

Beneš's Partial Model of NF: An Old Result Revisited

Edoardo Rivello

Abstract A paper by Beneš, published in 1954, was an attempt to prove the consistency of NF (Quine's set theory "New Foundations") via a partial model of Hailperin's finite axiomatization of NF. Here, I offer an analysis of Beneš's proof in a De Giorgi-style setting for set theory. This approach leads to an abstract version of Beneš's theorem that emphasizes the *monotone* and *invariant* content of the axioms proved to be consistent, in a sense of *monotony* and *invariance* that this paper intends to state rigorously and to help clarify. Moreover, some tentative speculation will be made about possible developments of the topic in the following two directions: (1) which set theories can be proved to be consistent via Beneš-like constructions, and (2) how can we elaborate on Beneš's model to get a consistency proof for full NF?

1 Introduction

Quine [6] presented an axiomatization of a set theory (now known as "New Foundations," NF, from the title of the paper) based on the notion of *stratification* for formulas in the first-order language of membership (see Section 2 below for a quick overview of Quine's NF). The consistency of NF (relative to some system of set theory, like ZFC or its extensions) is still an open problem. Hailperin [4] showed that NF can be finitely axiomatized by extensionality and by nine (stratified) instances of the axiom schema of comprehension. Ten years later, Beneš [1] gave a proof of the consistency of the subsystem formed by extensionality and Hailperin's first eight axioms.

Most of Beneš's paper is devoted to developing formally the inductive machinery that is needed in the proof, strictly following the notation in Gödel's 1940 monograph on the continuum hypothesis. This choice is likely to have been motivated by the purpose of lending support to Beneš's claim that "NF certainly suffices for the . . . construction" [1, p. 199]. Gödel's methods of closing a set under a finite family of operations and of giving an explicit enumeration of the resulting closure are, at least

Received April 26, 2012; accepted November 24, 2012

2010 Mathematics Subject Classification: Primary 03E70; Secondary 03C30, 03C62

Keywords: new foundations, set theory, comprehension

© 2014 by University of Notre Dame 10.1215/00294527-2688087

nowadays, standard ZF procedures. In this paper I will focus on the remaining part of Beneš's proof, where he claims that his construction is, as expected, a model of the selected instances of comprehension. Since for the first eight Hailperin axioms this fact is almost evident, checking that all is correct is left to the reader.

The aim of this conceptual analysis of Beneš's paper is to extract some general principles from his proof by explaining why the construction works in this particular case. The result is an abstract (and also generalized) version of Beneš's theorem that isolates a notion of *monotone* and *invariant* formulas. This notion will hopefully prove useful in work on the axiom of comprehension.

A central role in this analysis of Beneš's proof is played by the *family of extensions* that is naturally associated to any model of a set theory and from which every model of a set theory can be recovered. Such a translation between membership relations and families of extensions will be called the *De Giorgi translation*, following a similar terminology that can sometimes be found in the literature (see, for instance, the beginning of Forster [3, Chapter 8.1]). Section 3 will be devoted to introducing a suitable notation for this unorthodox way of presenting set theory.

Sections 4 and 5 give, respectively, the generalized version of Beneš's theorem and an account of the original proof in this more abstract setting. Finally, the last two sections host some speculation about possible applications of the present approach to the NF consistency problem and to the general study of the axiom of comprehension, by taking, in both cases, Beneš's result as a paradigmatic leading example.

2 Quine's NF

Quine's paper [6] starts with a formulation of Frege's system in the language of membership, that is, with the first-order theory axiomatized in the language with just one binary relational symbol \in (as the only nonlogical one) and consisting of the following axioms:

- (1) principle of extensionality (Ext):

$$\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y;$$

- (2) principle of abstraction:

$$\exists y \forall x (x \in y \leftrightarrow \varphi),$$

where φ is a formula of the language of membership and y does not occur free in φ .

In fact, the principle of abstraction has the form of an axiom *schema*, which we will refer to as the axiom schema of *comprehension*.

In 1902, Russell proved the inconsistency of Frege's system by exhibiting the so-called *Russell paradox*, that is, the antinomy that arises by taking the formula $x \notin x$ as φ in (2).

This difficulty led, in 1908, to two different proposals for amending Frege's system: Russell's theory of types (later developed in Russell and Whitehead's *Principia Mathematica*) and Zermelo's set theory (the source of the current "standard set theory" known as Zermelo–Fraenkel set theory). Quine's own proposal maintains Zermelo's language (LST, the acronym for "language of set theory," that is, the language of membership) and attempts to avoid inconsistencies by restricting the axiom schema of comprehension to the "stratified" formulas of the language, that is, to all

those formulas which can be translated into a corresponding formula of the language of the theory of types by assigning suitable types to their variables.

More precisely, a formula φ of LST is called a *stratified* formula if and only if there exists a map s from the set of the variables occurring in φ to the set of the natural numbers such that

- $s(x) = s(y)$ whenever $x = y$ is an atomic subformula of φ , and
- $s(x) = s(y) + 1$ whenever $y \in x$ is an atomic subformula of φ .

The purpose of Quine's NF, as stated in the original paper, was to provide a foundational theory that would avoid inconsistencies, as well as the "unnatural and inconvenient consequences" [6, p. 78] (in Quine's words) of Russell's theory of types. Indeed, in NF we are allowed to speak about objects such as the universal set, the complement operation, the natural numbers, and so on, in the same way as we do in the theory of types. However, in NF these objects really are *the* universal set \mathbf{V} , *the* natural numbers, and so on, without any reference to the awkward infinite copies (one for each type) of these objects whose existence, by contrast, is imposed by the theory of types. Moreover, as noted by Quine, "the deductive power of [NF] outruns that of the Principia" [5, p. 80], since within NF it is also possible to prove some unstratified instances of comprehension.

