ can differ from $\Pi_2''(\bar{t})$ in the sign of the extension they assign to some predicate P only if $\bar{t}$ contains only denoting terms. But such conflicts are inconsequential, meaning that they make no difference as to the truth values of sentences, and therefore can be settled in some arbitrary manner (in favor of Π_1'', say), preserving the truth values assigned to sentences. Therefore, we can "identify" (Π_1'', ρ_1'') and (Π_2'', ρ_2''), resolving inconsequential conflicts as indicated.

More precisely, this shows that it is possible to expand (Π_1'', ρ_1'') to a $\mathcal{L}_1' \cup \mathcal{L}_2'$-model of $T_\infty \cup U_\infty$. □

3.3. *Beth Definability*

As is well-known, in classical first-order logic, Beth's definability theorem follows easily from Craig's Interpolation theorem. In this subsection, we offer the analogous proof for positive free logic with proto-semantics (again, the reader is invited to consult [3] for comparison). First, a few definitions.

DEFINITION 3.9. Fix a language $\mathcal{L}$ and consider two n-place predicate symbols P and P' not in $\mathcal{L}$. Let $\Sigma(P)$ be a set of sentences in $\mathcal{L} \cup \{P\}$, and $\Sigma(P')$ the set of sentences of $\mathcal{L} \cup \{P'\}$ obtained by writing P' for P throughout.

(1) $\Sigma(P)$ *weakly implicitly defines* P if and only if

$$\Sigma(P) \cup \Sigma(P') \vDash \forall \bar{x}(P(\bar{x}) \leftrightarrow P'(\bar{x}));$$

(2) $\Sigma(P)$ *weakly explicitly defines* P if and only if there is formula $\varphi(\bar{x})$ of $\mathcal{L}$ such that

$$\Sigma(P) \vDash \forall \bar{x}(P(\bar{x}) \leftrightarrow \varphi(\bar{x}));$$

(3) $\Sigma(P)$ *implicitly defines* P if and only if

$$\Sigma(P) \cup \Sigma(P') \vDash P(\bar{c}) \leftrightarrow P'(\bar{c});$$

where $\bar{c}$ is an n-tuple of new constants not in $\mathcal{L}$;

(4) $\Sigma(P)$ *explicitly defines* P if and only if for some formula $\varphi(\bar{x}) \in \mathcal{L}$

$$\Sigma(P) \vDash P(\bar{c}) \leftrightarrow \varphi(\bar{c}),$$

where $\bar{c}$ is an n-tuple of new constants not in $\mathcal{L}$.

It is worth noting that in classical first order logic implicit and explicit definitions are equivalent to their weak counterparts, but these notions come apart in a free framework.

THEOREM 3.10 (Beth definability). *Fix a language $\mathcal{L}$ and two n-place predicates P and P' not in $\mathcal{L}$. As before, $\Sigma(P)$ is a set of sentences from $\mathcal{L} \cup \{P\}$, and $\Sigma(P')$ the result of replacing P by P' throughout $\Sigma(P)$.*

(1) *If $\Sigma(P)$ weakly implicitly defines P, then $\Sigma(P)$ weakly explicitly defines P.*
(2) *If $\Sigma(P)$ implicitly defines P, then $\Sigma(P)$ explicitly defines P.*

Proof. First we show that part (1) follows from part (2). With no loss in generality we do the case where P is a 1-place predicate symbol. So assume that $\Sigma(P)$ weakly implicitly defines P:

$$\Sigma(P) \cup \Sigma(P') \vDash \forall x(P(x) \leftrightarrow P'(x)),$$

and abbreviate $\exists x(x \equiv c)$ by $E!(c)$. Then, where c is a new constant,

$$\Sigma(P) \cup \Sigma(P') \cup \{E!(c)\} \vDash P(c) \leftrightarrow P'(c).$$

This implies that $\Sigma(P) \cup \{E!(c)\}$ implicitly defines P, whence by part (2) there is a formula $\varphi(x) \in \mathcal{L}$ such that

$$\Sigma(P) \cup \{E!(c)\} \vDash P(c) \leftrightarrow \varphi(c),$$

whence $\Sigma(P) \vDash \forall x(P(x) \leftrightarrow \varphi(x))$, as required. Now we move on to part (2). Assume that

$$\Sigma(P) \cup \Sigma(P') \vDash P(c) \leftrightarrow P'(c).$$

By compactness, there are finite subsets $\Sigma_0'(P)$ and $\Sigma_0''(P)$ of $\Sigma(P)$ such that

