

HOL in HOL over a set-theoretic universe: consistency of NK with infinity and choice

Christoph Benzmüller

Daniel Kirchner

October 3, 2026

Abstract

This companion to the entry *A deep embedding of HOL in HOL* (HOL_in_HOL_Deep) [2] certifies the natural-deduction calculus NK of that entry consistent with a genuine Dedekind-style *axiom* of infinity—an injective, non-surjective self-map of the individuals—and, because its standard model has full function spaces, with the relational axiom of *choice* as well. Such an axiom has *no* finite models, so the compactness argument of the main entry—which lifts arbitrarily large *finite* models to an infinity *scheme*—does not apply: a genuinely *infinite* model is required. We construct one over Paulson’s ZFC_in_HOL [8]: the individuals are interpreted as the finite ordinals ω , with full function spaces at every type, yielding a *standard* Σ -model of NK over the set-theoretic universe V . As this meta-theory is stronger and non-conservative, the guarantee is *relative* consistency at ZF strength; the machine-checked formalisation lives in Isabelle/HOL with Paulson’s ZFC_in_HOL (the ZF reading is metatheoretic, made precise in the introduction). Because NK already provides extensionality and typed description (after Andrews), this certifies the consistency of *full* classical higher-order logic in Church’s sense, choice included as a separate relational scheme.

The infinity axiom and the weaker infinity *scheme* of the main entry turn out to be logically independent. That the scheme does not entail the axiom is shown in the main entry; the present entry settles the other direction, since the axiom does not entail any of the scheme’s individual inequalities: its set-theoretic model interprets all individual constants alike, so those inequalities fail while the axiom holds. Neither sentence, then, implies the other.

The plain-HOL results—soundness, Henkin completeness, and finite-model consistency, including the compactness route to the infinity scheme—are proved in the main entry [2] and imported here *unchanged*. What the present entry *adds* is the set-theoretic route to axioms that plain HOL alone cannot model.

This contribution is part of the LogiKEy project (<https://logikey.org>).

Use of generative AI. Parts of this formalisation and its documentation were drafted with a generative AI assistant (Anthropic’s Claude) under continuous human direction. All proofs are machine-checked by Isabelle; the entire development, including all comments and references, was reviewed, revised and simplified by the authors, who take full responsibility for its content.

Contents

1	Introduction	3
2	Full classical HOL, abstractly: the choice scheme, the axioms’ satisfaction, and <i>hol-universe</i>	5

3	Consistency of full classical HOL: infinity and choice	5
4	Minimising the meta-theory: a weaker universe suffices	8
5	Minimising further: a Harrison-style universe axiom	8
6	The set-theoretic instantiation: V as a <i>hol-universe</i>	11

1 Introduction

Two entries, and where each result is proved. This is the second of two Archive of Formal Proofs entries on a deep embedding of classical higher-order logic—Church’s simple theory of types, following Benzmüller, Brown and Kohlhase [1]—inside Isabelle/HOL. The first entry, *A deep embedding of HOL in HOL* (`HOL_in_HOL_Deep`) [2], stays *entirely within plain Isabelle/HOL, with no set-theoretic universe*. There the natural-deduction calculus NK is proved sound, Henkin-complete and consistent; consistency comes from a concrete finite standard model, and, by compactness over arbitrarily large *finite* models, NK is shown consistent even with an infinity *scheme* $\{c_i \neq c_j \mid i \neq j\}$ —still without ever constructing an infinite model. The present entry adds one result that plain HOL cannot reach: the consistency of NK with a genuine Dedekind-style *axiom* of infinity, and with the axiom of *choice*. It imports the first entry unchanged and supplies the set-theoretic model that this result needs.

We are careful throughout—here, in the theory comments, and in both abstracts—to say which of two tiers a statement belongs to: what holds in plain HOL, and what needs the set-theoretic universe. Within this entry the split is visible in the theory structure itself. The set-theory-free half, theory `HOL_Universe`, imports only the first entry and never `ZFC_in_HOL`. It contains the choice scheme with its satisfaction condition (the infinity axiom and its satisfaction are inherited from the first entry), the abstract relative-consistency theorem, the completeness results, and two universe interfaces: the locale `hol_universe`, and the (at most as strong) Harrison-style axiom `harrison_universe`, after Harrison’s self-verification of HOL Light [4], from which `hol_universe` is in turn *constructed*. The set-theoretic universe itself appears only in the second theory, which instantiates both interfaces with Paulson’s set-theoretic universe V (the type of all sets of `ZFC_in_HOL`, introduced below).

Why the meta-theory must be strengthened. A single Dedekind-style axiom of infinity has *no* finite models, by the pigeonhole principle. So the route of the first entry—compactness over finite models—yields nothing here, and a genuinely *infinite* model is required. We do not try to build one inside plain HOL, for two reasons. First, no type that plain HOL can construct is large enough to carry a universe of full function spaces: by Cantor’s theorem each such type is outgrown by the tower of function spaces built over it, and that no other plain-HOL route succeeds follows from Gödel’s second theorem (discussed below). Second, the term-model construction of the completeness proof would be circular for a single sentence. Instead we borrow a ready-made, stronger meta-theory: Paulson’s `ZFC_in_HOL` [8], which axiomatises a Zermelo–Fraenkel universe as a type V . This is *not* a conservative extension of HOL— V is postulated by axioms, and ZF is proof-theoretically stronger than plain HOL—so what we obtain is a *relative* consistency result: NK with infinity is consistent *relative to* ZF. (The formalisation rests on the full `ZFC_in_HOL`, whose set-theoretic choice axiom we never invoke essentially; that the construction in fact needs only ZF strength—indeed less—is made precise in *How little of ZF is used* below. This is why we write ZF, not ZFC.) This pins down exactly where set theory is used, namely only to certify consistency (an axiom without finite models has no finite consistency witness). The model that realises the consistency is, by contrast, a *countable* Henkin model built in plain HOL, by the (ZFC-free) model-existence half of the first entry. For that single axiom the stronger meta-theory supplies the consistency premise and nothing else.