On the other hand, NF seems to avoid the known paradoxes of set theory (such as Russell's, Burali-Forti's, etc.). In particular, within NF one can recast Cantor's paradox (or even Russell's paradox; see Rosser [7]) as a proof of the nonexistence of the map $\iota : x \mapsto \{x\}$.

3 De Giorgi's Translation

Let $\mathcal{M} = \langle M, E \rangle$ be any structure in the language of set theory (LST). For each element $x \in M$, define $x_E = \{y \in M \mid y E x\}$ to be the *extension* of x in \mathcal{M} . Consider the map $h_E : x \mapsto x_E$. It is immediate that $\langle M, E \rangle$ is a model of extensionality if and only if h_E is injective.

Conversely, given any function $h : M \rightarrow \mathcal{P}(M)$, where $\mathcal{P}(M)$ denotes the power set of M , we can define a binary relation on M as

$$y E_h x \leftrightarrow y \in h(x).$$

Again, $\langle M, E_h \rangle$ is a model of extensionality if and only if h is injective.

Fact 3.1 The maps $E \mapsto h_E$ and $h \mapsto E_h$ are one-to-one and inverse to each other.

We will call any injective function from a set M to its powerset $\mathcal{P}(M)$ a *De Giorgi map*.

In what follows the range of a De Giorgi map will play an essential role, so we need the following definition.

Definition 3.2 We say that a subset $\mathcal{E} \subseteq \mathcal{P}(M)$ is a *De Giorgi family* if and only if \mathcal{E} has the same cardinality as M .

For ease of notation, we shall only consider LST-structures of the form $\langle \omega, E \rangle$, where ω denotes the set of natural numbers. Accordingly, in this paper a De Giorgi family will simply be a *denumerable subset* of $\mathcal{P}(\omega)$, and a *membership relation* E will be any extensional binary relation on ω .

Given a membership relation E , we denote by $E(\omega) = \{x_E \mid x \in \omega\}$ the De Giorgi family defined by E , that is, the set of all extensions (in the sense of E) of elements of ω . Given a De Giorgi family \mathcal{E} and a membership relation E , we say that E is an \mathcal{E} -membership relation if and only if $E(\omega) = \mathcal{E}$ or, in other words, if and only if there exists a (unique) injective enumeration h of \mathcal{E} such that $E = E_h$.

Fact 3.1 allows us to translate results about extensional LST-structures $\langle \omega, E \rangle$ in terms of injective functions from ω into $\mathcal{P}(\omega)$. In particular, we can recast comprehension as follows.

Let \vec{z} denote a generic finite sequence of variables z_1, \dots, z_n occurring free in a formula φ , so that we write $\forall \vec{z}$ for $\forall z_1, \dots, \forall z_n$.

Given a formula $\varphi(x, \vec{z})$ of LST, denote by σ_φ the corresponding instance of the comprehension axiom schema

$$\forall \vec{z} \exists y \forall x (x \in y \leftrightarrow \varphi(x, \vec{z})).$$

Given a membership relation E on ω , denote by $F_{\varphi, E}$ the map

$$F_{\varphi, E}(\vec{z}) = \{x \mid \langle \omega, E \rangle \models \varphi[x, \vec{z}]\}.$$

Let \vec{Z} denote a finite sequence Z_1, \dots, Z_n of subsets of ω . Then, given any function f , we denote the finite sequence $f(Z_1), \dots, f(Z_n)$ by $f(\vec{Z})$.

Define, for every $\vec{Z} \in \mathcal{E} = E(\omega)$,

$$\mathcal{F}_{\varphi, E}(\vec{Z}) = F_{\varphi, E}(h_E^{-1}(\vec{Z})).$$

Fact 3.3 $\langle \omega, E \rangle$ is a model of σ_φ if and only if $\mathcal{E} = E(\omega)$ is closed under the function $\mathcal{F}_{\varphi, E} : \mathcal{E} \rightarrow \mathcal{P}(\omega)$.

So, in De Giorgi's terms, the instances of comprehension that are true in a given LST-structure $\langle \omega, E \rangle$ correspond to closure properties of the family of extensions $\mathcal{E} = E(\omega)$.

The procedure followed by Beneš in constructing his partial model of NF—as outlined in the second paragraph of [1]—can roughly be summarized in De Giorgi's terms as follows.

First, we observe that the axioms P1–P8 of Hailperin's finite axiomatization of NF (*Beneš axioms*) are instances σ_φ of comprehension that correspond (in a sense that will be made precise) to operations \mathcal{F}_1 – \mathcal{F}_8 in $\mathcal{P}(\omega)$ (*Beneš operations*). Then we close a suitable family of subsets of ω under these operations. A standard procedure for closing a set under a family of finitary operations gives a denumerable subset \mathcal{E} of $\mathcal{P}(\omega)$ —a De Giorgi family—together with an injective enumeration h of \mathcal{E} . Since \mathcal{E} is closed under \mathcal{F}_1 – \mathcal{F}_8 , the extensional structure $\langle \omega, E_h \rangle$ is a model of the instances of comprehension P1–P8.

In the next section we will extract, from this informal procedure, a formal theorem that provides a general account of Beneš's construction.

4 A General Consistency Theorem

Let \mathcal{H} be a map that assigns to every De Giorgi family $\mathcal{E} \in \text{dom}(\mathcal{H})$ a set $H = \mathcal{H}(\mathcal{E})$ of injective enumerations of \mathcal{E} (or, equivalently, a set of \mathcal{E} -membership relations: we will use the same notation, $H_{\mathcal{E}}$, both for the set of the injective functions and for the corresponding set of membership relations).

Definition 4.1 A formula $\varphi(x, \vec{z})$ is said to be \mathcal{H} -monotone if and only if, for every $\mathcal{E}, \mathcal{E}' \in \text{dom}(\mathcal{H})$ and for every $E \in H_{\mathcal{E}}, E' \in H_{\mathcal{E}'}$,

$$\mathcal{E} \subseteq \mathcal{E}' \rightarrow \mathcal{F}_{\varphi, E} \subseteq \mathcal{F}_{\varphi, E'}.$$

Note that if a formula $\varphi(x, \vec{z})$ is \mathcal{H} -monotone, then it is, in particular, \mathcal{H} -invariant, that is,

$$\mathcal{F}_{\varphi, E} = \mathcal{F}_{\varphi, E'}$$

for every $\mathcal{E} \in \text{dom}(\mathcal{H})$ and for every $E, E' \in H_{\mathcal{E}}$.