$$\Sigma_0'(P) \cup \Sigma_0''(P') \vDash P(c) \rightarrow P'(c).$$

By putting $\Sigma_0(P) = \Sigma_0'(P) \cup \Sigma_0''(P)$ we have

$$\Sigma_0(P) \cup \Sigma_0(P') \vDash P(c) \rightarrow P'(c).$$

Rearranging:

$$\Sigma_0(P) \cup \{P(c)\} \vDash \bigwedge \Sigma_0(P') \rightarrow P'(c),$$

and by interpolation there is a formula $\varphi(x) \in \mathcal{L}$ such that:

$$\Sigma_0(P) \cup \{P(c)\} \vDash \varphi(c) \quad \text{and} \quad \varphi(c) \vDash \bigwedge \Sigma_0(P') \rightarrow P'(c).$$

In other words:

$$\Sigma_0(P) \vDash P(c) \rightarrow \varphi(c) \quad \text{and} \quad \Sigma_0(P') \vDash \varphi(c) \rightarrow P'(c).$$

From the second of these, since $\varphi \in \mathcal{L}$ and therefore does not contain either P or P', we have

$$\Sigma_0(P) \vDash \varphi(c) \rightarrow P(c),$$

which, together with $\Sigma_0(P) \vDash P(c) \rightarrow \varphi(c)$ gives

$$\Sigma_0(P) \vDash \varphi(c) \leftrightarrow P(c).$$

Since $\Sigma_0(P) \subseteq \Sigma(P)$, the desired conclusion follows. □

4. Completeness

In this section we introduce a particular system of axioms for positive free logic (due to Lambert [13]), and we show that proto-semantics is (sound and) complete for such an axiom system. In the presence of compactness, one way to establish completeness is to introduce a refutation procedure, such as truth trees, and establish that all validities are provable by showing how to obtain a counter-model from a failing or non-terminating refutation attempt. Such refutation procedures for free logic are available in the literature, and no doubt they could be used in the present proto-semantical framework. Nonetheless, the argument of this section relies on a Hilbert-style axiomatization and a direct argument for strong completeness.

4.1. *The Axioms*

The set of axioms for Positive Free Logic (PFL) is characterized as follows:

1. All tautologies are axioms;
2. All formulas of the form $\forall x(\varphi \rightarrow \psi) \rightarrow (\forall x\varphi \rightarrow \forall x\psi)$ are axioms;
3. All formulas of the form $\forall y(\forall x\varphi \rightarrow \phi(y))$ are axioms;
4. All formulas of the form $\forall x\forall y\varphi \rightarrow \forall y\forall x\varphi$ are axioms;
5. All formulas of the form $t \equiv t'$ are axioms;
6. All formulas of the form $t \equiv t' \rightarrow (\varphi(t) \rightarrow \varphi(t'))$ are axioms (where $\varphi(t')$ is the result of writing t' for zero or more occurrences of t in $\varphi(t)$).
7. If φ is an axiom and x a variable, then $\forall x\varphi$ is an axiom.
8. Nothing else is an axiom.

The system has only one rule of inference, namely *modus ponens*. If Γ is a set of formulas, we write $\Gamma \vdash \varphi$ to mean that there is a sequence of formulas of PFL, the last one of which is φ, and every formula in the sequence is either an axiom, or is from Γ, or is obtained by modus ponens from previous formulas in the sequence. When $\emptyset \vdash \varphi$, the formula φ is a *theorem* of PFL. Γ is *consistent* if $\Gamma \nvdash \psi \wedge \neg\psi$ or, equivalently, $\Gamma \nvdash \psi$ for some ψ.

It is known that in the presence of identity, axioms of the form $\forall y(\forall x\varphi \rightarrow \phi(y))$ are equivalent to the more standard, although less elegant,

$$\forall x\varphi(x) \wedge \exists x(x \equiv t) \rightarrow \varphi(t).$$

Also (a fact that will be needed below), existential elimination is admissible: if $\Delta \cup \{E!(c), \varphi(c)\} \vdash \psi$ (c a new constant) then already $\Delta \cup \{\exists x\varphi\} \vdash \psi$ (see [2, p. 199]; see also [10]).

One can show, by the usual inductive argument on the length of proofs, that the axiom system presented here is *sound* for proto-semantics, i.e., that if $\Gamma \vdash \varphi$ then $\Gamma \vDash \varphi$.

4.2. *Completeness*

The strategy to prove that $\Gamma \vDash \varphi$ implies $\Gamma \vdash \varphi$ is to suppose that $\Gamma \nvdash \varphi$ (equivalently: $\Gamma \cup \{\neg\varphi\}$ is consistent) and construct a model of Γ falsifying φ. In turn, as usual, it will suffice to establish the following theorem.