Choice, and full HOL. The same standard model settles more. Because its function spaces are *full*, it validates every instance of the relational axiom of *choice*: the required choice function is simply the set-theoretic λ -abstraction of a meta-level Hilbert choice. Hence NK with the

Dedekind axiom of infinity *and* choice is consistent relative to ZF. Since NK already carries extensionality and typed description, this is the consistency of full classical higher-order logic in the sense of Church. To be precise, by *full classical HOL in the sense of Church* we mean exactly this system: Church’s monomorphic simple theory of types [3]—the calculus NK with Boolean and functional extensionality and typed description, extended by the Dedekind axiom of infinity and the relational axiom of choice. It is *not* the polymorphic Isabelle/HOL of the meta-level, which goes beyond Church’s system with type variables, type definitions and overloading.

Two tiers of consistency, and Gödel’s second theorem. Write *O-HOL* for the embedded object logic (whose calculus is NK) and *M-HOL* for the meta-logic Isabelle/HOL in which it lives; the two must be kept apart. The first entry proves $M-HOL \vdash \text{con}(O-HOL)$: the bare object logic is consistent, and this is shown in plain HOL with *no* set universe. (With a single individual every domain is finite, so M-HOL exhibits a model directly; and O-HOL, having finite models, lacks the arithmetic strength that Gödel’s second theorem presupposes.) The present entry adds the *full* logic: $M-HOL + V \vdash \text{con}(O-HOL + \text{infinity} + \text{choice})$, that is, M-HOL extended by the set universe V proves full classical HOL consistent. This is as close as one can come to a self-standing $\vdash \text{con}(HOL)$, and it is emphatically *not* the object logic proving its own consistency.

That the meta-logic here cannot be plain M-HOL either is a *metatheoretic* observation, outside the machine-checked development. With infinity and choice, O-HOL is itself Church’s simple type theory with infinity and choice, and this interprets the M-HOL core (its individuals becoming an infinite type, its function types the function types). So the two share their proof-theoretic strength: by the classical analysis both are equiconsistent with a Zermelo-style set theory [7]. (Here “M-HOL” means the HOL *core*, whose definitional and overloading apparatus is separately known consistent [6].) Were plain M-HOL to prove $\text{con}(O-HOL + \text{infinity} + \text{choice})$, it would thereby prove its own consistency, which Gödel’s second incompleteness theorem forbids; so the step up to the set universe V is *necessary*, not a convenience. The same obstacle, and the same step to a stronger meta-theory, appears in Harrison’s self-verification of HOL Light [4] and its successors [5]. Throughout it is M-HOL reasoning *about* O-HOL, one level up—a difference of role, not of logical power: the object logic is in no way the weaker system. The certificate is tied to exactly the axioms modelled here. Extending O-HOL with a further axiom stays legitimate, but the machine-checked guarantee does not extend automatically: one must exhibit a model of the new axiom, and if the extension reaches the consistency strength of $M-HOL + V$, certifying it needs a yet stronger meta-theory—again by Gödel’s second theorem.

How little of ZF is used. A closer look pins down the exact surplus over plain HOL. The object-side argument is carrier-generic and uses no set theory at all: full HOL is consistent as soon as *some* standard Σ -frame has Dedekind-infinite individuals (an injective, non-surjective self-map). Plain HOL has infinity already; what it lacks is a single type gathering the full function spaces over an infinite base (a model formalised in HOL keeps all its domains inside one carrier type)—by Cantor these outgrow every type it can construct—and that is the whole reason for a stronger universe. Choice is *not* part of this surplus: the same full model validates it for free, so the price of full HOL over plain HOL is the axiom of infinity alone. Nor need that universe be full ZF. It is enough to assume full function spaces, booleans, extensionality, infinitely many individuals and a domain-respecting parameter interpretation—no replacement and no ordinals—and from these the logical constants (negation, disjunction, quantification, equality and description) are *constructed* rather than assumed. Every domain then has bounded rank, and this is machine-checked: the domain of individuals sits at rank ω , and each function-space layer adds only finitely

much, so $\text{rank}(D_\sigma) < \omega + \omega$ for every type σ , and every domain is an element of the set $V_{\omega \cdot 2}$. Two things should be kept apart here. What is *machine-checked* is precisely this rank bound together with Isabelle/HOL + ZFC_in_HOL $\vdash \text{con}(\text{O-HOL} + \text{infinity} + \text{choice})$. That the same consistency follows from far less than ZF is a *metatheoretic* reading of that proof, read off from the ranks of the objects it constructs: a single derivation mentions only finitely many types, so its refutation is validated using only sets of bounded rank—comfortably within Zermelo set theory, and without ever forming $V_{\omega \cdot 2}$ as a completed set (which would already call on a trace of replacement). This reading is not itself a formalised relative-consistency theorem. Under it the relative consistency factors as $\text{con}(\text{ZF}) \Rightarrow \text{con}(\text{weak universe}) \Rightarrow \text{con}(\text{full HOL})$; the first implication also absorbs the choice that the meta-theory uses throughout—by Gödel’s constructible universe L at ZF strength, while at Zermelo strength the corresponding elimination of choice without replacement is genuinely delicate (see Mathias [7]). The surplus strength of ZFC_in_HOL—which *exceeds* ZF, as it proves $\text{con}(\text{ZFC})$ —is never used.