Therefore, for a fixed \mathcal{H} , we can define a function $\mathcal{F}_{\varphi, \mathcal{E}} : \mathcal{E} \rightarrow \mathcal{P}(\omega)$ by setting

$$\mathcal{F}_{\varphi, \mathcal{E}}(\vec{Z}) = \mathcal{F}_{\varphi, E}(\vec{Z}),$$

since the definition does not depend on the choice of $E \in H_{\mathcal{E}}$.

Theorem 4.2 *Suppose that*

- (1) Φ is a collection of \mathcal{H} -monotone formulas;
- (2) $\text{dom}(\mathcal{H})$ is closed under countable unions;
- (3) $\text{dom}(\mathcal{H})$ is closed under the operator $\Gamma(\mathcal{E}) = \mathcal{E} \cup \bigcup_{\varphi \in \Phi} \text{ran}(\mathcal{F}_{\varphi, \mathcal{E}})$.

Then the theory $\mathbb{T} = \text{Ext} + \{\sigma_{\varphi} \mid \varphi \in \Phi\}$ is consistent.

Proof Since every $\varphi \in \Phi$ is \mathcal{H} -monotone, the operator Γ is well defined and monotone on $\text{dom}(\mathcal{H})$. Moreover, Γ is continuous; that is,

$$\Gamma\left(\bigcup_{n \in \omega} \mathcal{E}_n\right) = \bigcup_{n \in \omega} \Gamma(\mathcal{E}_n)$$

for any countable chain $\langle \mathcal{E}_n \mid n \in \omega \rangle$ of De Giorgi families in $\text{dom}(\mathcal{H})$.

Hence, from the general theory of monotone continuous operators (see, e.g., Davey and Priestley [2, Theorem 8.15, p. 183]) it follows that Γ has a fixed point; that is, there exists a De Giorgi family $\mathcal{E} \in \text{dom}(\mathcal{H})$ such that $\bigcup_{\varphi \in \Phi} \text{ran}(\mathcal{F}_{\varphi, \mathcal{E}}) \subseteq \mathcal{E}$. So, \mathcal{E} is closed under $\mathcal{F}_{\varphi, \mathcal{E}}$ for all $\varphi \in \Phi$. Hence, by Fact 3.3, any structure $\langle \omega, E \rangle$ induced by an injective enumeration h of \mathcal{E} belonging to $H = \mathcal{H}(\mathcal{E})$ is a model of (extensionality plus) every instance of comprehension σ_{φ} indexed by a formula in Φ . □

5 Beneš's Model

In this section we will show how Beneš's original consistency proof fits in the general framework provided by Theorem 4.2.

The statements of some of Hailperin's axioms¹ involve defined function symbols for the operations of singleton, ordered pair, and triple. So, in a sense, we need to fix these operations in ω before we define the membership relation E .

In order to do this, the key observation is that the operation of pairing— $(z, u) \mapsto \{z, u\}$ —is definable by an instance of comprehension indexed by a formula that *does not involve any membership relation*:

$$\{z, u\} = y \leftrightarrow \forall x (x \in y \leftrightarrow x = u \vee x = z).$$

This fact can also be understood in terms of “invariance” of the extension associated to the pairing operation as follows. Let G be any definable set operation, where $G(u, z)$ is defined as the unique y such that $\varphi(u, z, y)$ holds, for some formula φ of LST. Then $G^E(u, z)$ denotes the set defined by φ in a model $\langle \omega, E \rangle$ of the existence

and uniqueness condition for φ ; that is, if $\langle \omega, E \rangle$ models that there exists a unique y such that $\varphi(u, z, y)$ holds, then, by definition,

$$G^E(u, z) = y \leftrightarrow \langle \omega, E \rangle \models \varphi(u, z, y).$$

In particular, if $\langle \omega, E \rangle$ is a model of extensionality and pairing, then

$$\{u, z\}^E = y \leftrightarrow \langle \omega, E \rangle \models \text{“}\forall x (x \in y \leftrightarrow x = u \vee x = z)\text{”}.$$

Then, whatever the membership relation E , the extension $(\{u, z\}^E)_E$ of the pair $\{u, z\}$ (in the sense of E) is always the set $\{u, z\}$:

$$y = \{u, z\}^E \leftrightarrow y_E = \{u, z\}.$$

Hence a sufficient condition for a De Giorgi family \mathcal{E} to induce models of pairing is that \mathcal{E} should include all subsets of ω of the form $\{u, z\}$ for $u, z \in \omega$. More formally, let $\mathcal{J} = \{\{u, z\} \mid u, z \in \omega\}$, and let $j : J \rightarrow \mathcal{J}$ be any bijection between an infinite (not cofinite) subset J of ω and \mathcal{J} itself.² In addition, define the map

$$[u, z] = j^{-1}(\{u, z\}).$$

Then the following holds.

Fact 5.1 Let \mathcal{E} be any De Giorgi family such that $\mathcal{J} \subseteq \mathcal{E}$, and let h be any injective enumeration of \mathcal{E} such that $h \upharpoonright J = j$. Then for every $u, z \in \omega$,

$$[u, z] = \{u, z\}^{E_h}.$$

In particular, $\langle \omega, E_h \rangle$ is a model of pairing.

Operations for singleton, Wiener–Kuratowski ordered pair, and Hailperin triple³ are definable, as usual, just in terms of the pairing function.