THEOREM 4.1. *Any consistent set of formulas is satisfiable.*

Proof. After the model-theoretic *tour-de-force* of the preceding section, we can afford to be somewhat cavalier about the model construction. Let Γ be a consistent set of formulas. As in the proof of compactness, the first step is to introduce infinitely many new constants c_i's and extend Γ to a set Γ' by adjoining all sentences of the form

$$\exists x\varphi \rightarrow (\exists y(y \equiv c_i) \wedge \varphi[c_i/x]),$$

for each formula φ of the expanded language, where c_i is a suitably chosen new constant. Then the resulting Γ' is consistent if Γ is (by existential elimination).

Next, Γ' is extended to a maximally consistent set Θ in the usual way, by adjoining φ or $\neg\varphi$ according as the former or the latter preserves consistency. Finally, a model is extracted from Θ in the same way as in Section 3.1. □

ACKNOWLEDGEMENT

I am indebted to Joe Lambert for introducing me to this topic as well as for several illuminating conversations on free logic, a non-empty subset of which was crucially facilitated by a couple of pints of lager. I also thank an anonymous referee for this journal, whose insightful comments helped improve the paper.

REFERENCES

1. Bencivenga, E.: *Free Logics*, in D. Gabbay and F. Guenthner (eds), *Handbook of Philosophical Logic*, Vol. III, Kluwer Academic Publishers, Dordrecht, 1986, pp. 373–426.
2. Bencivenga, E., Lambert, J. K. and van Fraassen, B. C.: *Logic, Bivalence, and Denotation*, Ridgeview, Atascadero, Calif., 1991.

3. Chang, C. and Keisler, H.: *Model Theory*, 3rd edn, Elsevier North-Holland, Amsterdam and New York, 1990.
4. Cocchiarella, N.: A logic of possible and actual objects, *J. Symbolic Logic* **31** (1996), 388.
5. Feferman, S.: Logic for termination and correctness of functional programs, in Y. N. Moschovakis (ed.), *Logic from Computer Science*, Springer-Verlag, Berlin and New York, 1992, pp. 95–127.
6. van Fraassen, B. C.: Singular terms, truth-value gaps, and free logic, *J. Philosophy* **67** (1966), 481–495.
7. van Fraassen, B. C. and Lambert, J. K.: *Derivation and Counterexample*, Dickeson, 1972.
8. Gumb, R. D. and Lambert, J. K.: Definitions in nonstrict positive free logic, *Modern Logic* **7** (1979), 25–55 (errata and corrected version available from `ftp://ftp.uml.edu/gumb`).
9. Hoare, C. A. R.: An axiomatic basis of computer programming, *Comm. ACM* **12** (1969), 576–583.
10. Leblanc, H.: *Truth-Value Semantics*, North-Holland, Amsterdam 1976.
11. Leblanc, H. and Thomason, R. H.: Completeness theorems for some presupposition-free logic, *Fund. Math.* **62** (1968), 125–164.
12. Lambert, J. K.: The definition of $E!$ in free logic, in *Abstracts: The International Congress for Logic, Methodology and Philosophy of Science*, Stanford University Press, Palo Alto, 1960.
13. Lambert, J. K.: Existential import revisited, *Notre Dame J. Formal Logic* **4** (1963), 133–144.
14. Lambert, J. K.: Free logic and the concept of existence, *Notre Dame J. Formal Logic* **8** (1967), 133–144.
15. Lambert, J. K.: *Free Logics. Their Foundations, Character, and Some Applications Thereof*, Academia Verlag, Sankt Augustin, 1997.
16. Lambert, J. K.: Free logics, in L. Goble (ed.), *The Blackwell Guide to Philosophical Logic*, Blackwell, Oxford and Malde, MA, 2000 (to appear).
17. Lehman, S.: Strict fregean logic, *J. Philos. Logic* **23**(3) (1994), 307–336.
18. Leonard, H. S.: The logic of existence, *Philos. Stud.* **7** (1956), 49–64.
19. Meyer, R. K. and Lambert, J. K.: Universally free logic and standard quantification theory, *J. Symbolic Logic* **33** (1968), 8–26.
20. Scott, D. S.: Existence and description in formal logic, in A. Schoenman (ed.), *Russell, Philosopher of the Century*, Allen & Unwin, London, 1967, 181–200.
21. Smiley, T.: Sense without denotation, *Analysis* **20** (1960), 125–135.
22. Woodruff, P.: On supervaluations in free logic, *J. Symbolic Logic* **49** (1984), 943–950.

Logic and Philosophy of Science,
University of California,
Irvine, CA 92697-5100
(e-mail: aldo@uci.edu)