2 Full classical HOL, abstractly: the choice scheme, the axioms’ satisfaction, and *hol-universe*

This theory is the *set-theory-free* half of the companion development: it imports only the main entry HOL_in_HOL_Deep and never mentions a set-theoretic universe. The Dedekind axiom of infinity *DInf*, its satisfaction conditions and its separation from the infinity scheme are inherited from the main entry (theory *NK-Infinity*); here the relational choice scheme *ACrel* is added as an object formula with its satisfaction condition in an *arbitrary* Σ -standard model, the abstract relative-consistency theorem is derived — full classical *HOL* is consistent as soon as *some* standard model has Dedekind-infinite individuals — and the universe interfaces are isolated from which the required standard model is *constructed*: the locale *hol-universe*, and beneath it an at most as strong Harrison-style locale *harrison-universe*. Completeness for the extension by *DInf* is likewise established here, in plain HOL. That the assumptions of these locales are satisfiable is a separate question — no *plain-HOL* type answers it: Cantor’s tower outgrows every type plain HOL can construct, and Gödel’s second theorem bars every other plain-HOL route (see the document introduction) — and is deferred to this entry’s final theory, *HOL-in-HOL-Deep-Infinity-ZFC*, which instantiates both locales with Paulson’s axiomatic set-theoretic universe V . The split makes the dependency structure machine-visible: nothing in this theory rests on *ZFC-in-HOL*.

3 Consistency of full classical HOL: infinity and choice

NK already carries extensionality (property *f*) and typed description. Adding the Dedekind axiom of infinity *DInf* and the axiom of *choice* makes the object logic full classical higher-order logic in the sense of Church. The consistency of the whole package is reduced, in one carrier-generic argument, to the existence of a standard frame with Dedekind-infinite individuals.

3.1 The axiom of choice as a relational scheme

Choice is stated, without any new primitive, as the relational scheme $AC_{\sigma, \tau} = \Pi R. (\Pi X. \exists Y. R \cdot X \cdot Y) \supset (\exists F. \Pi X. R \cdot X \cdot (F \cdot X))$ at every pair of types.

definition *rR* where $rR = (5::nat)$

definition *rX* where $rX = (6::nat)$

definition *rY* where $rY = (7::nat)$

definition rF where $rF = (8::nat)$

definition $ACatom1 :: ty \Rightarrow ty \Rightarrow 'p\ tm$ where

$ACatom1\ \sigma\ \tau = (((rR^f\sigma \Rightarrow \tau \Rightarrow o) \cdot (rX^f\sigma)) \cdot (rY^f\tau))$

definition $ACatom2 :: ty \Rightarrow ty \Rightarrow 'p\ tm$ where

$ACatom2\ \sigma\ \tau = (((rR^f\sigma \Rightarrow \tau \Rightarrow o) \cdot (rX^f\sigma)) \cdot ((rF^f\sigma \Rightarrow \tau) \cdot (rX^f\sigma)))$

definition $ACprem :: ty \Rightarrow ty \Rightarrow 'p\ tm$ where

$ACprem\ \sigma\ \tau = \prod rX_\sigma. \exists rY_\tau. ACatom1\ \sigma\ \tau$

definition $ACconcl :: ty \Rightarrow ty \Rightarrow 'p\ tm$ where

$ACconcl\ \sigma\ \tau = \exists rF_\sigma \Rightarrow \tau. \prod rX_\sigma. ACatom2\ \sigma\ \tau$

definition $ACrel :: ty \Rightarrow ty \Rightarrow 'p\ tm$ where

$ACrel\ \sigma\ \tau = \prod rR_\sigma \Rightarrow \tau \Rightarrow o. (ACprem\ \sigma\ \tau \supset ACconcl\ \sigma\ \tau)$

$\langle proof \rangle \langle proof \rangle \langle proof \rangle \langle proof \rangle \langle proof \rangle \langle proof \rangle \langle proof \rangle$

The scheme in full, as for $DInf$: assembled by definitional unfolding, and in its locally-nameless normal form:

lemma $ACrel-unfolded$:

$(ACrel\ \sigma\ \tau :: 'p\ tm)$

$= (\prod rR_\sigma \Rightarrow \tau \Rightarrow o.$

$((\prod rX_\sigma. \exists rY_\tau. (((rR^f\sigma \Rightarrow \tau \Rightarrow o) \cdot (rX^f\sigma)) \cdot (rY^f\tau)))$

$\supset (\exists rF_\sigma \Rightarrow \tau. \prod rX_\sigma.$

$((((rR^f\sigma \Rightarrow \tau \Rightarrow o) \cdot (rX^f\sigma)) \cdot ((rF^f\sigma \Rightarrow \tau) \cdot (rX^f\sigma))))))$

$\langle proof \rangle$

lemma $ACrel-norm$:

$(ACrel\ \sigma\ \tau :: 'p\ tm)$

$= \prod \sigma \Rightarrow \tau \Rightarrow o$

$((\prod \sigma (\exists \tau ((Bnd\ (Suc\ (Suc\ 0)) \cdot Bnd\ (Suc\ 0)) \cdot Bnd\ 0)))$

$\supset (\exists \sigma \Rightarrow \tau (\prod \sigma ((Bnd\ (Suc\ (Suc\ 0)) \cdot Bnd\ 0) \cdot (Bnd\ (Suc\ 0) \cdot Bnd\ 0))))$

$\langle proof \rangle$

3.2 Satisfaction of the choice scheme in an arbitrary standard model

Every instance of the scheme holds in *any* Σ -standard model: the full function space supplies the choice function as the λ -abstraction of a meta-level Hilbert choice. (The corresponding satisfaction analysis for $DInf$ — $DInf$ -sat and its relatives — lives in the main entry's *NK-Infinity* and is used below unchanged.)