Definition 5.2 Let $[u, z]$ be a pairing function. Then define

- $[z] = [z, z]$,
- $\ulcorner u, z \urcorner = \llbracket [u], [u, z] \rrbracket$,
- $\ulcorner u, w, z \urcorner = \ulcorner \llbracket [u] \rrbracket, \ulcorner w, z \urcorner \urcorner$.

Fact 5.1 also extends to these operations.

Fact 5.3 Let \mathcal{E} be any De Giorgi family such that $\mathcal{J} \subseteq \mathcal{E}$, and let h be any injective enumeration of \mathcal{E} such that $h \upharpoonright J = j$. Then for every $u, w, z \in \omega$,

1. $[z] = \{z\}^E$,
2. $\ulcorner u, z \urcorner = \langle u, z \rangle^E$,
3. $\ulcorner u, w, z \urcorner = \langle u, w, z \rangle^E$.

All the Hailperin axioms, except extensionality, are instances of comprehension indexed by formulas that involve equality, membership, singleton, ordered pair, and triple. For each φ that indexes a Hailperin axiom, Beneš’s paper provides an operation $\mathcal{F}_\varphi : \mathcal{P}(\omega) \rightarrow \mathcal{P}(\omega)$ such that

$$\mathcal{F}_{\varphi, E} = \mathcal{F}_\varphi \upharpoonright E(\omega)$$

for every membership relation E such that $\{u, z\}^E = [u, z]$.

To be more precise, since the formulas that index Hailperin’s axioms P1.1 (universal set) and P.8 (Frege’s cardinal number 1) have no parameters, the corresponding functions \mathcal{F}_φ are in fact fixed subsets of ω , that is, ω itself and the set

$I = \{[x] \mid x \in \omega\}$, respectively. Therefore, we will only use the term *Beneš operations* to refer to those \mathcal{F}_φ -functions which correspond to Hailperin axioms other than P1.1 and P.8.⁴

Definition 5.4 We say that a De Giorgi family \mathcal{E} is a *Beneš family* if and only if

1. $\{\omega, I\} \cup \mathcal{J} \subseteq \mathcal{E}$,
2. \mathcal{E} is closed under all Beneš operations.

In what follows, *Beneš set theory* is short for the theory given by extensionality plus Hailperin's axioms P1–P8.

The following proposition follows immediately from Fact 3.3.

Proposition 5.5 *Let $\mathcal{E} \subseteq \mathcal{P}(\omega)$ be a Beneš family, and let $h : \omega \rightarrow \mathcal{E}$ be any injective enumeration of \mathcal{E} such that $h \upharpoonright J = j$.*

Then $\mathcal{M} = \langle \omega, E_h \rangle$ is a model of Beneš set theory.

Finally, the main result in Beneš's paper can be recast in De Giorgi's terms as follows.

Theorem 5.6 (Beneš) *There exists a Beneš family.*

For a proof of the consistency of Beneš set theory based on the general Theorem 4.2, observe that Fact 5.3 immediately suggests the following choice for the map \mathcal{H} (the *Beneš map*):

- $\text{dom}(\mathcal{H}) = \{\mathcal{E} \mid \mathcal{E} \text{ is a De Giorgi's family and } \mathcal{J} \subseteq \mathcal{E}\}$,
- $H_{\mathcal{E}} = \{h \mid h \text{ is an injective enumeration of } \mathcal{E} \text{ and } h \upharpoonright J = j\}$.

This choice of \mathcal{H} ensures that in all models $\langle \omega, E \rangle$ in which $E \in H_{\mathcal{E}}$ and $\mathcal{E} \in \text{dom}(\mathcal{H})$ the pairing operation $\{z, u\}$ (and the other terms defined from it) is defined and is the same in all models.

As a matter of fact, Beneš's setting suits the hypotheses of Theorem 4.2 in a very strong sense. On the one hand, it is easy to check that \mathcal{H} fulfills both requirements (2) and (3) in Theorem 4.2 *regardless of which collection Φ of formulas we are considering*.⁵

On the other hand, the \mathcal{H} -monotony of the formulas in Φ (requirement (1)) follows immediately from the fact that $\mathcal{F}_{\varphi, \mathcal{E}} = \mathcal{F}_\varphi \upharpoonright \mathcal{E}$ for every $\mathcal{E} \in \text{dom}(\mathcal{H})$.

This proves the claim that the consistency of Beneš set theory can be proved as a particular instance of Theorem 4.2.

We remark that in the case of Beneš set theory, since Φ satisfies requirement (1) in a particularly nice way, we can prove Theorem 5.6 simply by closing the set $\mathcal{J} \cup \{\omega, I\}$ under the finite family of finitary functions $\{\mathcal{F}_\varphi \mid \varphi \in \Phi\}$ (the family of all Beneš operations) in the standard way. This standard closure procedure also provides an injective enumeration of the resulting De Giorgi family \mathcal{E} without any appeal to the axiom of choice. This feature is crucial to Beneš's claim that his consistency proof can be carried out entirely *inside* NF, thus providing an indirect argument for the independence of Hailperin's axiom P9 from the others (see below).

In our setting, the essential use of the axiom of countable choice (AC_ω) is just a payoff for the general form in which we have chosen to state Theorem 4.2, but in single applications of the theorem the possibility remains open that we can dispense with any form of choice.

6 An Approach to the NF Consistency Problem

The most natural question that arises from Beneš's paper is: how can we get a model of full NF from a Beneš-like model?

Obviously, I do not have an answer to this question, but a few remarks might help clarify the problem.

By Fact 3.3, Hailperin's axiom P9 holds in an extensional structure $\langle \omega, E \rangle$ if and only if the set

$$H_E = \{ \ulcorner [v], w \urcorner \mid v E w \},$$

called *Hailperin's membership set*, belongs to $\mathcal{E} = E(\omega)$.

Let us denote by \mathcal{E}_0 the Beneš family obtained by closing the set $\mathcal{J} \cup \{ \omega, I \}$ under all the Beneš operations, and denote by h_0 the standard injective enumeration of \mathcal{E}_0 , as described in Beneš's paper. Then the general question about the existence of a Beneš-like model of P9 (hence, a model of NF) can be arranged in a sequence of (apparently) progressively less demanding questions.

Question 6.1

- (1) Does $H_0 = \{ \ulcorner [v], w \urcorner \mid v \in h_0(w) \}$ belong to \mathcal{E}_0 ?
- (2) If not, is there any other enumeration h of \mathcal{E}_0 such that $H_h \in \mathcal{E}_0$?
- (3) If not, is there any other pair (\mathcal{E}, h) such that \mathcal{E} is a Beneš family and $H_h \in \mathcal{E}$?
(Obviously, any countable model of NF would give rise to such a pair.)

Beneš model The claim of [1] is that NF is strong enough to construct (\mathcal{E}_0, h_0) . This fact gives an indirect proof of the independence of P9 from the other Hailperin axioms (unless NF is inconsistent). For, suppose towards a contradiction that P9 follows from Beneš theory and that NF is consistent. Given any model \mathcal{M} of NF we can construct (\mathcal{E}_0, h_0) , that is, a model of NF, inside \mathcal{M} using only the axioms of NF: but this contradicts Gödel's second incompleteness theorem for NF.

I think that this indirect argument does not lead to a definite negative answer to Question 6.1(1). Indeed, a positive answer to this question does not imply that P9 follows from Beneš theory, since the possibility remains open that other models of Beneš theory might exist in which P9 fails. Moreover, even if NF can construct (\mathcal{E}_0, h_0) , it might be that there is no proof in NF of the fact that P9 holds in (\mathcal{E}_0, h_0) . So, to give a negative answer to Question 6.1(1), we need a direct proof of the fact that $H_0 \notin \mathcal{E}_0$.

However, a positive answer to Question 6.1(1) seems unlikely. Suppose $H_0 \in \mathcal{E}_0$. Since H_0 is obviously different from ω , from I , and from all pairs in \mathcal{J} (by an immediate inspection of the form of its elements), it follows by construction that $H_0 = \mathcal{F}(X, Y)$ for some Beneš operation \mathcal{F} and some $X, Y \in \mathcal{E}$. Then $H_0 = h_0(p)$ for some $p \in \omega$ and $X = h(q), Y = h(q')$ for some $q, q' < p$, and so it would be surprising if H_0 correctly guessed the membership relation h also for elements $r > p$. As a matter of fact, a further examination of the form of the elements of H_0 forces us to rule out $\mathcal{F}_2, \mathcal{F}_4 - \mathcal{F}_7$, and \mathcal{F}_{10} , but I have not found any obvious reason (nor any inductive argument) to also rule out the remaining possibilities: $\mathcal{F}_1, \mathcal{F}_3$, and $\mathcal{F}_8 - \mathcal{F}_9$.

Permutations We can recast Question 6.1(2) in terms of *permutation models* as follows.

Recall that a *permutation of ω* is just a bijection $\tau : \omega \rightarrow \omega$ and that given any membership relation E and a permutation τ of ω we define

$$x E_\tau y \leftrightarrow x E \tau(y).$$

The structure $\langle \omega, E_\tau \rangle$ is said to be a *permutation model* of $\langle \omega, E \rangle$.

Given an injective enumeration h of \mathcal{E}_0 such that $h \upharpoonright J = h_0 \upharpoonright J = j$, let $\tau_h = h_0^{-1} \circ h$. Conversely, given a permutation τ of ω such that $\tau \upharpoonright J = \text{id}_J$, let $h_\tau = h_0 \circ \tau$.

Fact 6.2 The maps $h \mapsto \tau_h$ and $\tau \mapsto h_\tau$ are one-to-one and inverse to each other.

By Fact 6.2, \mathcal{H} -invariance can be stated equivalently either in terms of injections or in terms of permutations. The latter form emphasizes the fact that, even though the set $H_\mathcal{E} = \mathcal{E}$ obviously depends on the De Giorgi family $\mathcal{E} \in \text{dom}(\mathcal{H})$, the corresponding set of permutations of ω is always the same $G = \{\tau \mid \tau \upharpoonright J = \text{id}_J\}$. Moreover, G is a *group* of permutations.

Hence Question 6.1(2) amounts to requiring the existence of a permutation $\tau \in G$ of the Beneš model $\langle \omega, E_0 \rangle$ such that the set $H_\tau = \{\ulcorner v \urcorner, w^\top \mid v E_\tau w\}$ belongs to \mathcal{E}_0 .

Beneš families By construction, \mathcal{E}_0 is the *least Beneš family* on ω ; that is $\mathcal{E}_0 \subseteq \mathcal{E}$ for any other Beneš family \mathcal{E} . The order-theoretic properties of the space of all Beneš families are collected in more detail in the following.

Proposition 6.3 *Let \mathcal{B} be the poset of all Beneš families on ω ordered by inclusion.*

- (1) \mathcal{B} is a lattice with respect to intersections and closure of unions. In particular, \mathcal{E}_0 is the least element of \mathcal{B} .
- (2) \mathcal{B} is bounded complete.
- (3) (AC) \mathcal{B} is countable complete.
- (4) (AC) \mathcal{B} has cardinality greater than or equal to ω_1 .
- (5) (AC) There are no maximal Beneš families.

Obviously, we can look for a permutation model of NF—as described in the previous paragraph—by starting from any Beneš family \mathcal{E} , not just from \mathcal{E}_0 . But the converse strategy also gives rise to a question of some interest.

Question 6.4 Is it true that if NF is consistent, then NF has a model that comes from a Beneš-like model (where P9 is false) and a suitable permutation?

The obvious attempts to prove this conjecture fail. By contrast, suppose that P9 is \mathcal{H} -invariant in every Beneš-like model; then we could use this fact to get NF-independence proofs—for both stratified and unstratified sentences.