context *standard-model*

begin

lemma $ACrel-sat$:

assumes xi : $bkkA.asg\ \xi$

shows $den\ (ACrel\ \sigma\ \tau)\ \xi = Tv$

$\langle proof \rangle$

end

The relative-consistency theorem, abstractly. Suppose *some standard* Σ -model has an injective but not surjective self-map of its individuals — that is, Dedekind-infinitely many individuals. Then full classical *HOL* is consistent. Consistency is the predicate *con* of the main entry, defined by $con\ \Phi \longleftrightarrow \neg (\Phi \vdash \perp)$: the conclusion below states that *NK* cannot derive falsity from $DInf$ together with all instances of $ACrel$. The eigen-parameter starvation that makes *con* a weak reading for impure infinite contexts (main entry, *NK-Infinity*) cannot arise here: the axioms use no parameters at all, so the context is pure, and over an infinite signature $con\ \Phi$ and $\neg (\Phi$

$\models \perp$) provably coincide (*fprov-eq-bprov*). The proof works for an *arbitrary* such model, whatever type its elements come from, and uses no set theory. The one thing plain HOL cannot supply is the premise — a model with infinitely many individuals; producing one is the central task of the set-theoretic theory that concludes this entry.¹

theorem *con-full-HOL-rel*:

fixes $Dm :: ty \Rightarrow 'u \Rightarrow bool$ **and** $Ap :: 'u \Rightarrow 'u \Rightarrow 'u$
and $Lm :: ty \Rightarrow ('u \Rightarrow 'u) \Rightarrow 'u$ **and** $Tv Fv Ngv Dsv :: 'u$
and $Iv Ev Piv :: ty \Rightarrow 'u$ **and** $Jv :: 'p \Rightarrow ty \Rightarrow 'u$
assumes M : *standard-model* $Dm Ap Lm Tv Fv Ngv Dsv Iv Ev Piv Jv$
and *inf*: $\exists d. Dm (\iota \Rightarrow \iota) d$
 $\wedge (\forall a. Dm \iota a \longrightarrow (\forall b. Dm \iota b \longrightarrow (Ap d a = Ap d b \longrightarrow a = b)))$
 $\wedge (\exists c. Dm \iota c \wedge (\forall z. Dm \iota z \longrightarrow Ap d z \neq c))$
shows *con* (*insert* (*DInf* :: '*p tm*) (*range* ($\lambda(\sigma, \tau). ACrel \sigma \tau$)))

<proof>

Completeness carries over to the axiomatic extension, in its strongest form: $\{DInf\}$ is a finite context, so *completeness-hyps-open-finite* applies — free variables are allowed in the conclusion, the carrier and the signature are any infinite types, and purity is automatic. Any formula true in every infinite-carrier model of *DInf* is thus already derivable from *DInf* in *NK*, and with soundness the two coincide — established by the plain-HOL completeness of the main entry, with no set universe.

theorem *NK-completeness-DInf*:

fixes $A :: 'p::infinite\ tm$
assumes $wff_o(A)$ **and** $\{DInf\} \models ('u::infinite) A$
shows $\{DInf\} \vdash A$

<proof>

theorem *sound-and-complete-DInf*:

fixes $A :: 'p::infinite\ tm$
assumes $wff_o(A)$
shows $\{DInf\} \vdash A \longleftrightarrow \{DInf\} \models ('u::infinite) A$
<proof>

¹Why not? Plain HOL has infinite *types*, and every single level *nat*, $nat \Rightarrow bool$, ... of the full tower exists as a type. What it lacks is *one carrier* for the whole tower at once. In BKK's sense "standard" means only that the function domains are *full*; but a model formalised inside plain HOL necessarily has one further feature: the object types are values of the datatype *ty*, so any formalised model — the locale *standard-model* included — interprets all domains inside a single carrier type '*u*' (simple types cannot assign each object type its own meta-type). Over infinite individuals with full function spaces that one carrier must outgrow every level of the tower. No plain-HOL type reaches that far: every type plain HOL can construct is bounded in cardinality by some finite tower of function spaces over its one primitive infinite type *ind* (of which *nat* is the familiar copy). One might hope to escape with a *Henkin* model, whose function domains need not be full and which therefore fits into a small carrier. But a sparse function domain need not provide what the axiom asks for. *DInf* says: *there exists* an injective, non-surjective self-map of the individuals. In a Henkin model that map has to be an element of the (now sparse) function domain $D\iota \Rightarrow \iota$ — and it may simply not be there, even when the individuals are infinite. This is not a hypothetical: the main entry's *henkin-scheme-refutes-DInf* constructs a Henkin model with infinitely many individuals in which *DInf* is nevertheless *false*. And the canonical way to build a Henkin model that does contain the witness — the term-model construction — takes the consistency of *DInf* as its *premise*, and so cannot be the way to prove it (that *every* other route fails as well is Gödel's second theorem, see the introduction).

4 Minimising the meta-theory: a weaker universe suffices

We weaken the premise of *con-full-HOL-rel* once more. Instead of *assuming* the logical constants (as *standard-model* does), the locale *hol-universe* adds Dedekind-infinite individuals to the constants-construction interface *lambda-universe* of the main entry: a universe with full function spaces, booleans, extensionality and a domain-respecting parameter interpretation, over which the constants are *constructed*. Any model of these weaker assumptions — the final theory provides one over V — thus yields the relative consistency of full *HOL*.

```

locale hol-universe = lambda-universe Dm Ap Lm Tv Fv Jv
  for Dm :: ty  $\Rightarrow$  'u  $\Rightarrow$  bool and Ap :: 'u  $\Rightarrow$  'u  $\Rightarrow$  'u
    and Lm :: ty  $\Rightarrow$  ('u  $\Rightarrow$  'u)  $\Rightarrow$  'u and Tv Fv :: 'u and Jv :: 'p  $\Rightarrow$  ty  $\Rightarrow$  'u +
  assumes ind-inf:  $\exists d. Dm (\iota \Rightarrow \iota) d$ 
     $\wedge (\forall a. Dm \iota a \longrightarrow (\forall b. Dm \iota b \longrightarrow (Ap d a = Ap d b \longrightarrow a = b)))$ 
     $\wedge (\exists c. Dm \iota c \wedge (\forall z. Dm \iota z \longrightarrow Ap d z \neq c))$ 
begin