Back to the original problem of the consistency of NF, we observe that Beneš's construction can also be understood as a solution h_0 of the following *system of equations*:

- $h(0) = \omega$,
- $h(1) = I$,
- $h(2p + 2) = j(2p + 2)$,
- $h(2p + 3) = F(\min\{w \mid F(w) \notin \mathcal{A} \cup \{h(2q + 3) \mid q < p\}\})$,

where the last equation is just a formalization of the standard procedure to enumerate the closure $\mathcal{E}_0 = \text{Cl}_{\mathcal{F}}(\mathcal{A})$ of the set $\mathcal{A} = \mathcal{J} \cup \{\omega, I\}$ under the family \mathcal{F} of all Beneš operations. This formalization is obtained using a fixed map F from the set of words on the alphabet $(\mathcal{A}, \mathcal{F})$ into \mathcal{E}_0 .

The existence of a solution follows from the recursion theorem. In fact, ω -recursion suffices, and the solution is unique, that is, the system of equations is a *recursive definition* of h_0 .

The freedom in the choice of the Beneš family we want to build allows us to recast the NF consistency problem as a fixed point problem, by mimicking Beneš's system of equations, as follows. Let \mathcal{S} be the following *system of equations*:

- $h(0) = \omega$,
- $h(1) = I$,
- $h(2) = H_h = \{\ulcorner [m], n \urcorner \mid m \in h(n)\}$,
- $h(2n + 4) = j(2n + 4)$,
- $h(2n + 3) = F(\min\{w \in W \mid F(w) \notin \{h(2m) \mid m \in \omega\} \cup \{h(2m + 1) \mid m \leq n\}\})$,

where $W = \text{Words}(S, a)$, (S, a) is the alphabet induced by $(\mathcal{A}, \mathcal{F})$, \mathcal{F} is the family of all Beneš operations, and $\mathcal{A} = \{\omega, I, \mathbf{a}\} \cup \mathcal{J}$, where \mathbf{a} is just a *symbol*, that is, any set not involved in the construction.

The function F is defined by recursion as

$$F(w) = \begin{cases} w & \text{if } w \in \{\omega, I\} \cup \mathcal{J}, \\ H_h & \text{if } w = \mathbf{a}, \\ \mathcal{F}(F(w_1), F(w_2)) & \text{if } w = \langle \mathcal{F} \rangle \hat{w}_1 \hat{w}_2. \end{cases}$$

So the system of equations \mathcal{S} is *circular* not only because h explicitly occurs on the right-hand side of the third and the last equation but also because the “parameters” $\mathcal{E} = \text{Cl}_{\mathcal{F}}(\mathcal{A})$ and F depend on h .

Fact 6.5 Any solution h of the system of equations \mathcal{S} gives rise to a model of NF.

7 Beneš-Like Constructions

Another natural question about Beneš's construction is the following: which collections of instances of comprehension can be proved to be consistent using Beneš's map \mathcal{H} and Theorem 4.2?

In other words, we are looking at classes of formulas that can be proved to be \mathcal{H} -monotone with our choice of \mathcal{H} .

Consider the algebraic expansion \mathcal{L} of LST by function symbols for singleton, ordered pair, and triple, and define an \mathcal{L} -stratification for an \mathcal{L} -formula $\varphi(x, \vec{z})$ to be an assignment of natural numbers (types) to the \mathcal{L} -terms that occur in φ such that

- a term and all its subterms receive the same type;
- if $s = t$ occurs in φ , then s and t receive the same type;
- if $s \in t$ occurs in φ , then the type of t is equal to the type of s plus 1.

In other words, we regard the terms of \mathcal{L} as *homogeneous* and *flat* operations; namely, all their arguments must receive the same type, and these operations do not raise the type of their arguments.⁶

The intended interpretation of an \mathcal{L} -formula φ is a structure \mathcal{M} of the form $\langle \omega, E, [\cdot, \cdot] \rangle$, where E is a membership relation and $[\cdot, \cdot]$ is a fixed pairing operation defined independently of E . The symbols for singleton, ordered pair, and triple of \mathcal{L} are then interpreted in \mathcal{M} as the corresponding operations defined by equality and $[\cdot, \cdot]$ alone, as in Definition 5.2.

Beneš's map \mathcal{H} forces us to consider only those membership relations E that agree with $[\cdot, \cdot]$. More in detail, by Fact 5.1, for every $\mathcal{E}, \mathcal{E}' \in \mathcal{H}$ and for every $E \in H_{\mathcal{E}}, E' \in H_{\mathcal{E}'}$,

$$\{x, y\}^E = [x, y] = \{x, y\}^{E'}.$$

So the interpretations of singleton, ordered pair, and triple are the same, too, and, in general, given any formula φ of LST, the interpretations of φ as an LST-formula and as an \mathcal{L} -formula coincide:

$$\langle \omega, E, [\cdot, \cdot] \rangle \models \varphi[\vec{x}] \leftrightarrow \langle \omega, E \rangle \models \varphi[\vec{x}].$$

We say that an \mathcal{L} -formula φ is *$\mathcal{L}k$ -stratifiable* if and only if φ admits an \mathcal{L} -stratification that uses no more than k types.

Lemma 7.1 *Let $\varphi(x, \vec{z})$ be an $\mathcal{L}2$ -stratifiable \mathcal{L} -formula such that (in the canonical stratification of φ with types 0 and 1):*

1. *the variable x , all compound terms, and all bounded variables receive type 0;*
2. *all parameters \vec{z} receive type 1.*

Then φ is \mathcal{H} -monotone.

Proof In order to argue by induction on the complexity of the formula φ , we start by proving the following slightly more general claim.