```

```

theorem con-full-HOL-universe:
  con (insert (DInf :: 'p tm) (range ( $\lambda(\sigma, \tau). ACrel \sigma \tau$ )))
  <proof>

```

```

end

```

5 Minimising further: a Harrison-style universe axiom

Harrison's self-verification of HOL Light [4] proves the soundness of HOL-with-infinity in a meta-logic strengthened by a single *universe axiom*: an infinite type closed under the set-forming operations that full function spaces require. The locale *harrison-universe* is our rendering of that axiom, one level below *hol-universe*: it mentions neither types, domains, application nor abstraction. Fixed are an injective pairing *pr*, an injective coding *enc* of the *small* collections *Sm*, and a small Dedekind-infinite collection *D0* of individuals; smallness is closed downwards, under products and — the strong-limit core of Harrison's axiom — under *coded power collections* (*Sm-pow*).² From these data alone the whole *hol-universe* structure is *constructed*: domains as collections of coded functional graphs, abstraction by coding, application by decoding.

```

locale harrison-universe =
  fixes pr :: 'u  $\Rightarrow$  'u  $\Rightarrow$  'u and enc :: ('u  $\Rightarrow$  bool)  $\Rightarrow$  'u
    and Sm :: ('u  $\Rightarrow$  bool)  $\Rightarrow$  bool and D0 :: 'u  $\Rightarrow$  bool
  assumes pr-inj:  $pr a b = pr c d \implies a = c \wedge b = d$ 
    and enc-inj:  $\llbracket Sm X; Sm Y; enc X = enc Y \rrbracket \implies X = Y$ 
    and Sm-sub:  $\llbracket Sm X; Y \leq X \rrbracket \implies Sm Y$ 
    and Sm-prod:  $\llbracket Sm X; Sm Y \rrbracket \implies Sm (\lambda z. \exists a b. X a \wedge Y b \wedge z = pr a b)$ 
    and Sm-pow:  $Sm X \implies Sm (\lambda z. \exists Y. Y \leq X \wedge Sm Y \wedge z = enc Y)$ 
    and Sm-D0:  $Sm D0$ 
    and D0-inf:  $\exists f c. (\forall a. D0 a \longrightarrow D0 (f a))$ 
       $\wedge (\forall a b. D0 a \longrightarrow D0 b \longrightarrow f a = f b \longrightarrow a = b)$ 
       $\wedge D0 c \wedge (\forall a. D0 a \longrightarrow f a \neq c)$ 

```

²In effect this axiom postulates a fresh base type v that gathers the domains of the whole type hierarchy over ι and o , so that every finite-type domain embeds into it. The subtlety is that v is closed only under the operations those finite types require — pairing and *small* (coded) power collections — and *not* under its own full function space $v \Rightarrow o$, which by Cantor would be strictly larger. This restricted closure, rather than a full powerset, is exactly what keeps the axiom consistent and of merely strong-limit strength.

begin

Smallness of the empty collection need not be assumed: it is a sub-collection of the small $D0$.

lemma $Sm\text{-}empty$: $Sm (\lambda\cdot. False)$
 $\langle proof \rangle$

5.1 Truth values, domains, abstraction, application

The truth values are the first stages of the cumulative hierarchy in miniature: HTv codes the empty collection, HFv its singleton. Both collections are small — the empty one by $Sm\text{-}empty$, its singleton as the coded power collection of the empty one ($Sm\text{-}pow$) — and $enc\text{-}inj$ keeps the two codes apart. A function domain consists of the codes of the functional graphs over the argument domain; application decodes, abstraction codes.

definition $HTv :: 'u \text{ where } HTv = enc (\lambda\cdot. False)$

definition $HFv :: 'u \text{ where } HFv = enc (\lambda z. z = HTv)$

primrec $HD :: ty \Rightarrow 'u \Rightarrow bool \text{ where}$
 $HD \iota = D0$
 $| HD \circ = (\lambda x. x = HTv \vee x = HFv)$
 $| HD (\sigma \Rightarrow \tau) = (\lambda x. \exists g. (\forall a. HD \sigma a \longrightarrow HD \tau (g a))$
 $\quad \wedge x = enc (\lambda z. \exists a. HD \sigma a \wedge z = pr a (g a)))$

definition $Hgr :: ty \Rightarrow ('u \Rightarrow 'u) \Rightarrow 'u \Rightarrow bool \text{ where}$

$Hgr \sigma g = (\lambda z. \exists a. HD \sigma a \wedge z = pr a (g a))$

definition $HLm :: ty \Rightarrow ('u \Rightarrow 'u) \Rightarrow 'u \text{ where } HLm \sigma g = enc (Hgr \sigma g)$

The THE below is Isabelle's *definite description* at the meta-level — unique choice, not a choice principle; decoding is well-defined by $enc\text{-}inj$. (The $SOME$ in HJv is meta-level Hilbert choice of Isabelle/HOL. Neither adds an axiom to the embedded object logic.)

definition $Hdec :: 'u \Rightarrow 'u \Rightarrow bool \text{ where } Hdec x = (THE X. Sm X \wedge enc X = x)$

definition $HAp :: 'u \Rightarrow 'u \Rightarrow 'u \text{ where } HAp f a = (THE b. Hdec f (pr a b))$

definition $HJv :: 'p \Rightarrow ty \Rightarrow 'u \text{ where } HJv p \sigma = (SOME d. HD \sigma d)$

lemma $HD\text{-}Fun$: $HD (\sigma \Rightarrow \tau) x \longleftrightarrow (\exists g. (\forall a. HD \sigma a \longrightarrow HD \tau (g a)) \wedge x = HLm \sigma g)$
 $\langle proof \rangle$