Claim *Let φ be any $\mathcal{L}2$ -stratifiable \mathcal{L} -formula such that (in the canonical stratification of φ with types 0 and 1) all compound terms and all bounded variables receive type 0. Denote by \vec{x} the list of the free variables of φ that receive type 0 and by \vec{z} the list of the free variables of φ that receive type 1. Then, for every $\mathcal{E}, \mathcal{E}' \in \text{dom}(\mathcal{H})$ and for every $h \in H_{\mathcal{E}}, h' \in H_{\mathcal{E}'}$, if $\mathcal{E} \subseteq \mathcal{E}'$, then*

$$\langle \omega, E_h \rangle \models \varphi[\vec{a}, h^{-1}(\vec{Z})] \leftrightarrow \langle \omega, E_{h'} \rangle \models \varphi[\vec{a}, (h')^{-1}(\vec{Z})]$$

for every $\vec{Z} \in \mathcal{E}$ such that $h^{-1}(\vec{Z})$ and $(h')^{-1}(\vec{Z})$ have to be assigned to the \vec{z} -variables and for every $\vec{a} \in \omega$ that has to be assigned to the \vec{x} -variables.

Proof of the claim The proof is by induction on φ . Let $E = E_h$ and $E' = E_{h'}$.

If φ is atomic of the form $s(\vec{v}) = t(\vec{w})$, then by $\mathcal{L}2$ -stratifiability φ must be of the form $s(\vec{x}) = t(\vec{x})$ (i.e., all the terms involved and, in particular, all the variables have type 0). Recall that all the terms of \mathcal{L} receive the same interpretation under E and under E' , so in this case the claim holds trivially.

If φ is atomic of the form $s(\vec{v}) \in t(\vec{w})$, then $\mathcal{L}2$ -stratifiability forces φ to be of the form $s(\vec{x}) \in z$. If s again denotes the common interpretation of the term $s(\vec{x})$ under E and under E' , then

$$\begin{aligned} \langle \omega, E \rangle \models \varphi[\vec{a}, h^{-1}(\vec{Z})] &\leftrightarrow s(\vec{a}) \in h^{-1}(Z) \leftrightarrow s(\vec{a}) \in Z \\ &\leftrightarrow s(\vec{a}) \in (h')^{-1}(Z) \leftrightarrow \langle \omega, E' \rangle \models \varphi[\vec{a}, (h')^{-1}(\vec{Z})]. \end{aligned}$$

The negation, conjunction, and existential quantifier cases follow from the inductive hypothesis. In the existential case ($\varphi \equiv \exists y \psi(y, \vec{x}, \vec{z})$) the inductive step follows

from the fact that the variable y is bounded in φ , so by hypothesis it receives type 0 like the variables in \vec{x} . ■

Let $\varphi(x, \vec{z})$ be as in the hypotheses of the lemma. Then \mathcal{H} -monotony of φ follows immediately from the claim, since φ satisfies the assumptions in the claim. □

Lemma 7.1 provides a syntactical condition that suffices to prove \mathcal{H} -monotony for all the stratified formulas that index Hailperin's axioms P1–P8. I am indebted to the anonymous referees for the observation that this sufficient condition is also likely to be *necessary* for \mathcal{H} -monotony, but I have no proof of this conjecture.

By contrast, we observe that the notion of \mathcal{H} -monotony must not be understood as a kind of restriction of the notion of stratified formula.

Formulas in the language of pure equality are trivially stratified. But operations definable via instances of such formulas involve the membership relation in their definitions, so terms are no longer necessarily stratifiable. However, they can be both invariant and monotone.

Beneš suggests an alternative way to prove the independence of P9, and in the process he provides a nice example of an \mathcal{H} -monotone unstratifiable formula.

At the end of his paper, Beneš recalls that the class

$$U = \{\{\{x\}, x\} \mid x \in \mathbf{V}\}$$

is not a set in NF (it represents the ι -operation) and suggests that we can build a model of extensionality, Hailperin's P1–P8, and the existence of U by an obvious modification of his own construction.

Indeed, the formula φ that defines U is easily seen to be unstratified (when translated in LST). However, as an \mathcal{L} -formula φ fulfills the requirements in Lemma 7.1. Hence φ is \mathcal{H} -monotone, so we can apply Theorem 4.2 to get a model of Beneš theory that also contains the graph of the ι -operation as a set.

Since we can prove both invariance and monotony for the unstratified formula defining U , we can conclude that both these properties are not subproperties of stratifiability.

Hence, the syntactical \mathcal{H} -monotony condition provided by Lemma 7.1 characterizes a consistent class of instances of comprehension strictly larger than Beneš theory and incompatible with NF.

8 Conclusion

By taking the construction described in Beneš's paper as a case study, Sections 6 and 7 illustrate a few questions about models of set theories that can be conveniently addressed within the scope of Theorem 4.2.

However, Beneš map \mathcal{H} should be understood just as a paradigm, and it is possible that investigations both about alternative set theories and about the NF consistency problem may profit from the freedom in the choice of the map \mathcal{H} allowed by the theorem.

More restrictive conditions on the map \mathcal{H} might allow us to apply Theorem 4.2 to a larger collection of \mathcal{H} -monotone formulas. This would be useful for the general study of the consistency of fragments of comprehension as well as for tackling the problem of the consistency of NF by starting from a different partial model of this theory.

Appendix A Hailperin's Axioms

We have the following:

- P0. $\forall x, y (\forall u (u \in x \leftrightarrow u \in y) \rightarrow x = y)$,
- P1. $\forall x, y \exists z \forall u (u \in z \leftrightarrow u \in x \mid u \in y)$,
- P2. $\forall x \exists z \forall u, v (\langle \iota u, \iota v \rangle \in z \leftrightarrow \langle u, v \rangle \in x)$,
- P3. $\forall x \exists z \forall u, v, w (\langle u, v, w \rangle \in z \leftrightarrow \langle u, v \rangle \in x)$,
- P4. $\forall x \exists z \forall u, v, w (\langle u, w, v \rangle \in z \leftrightarrow \langle u, v \rangle \in x)$,
- P5. $\forall x \exists z \forall u, v (\langle v, u \rangle \in z \leftrightarrow u \in x)$,
- P6. $\forall x \exists z \forall u (u \in z \leftrightarrow \forall v (\langle v, \iota u \rangle \in x))$,
- P7. $\forall x \exists z \forall u, v (\langle v, u \rangle \in z \leftrightarrow \langle u, v \rangle \in x)$,
- P8. $\exists z \forall u (u \in z \leftrightarrow \exists v (u = \iota v))$,
- P9. $\exists z \forall u, v (\langle \iota u, v \rangle \in z \leftrightarrow u \in v)$.

Axiom P0 is just the axiom of extensionality. Axiom P1 uses Sheffer's stroke \mid to state that the universe is closed under all Boolean operations. In the current literature this axiom is usually split into the conjunction of the universal set and difference axioms

- P1.1. $\exists z \forall u (u \in z \leftrightarrow u = u)$,
- P1.2. $\forall x, y \exists z \forall u (u \in z \leftrightarrow u \in x \ \& \ u \notin y)$.

In axioms P2–P9 the symbols ι , $\langle \cdot, \cdot \rangle$, $\langle \cdot, \cdot, \cdot \rangle$ denote singleton, Wiener–Kuratowski ordered pair, and Hailperin triple, respectively. These terms are definable as follows, assuming the existence of the unordered pair $\{x, y\}$:

- $\iota(x) = \{x, x\}$,
- $\langle x, y \rangle = \{\iota(x), \{x, y\}\}$,
- $\langle x, y, z \rangle = \langle \iota(x), \langle y, z \rangle \rangle$.

Axiom P8 states the existence of Frege's cardinal number 1. Axiom P9 states the existence of the Hailperin's membership set, a set that codes the graph of the membership relation.

Appendix B Beneš Operations

We have the following:

- $\mathcal{F}_1(X, Y) = X - Y$,
- $\mathcal{F}_2(X, Y) = \{\ulcorner m, n \urcorner \mid n \in X\}$,
- $\mathcal{F}_3(X, Y) = \{\ulcorner m, n \urcorner \mid \ulcorner n, m \urcorner \in X\}$,
- $\mathcal{F}_4(X, Y) = \{\ulcorner m, n, p \urcorner \mid \ulcorner m, n \urcorner \in X\}$,
- $\mathcal{F}_5(X, Y) = \{\ulcorner m, n, p \urcorner \mid \ulcorner m, p \urcorner \in X\}$,
- $\mathcal{F}_6(X, Y) = \{\ulcorner m, n, p \urcorner \mid \ulcorner n, p, m \urcorner \in X\}$,
- $\mathcal{F}_7(X, Y) = \{\ulcorner m, n, p \urcorner \mid \ulcorner m, p, n \urcorner \in X\}$,
- $\mathcal{F}_8(X, Y) = \{k \mid \exists n (\ulcorner n, k \urcorner \in X)\}$,
- $\mathcal{F}_9(X, Y) = \{k \mid \exists n (\ulcorner n, [k] \urcorner \in X)\}$,
- $\mathcal{F}_{10}(X, Y) = \{\ulcorner [m], [n] \urcorner \mid \ulcorner m, n \urcorner \in X\}$.

Notes

1. See Appendix A for the complete list of Hailperin's axioms as originally stated in [4].

2. The details of this bijection are not relevant; to strictly follow Beneš's paper we had to take as J the set of even numbers greater than 0 and to define j using Gödel's coding of the ordered pairs of natural numbers.
3. See Appendix A.
4. See Appendix B for the complete list of Beneš operations as originally stated in [1]. As remarked by Beneš himself, this list also includes some operations other than those strictly derived from Hailperin's axioms but at no extra cost.
5. In fact, to verify requirement (2) in the current more general setting, we need at least AC_ω , while this assumption is not needed for Beneš's original proof; see further on this point.
6. All of this has no evident relation with the notion of a homogeneous and flat operation *definable* in LST (like Quine's ordered pair, for instance) nor with the notion of a *stratified language* introduced in Pétry [5].

References

- [1] Beneš, V. E., "A partial model for Quine's 'New Foundations,'" *Journal of Symbolic Logic*, vol. 19 (1954), pp. 197–200. [MR 0064716](#). [397](#), [400](#), [404](#), [410](#)
- [2] Davey, B. A., and H. A. Priestley, *Introduction to Lattices and Order*, 2d ed., Cambridge University Press, New York, 2002. [MR 1902334](#). [401](#)
- [3] Forster, T. E., *Logic, Induction and Sets*, vol. 56 of *London Mathematical Society Student Texts*, Cambridge University Press, Cambridge, 2003. [Zbl 1026.03002](#). [MR 1996832](#). [398](#)
- [4] Hailperin, T., "A set of axioms for logic," *Journal of Symbolic Logic*, vol. 9 (1944), pp. 1–19. [MR 0009753](#). [397](#), [409](#)
- [5] Pétry, A., "Stratified languages," *Journal of Symbolic logic*, vol. 57 (1992), pp. 1366–76. [MR 1195275](#). [DOI 10.2307/2275371](#). [399](#), [410](#)
- [6] Quine, W. V., "New foundations for mathematical logic," *American Mathematical Monthly*, vol. 44 (1937), pp. 70–80. [MR 1523848](#). [DOI 10.2307/2300564](#). [397](#), [398](#), [399](#)
- [7] Rosser, J. B., "On the consistency of Quine's 'New Foundations for Mathematical Logic,'" *Journal of Symbolic Logic*, vol. 4 (1939), pp. 15–24. [399](#)

Acknowledgments

Author's work supported by the European Science Foundation "New Frontiers of Infinity: Mathematical, Philosophical and Computational Prospects" grant 3291. The content of this paper was presented at the "75th NF Birthday Conference" held at Cambridge University, March 26–31, 2012. I wish to thank T. E. Forster and all members of the University of Cambridge Set Theory Group for teaching me NF during my visit in 2011.

Classe di Lettere
Scuola Normale Superiore
Piazza dei Cavalieri, 7
56126 Pisa
Italy
rivello.edoardo@gmail.com