5.2 Every domain is small and inhabited

lemma $Sm\text{-}singleton\text{-}HTv$: $Sm (\lambda z. z = HTv)$
 $\langle proof \rangle$

lemma $HTv\text{-}neq\text{-}HFv$: $HTv \neq HFv$
 $\langle proof \rangle$

lemma $Sm\text{-}bool$: $Sm (HD \circ)$
 $\langle proof \rangle$

The function-space case is the Harrison core: every coded graph is a small subcollection of the product of the two domains, so the domain of codes sits inside the coded power collection of that product.

lemma $Sm\text{-}HD$: $Sm (HD \sigma)$
 $\langle proof \rangle$

lemma *HLm-dom*: **assumes** $\bigwedge d. HD \sigma d \implies HD \tau (h d)$ **shows** $HD (\sigma \Rightarrow \tau) (HLm \sigma h)$
 $\langle proof \rangle$

lemma *HD-ne*: $\exists x. HD \sigma x$
 $\langle proof \rangle$

5.3 Decoding is inverse to coding, and the frame conditions follow

lemma *Hdec-enc*: **assumes** $Sm X$ **shows** $Hdec (enc X) = X$
 $\langle proof \rangle$

lemma *Hgr-pr*: **assumes** $HD \sigma a$ **shows** $Hgr \sigma h (pr a b) \longleftrightarrow b = h a$
 $\langle proof \rangle$

lemma *Sm-Hgr*: **assumes** $\bigwedge d. HD \sigma d \implies HD \tau (h d)$ **shows** $Sm (Hgr \sigma h)$
 $\langle proof \rangle$

lemma *HAp-HLm*: **assumes** $sm: Sm (Hgr \sigma h)$ **and** $a: HD \sigma a$ **shows** $HAp (HLm \sigma h) a = h a$
 $\langle proof \rangle$

lemma *HD-FunE*:
assumes $f: HD (\sigma \Rightarrow \tau) f$
shows $\exists g. (\forall a. HD \sigma a \longrightarrow HD \tau (g a)) \wedge f = HLm \sigma g \wedge (\forall a. HD \sigma a \longrightarrow HAp f a = g a)$
 $\langle proof \rangle$

lemma *HAp-dom*: **assumes** $f: HD (\sigma \Rightarrow \tau) f$ **and** $a: HD \sigma a$ **shows** $HD \tau (HAp f a)$
 $\langle proof \rangle$

lemma *HAp-ext*:
assumes $g: HD (\sigma \Rightarrow \tau) g$ **and** $k: HD (\sigma \Rightarrow \tau) k$
and $e: \bigwedge a. HD \sigma a \implies HAp g a = HAp k a$
shows $g = k$
 $\langle proof \rangle$

lemma *HJv-dom*: $HD \sigma (HJv p \sigma)$
 $\langle proof \rangle$

Dedekind infinity transfers from the meta-level self-map f of $D0$ to the object level: the witness in $HD (\iota \Rightarrow \iota)$ is its coded graph $HLm \iota f$.

lemma *HD-inf*:
 $\exists d. HD (\iota \Rightarrow \iota) d$
 $\wedge (\forall a. HD \iota a \longrightarrow (\forall b. HD \iota b \longrightarrow (HAp d a = HAp d b \longrightarrow a = b)))$
 $\wedge (\exists c. HD \iota c \wedge (\forall z. HD \iota z \longrightarrow HAp d z \neq c))$
 $\langle proof \rangle$

5.4 Every Harrison universe is a hol-universe

theorem *harrison-hol-universe*: $hol-universe HD HAp HLm HTv HFv (HJv :: 'p \Rightarrow ty \Rightarrow 'u)$
 $\langle proof \rangle$

lemmas *harrison-standard-model* =
 $lambda-universe.is-standard-model[OF hol-universe.axioms(1)][OF harrison-hol-universe]]$

theorem *con-full-HOL-harrison:*

con (insert (DInf :: 'p tm) (range (λ(σ,τ). ACrel σ τ)))
<proof>

end

6 The set-theoretic instantiation: V as a *hol-universe*

This theory supplies the one ingredient the abstract half *HOL-Universe* leaves open: a type satisfying the *hol-universe* assumptions. Over Paulson's `ZFC_in_HOL` the individuals are interpreted as the finite ordinals ω and every function space is full, so the successor map witnesses Dedekind infinity and all assumptions of the locale are discharged. The consistency of full classical *HOL* and of *NK* with *DInf* then fall out of the abstract theorems; V also validates the at most as strong Harrison-style interface *harrison-universe* outright — and that most abstract interface carries the exported consistency theorem; a rank bound closes the development, showing every domain of the witnessing model lives inside the set $Vset (\omega + \omega)$. Only this theory imports *ZFC-in-HOL*.

6.1 Domains, application, abstraction

primrec $D :: ty \Rightarrow V$ **where**

$D \iota = \omega$
 $| D \circ = set \{0, 1\}$
 $| D (\sigma \Rightarrow \tau) = VPi (D \sigma) (\lambda \cdot. D \tau)$

definition $VD :: ty \Rightarrow V \Rightarrow bool$ **where** $VD \sigma v \equiv v \in elts (D \sigma)$

definition $VLM :: ty \Rightarrow (V \Rightarrow V) \Rightarrow V$ **where** $VLM \sigma h \equiv VLambda (D \sigma) h$

lemma *VD-bool*: $VD \sigma v \longleftrightarrow v = 0 \vee v = 1$ *<proof>*

lemma *VD-nonempty*: $\exists v. VD \sigma v$
<proof>

lemma *D-nonempty-elts*: $\exists v. v \in elts (D \sigma)$
<proof>

lemma *VLM-beta*: $VD \sigma a \Longrightarrow app (VLM \sigma h) a = h a$ *<proof>*

lemma *VLM-dom*: $(\bigwedge d. VD \sigma d \Longrightarrow VD \tau (h d)) \Longrightarrow VD (\sigma \Rightarrow \tau) (VLM \sigma h)$
<proof>

lemma *VAp-dom*: $VD (\sigma \Rightarrow \tau) f \Longrightarrow VD \sigma a \Longrightarrow VD \tau (app f a)$
<proof>

lemma *VAp-ext*:

$VD (\sigma \Rightarrow \tau) f \Longrightarrow VD (\sigma \Rightarrow \tau) g \Longrightarrow (\bigwedge a. VD \sigma a \Longrightarrow app f a = app g a) \Longrightarrow f = g$
<proof>

definition $VJv :: 'p \Rightarrow ty \Rightarrow V$ **where** $VJv p \sigma = (SOME v. v \in elts (D \sigma))$

lemma *VJv-dom*: $VD \sigma (VJv p \sigma)$
<proof>

lemma *VJv-param-independent*: $VJv p \sigma = VJv q \sigma$

$\langle \text{proof} \rangle$

The V -frame discharges the Dedekind-infinity assumption *ind-inf* of *hol-universe* (verbatim the frame premise of *con-full-HOL-rel*): the successor on ω is an injective, non-surjective self-map of the individuals.

lemma *V-inf*:

$\exists d. VD (\iota \Rightarrow \iota) d$
 $\wedge (\forall a. VD \iota a \longrightarrow (\forall b. VD \iota b \longrightarrow (app\ d\ a = app\ d\ b \longrightarrow a = b)))$
 $\wedge (\exists c. VD \iota c \wedge (\forall z. VD \iota z \longrightarrow app\ d\ z \neq c))$

$\langle \text{proof} \rangle$

V is such a universe: the *hol-universe* assumptions are satisfiable (relative to ZF) — yet they demand far less than full ZF .

theorem *V-hol-universe*:

hol-universe $VD\ app\ V\ Lm\ (1::V)\ 0\ (VJv :: 'p \Rightarrow ty \Rightarrow V)$

$\langle \text{proof} \rangle$

6.2 The standard model, by instantiation

The constructed constants of the locale turn the V -frame into a standard model outright — no hand-built logical constants over V are needed. This natural model carries the Dedekind witness and the rank bound below; the consistency *theorem* itself is exported once, from the Harrison-style interface in the next subsection.

lemmas *V-standard-model* =

lambda-universe.is-standard-model[*OF hol-universe.axioms*(1)][*OF V-hol-universe*]

6.3 V validates the Harrison-style axiom, and consistency follows

The at most as strong interface *harrison-universe* — our rendering of Harrison’s universe axiom — is likewise validated by V directly: pairing is *vpair*, a collection is small in the sense of the locale iff it is *small*, its code is Paulson’s *set*, and the individuals are again the finite ordinals. The coded-power closure is exactly *VPow* seen through the coding. The locale rebuilds domains, application and abstraction from these four constants alone, so the headline consistency theorem rests on nothing beyond them — one spine, through the most abstract interface.

lemma *V-enc-inj*:

assumes *small* $\{x. X\ x\}$ **and** *small* $\{x. Y\ x\}$ **and** *set* $\{x. X\ x\} = \text{set } \{x. Y\ x\}$
shows $X = Y$

$\langle \text{proof} \rangle$

lemma *V-small-sub*: **assumes** *small* $\{x. X\ x\}$ **and** $Y \leq X$ **shows** *small* $\{x. Y\ x\}$

$\langle \text{proof} \rangle$

lemma *V-small-prod*:

assumes *small* $\{x. X\ x\}$ **and** *small* $\{x. Y\ x\}$
shows *small* $\{z. \exists a\ b. X\ a \wedge Y\ b \wedge z = \text{vpair } a\ b\}$

$\langle \text{proof} \rangle$

lemma *V-small-pow*:

assumes *sX*: *small* $\{x. X\ x\}$
shows *small* $\{z. \exists Y. Y \leq X \wedge \text{small } \{x. Y\ x\} \wedge z = \text{set } \{x. Y\ x\}\}$

$\langle \text{proof} \rangle$

lemma *V-omega-inf*:

$\exists f c. (\forall a. a \in \text{elts } \omega \longrightarrow f a \in \text{elts } \omega)$
 $\wedge (\forall a b. a \in \text{elts } \omega \longrightarrow b \in \text{elts } \omega \longrightarrow f a = f b \longrightarrow a = b)$
 $\wedge c \in \text{elts } \omega \wedge (\forall a. a \in \text{elts } \omega \longrightarrow f a \neq c)$
 $\langle \text{proof} \rangle$

theorem *V-harrison-universe*:

harrison-universe vpair $(\lambda X. \text{set } \{x. X x\}) (\lambda X. \text{small } \{x. X x\}) (\lambda x. x \in \text{elts } \omega)$
 $\langle \text{proof} \rangle$

theorem *con-full-HOL*:

con $(\text{insert } (DInf :: 'p \text{ tm}) (\text{range } (\lambda(\sigma, \tau). ACrel \sigma \tau)))$
 $\langle \text{proof} \rangle$

6.4 The Dedekind axiom of infinity and a countable Henkin witness

The infinity-only consistency *con* $\{DInf\}$ is the same *V*-model argument, now feeding the successor witness on ω to the carrier-generic *DInf-sat*; it holds for *every* parameter type *'p*. The *countable* model that witnesses it is the ZFC-free term model of *countable-henkin-sat*, so a set-theoretic universe enters *only* through the consistency premise — the model itself has a countable total domain, carved out of the carrier *'p tm set*, and never mentions *V*.

theorem *con-DInf*: *con* $(\{DInf\} :: 'p \text{ tm set})$

$\langle \text{proof} \rangle$

theorem *countable-henkin-model-DInf*:

obtains *Dm Ap*
and *Ee* :: $(\text{nat} \Rightarrow \text{ty} \Rightarrow 'p::\{\text{countable}, \text{infinite}\} \text{ tm set}) \Rightarrow 'p \text{ tm} \Rightarrow 'p \text{ tm set}$
and *vl* ξ
where *bkk-model* *Dm Ap Ee vl countable* $\{x. \exists \tau. Dm \tau x\}$
app-struct.asg *Dm* $\xi \text{ vl } (Ee \xi (DInf :: 'p \text{ tm}))$
 $\langle \text{proof} \rangle$

The converse half of the separation of axiom and scheme. The main entry proves that the infinity *scheme* does not yield the axiom (*Diag-not-derives-DInf*, *henkin-scheme-refutes-DInf* in its *NK-Infinity*). The *V*-model settles the other direction: its parameter interpretation *VJv* is constant at each type (*VJv-param-independent*), so all diagram constants denote the same individual and every inequation *dneq f i j* is refuted, while *DInf* holds. By soundness the axiom derives no diagram inequation — as *sentences*, axiom and scheme are incomparable.

theorem *DInf-not-derives-dneq*:

fixes *f* :: $\text{nat} \Rightarrow 'p::\text{infinite}$
shows $\neg (\{DInf\} \vdash \text{dneq } f \text{ i } j)$
 $\langle \text{proof} \rangle$

6.5 The universe has bounded rank: no replacement, no proper class

The rank bound is now made precise: *every* domain *D* σ has rank below $\omega + \omega$ (that is, $\omega \cdot 2$), so every domain is an element of the set *Vset* $(\omega + \omega)$. The domain of individuals sits at rank ω , and each function-space layer *D* $(\sigma \Rightarrow \tau) = VPi (D \sigma) (\lambda-. D \tau)$ — contained in *VPow* (*VPow* (*VPow* (*D* $\sigma \sqcup D \tau$))) — raises the rank by only three, which the limit ordinal $\omega + \omega$ absorbs.

lemma *rank-D-iota*: *rank* (*D* ι) = ω $\langle \text{proof} \rangle$

lemma *rank-VPow-le*: $\text{rank } (\text{VPow } x) \leq \text{succ } (\text{rank } x)$
 $\langle \text{proof} \rangle$

lemma *sup-Sup-pair*: $(a :: V) \sqcup b = \bigsqcup \{a, b\}$
 $\langle \text{proof} \rangle$

lemma *rank-sup-eq*: $\text{rank } (a \sqcup b) = \text{rank } a \sqcup \text{rank } b$
 $\langle \text{proof} \rangle$

lemma *VPi-le-VPow*: $\text{VPi } A \ B \leq \text{VPow } (\text{VSigma } A \ B)$
 $\langle \text{proof} \rangle$

lemma *VSigma-const-le*: $\text{VSigma } A \ (\lambda \cdot. C) \leq \text{VPow } (\text{VPow } (A \sqcup C))$
 $\langle \text{proof} \rangle$

lemma *Limit-ww*: $\text{Limit } (\omega + \omega) \ \langle \text{proof} \rangle$

lemma *succ-lt-ww*:
assumes $\text{Ord } \beta$ **and** $\beta < \omega + \omega$ **shows** $\text{succ } \beta < \omega + \omega$
 $\langle \text{proof} \rangle$

lemma *omega-lt-ww*: $\omega < \omega + \omega$
 $\langle \text{proof} \rangle$

lemma *rank-VPi-const-lt*:
assumes $\text{rank } A < \omega + \omega$ **and** $\text{rank } C < \omega + \omega$
shows $\text{rank } (\text{VPi } A \ (\lambda \cdot. C)) < \omega + \omega$
 $\langle \text{proof} \rangle$

theorem *rank-D-lt*: $\text{rank } (D \ \sigma) < \omega + \omega$
 $\langle \text{proof} \rangle$

Hence $D \ \sigma \in \text{elts } (\text{Vset } (\omega + \omega))$ for every type: each domain of the witnessing model is an element of the set $\text{Vset } (\omega + \omega)$.

lemma *D-in-Vset-ww*: $D \ \sigma \in \text{elts } (\text{Vset } (\omega + \omega))$
 $\langle \text{proof} \rangle$

References

- [1] C. Benzmüller, C. E. Brown, and M. Kohlhase. Higher-order semantics and extensionality. *Journal of Symbolic Logic*, 69(4):1027–1088, 2004.
- [2] C. Benzmüller and D. Kirchner. A deep embedding of HOL in HOL: Soundness, completeness, consistency. *Archive of Formal Proofs*, 2026. https://isa-afp.org/entries/HOL_in_HOL_Deep.html, Formal proof development.
- [3] A. Church. A formulation of the simple theory of types. *Journal of Symbolic Logic*, 5(2):56–68, 1940.

- [4] J. Harrison. Towards self-verification of HOL Light. In *Automated Reasoning (IJCAR 2006)*, volume 4130 of *Lecture Notes in Computer Science*, pages 177–191. Springer, 2006.
- [5] R. Kumar, R. Arthan, M. O. Myreen, and S. Owens. Self-formalisation of higher-order logic: Semantics, soundness, and a verified implementation. *Journal of Automated Reasoning*, 56(3):221–259, 2016.
- [6] O. Kunčar and A. Popescu. A consistent foundation for Isabelle/HOL. *Journal of Automated Reasoning*, 62(4):531–555, 2019.
- [7] A. R. D. Mathias. The strength of Mac Lane set theory. *Annals of Pure and Applied Logic*, 110(1–3):107–234, 2001.
- [8] L. C. Paulson. Zermelo fraenkel set theory in higher-order logic. *Archive of Formal Proofs*, 2019. https://isa-afp.org/entries/ZFC_in_HOL.html, Formal proof development.