Weak double extension set theory is inconsistent

Peter Zhang
peterzhangsg@gmail.com

November 18, 2026

Abstract

We prove that Kisielewicz's weak double extension set theory DEST is inconsistent: already two of its axioms, mixed extensionality and comprehension, contradict each other. DEST has two membership relations, $\in$ and ε; it interprets ZF, and its consistency was an open problem.

Holmes showed, using the regularity axiom, that the ordinals of the two relations differ: with the relations suitably named, every ε-ordinal is an $\in$-ordinal, and the object $\mathrm{Ord} = \{x \mid x \text{ is an ordinal}\}$ is an $\in$-ordinal but not an ε-ordinal. We show first that comprehension alone refutes such an asymmetry, in intuitionistic logic. Assume this asymmetry. If V is the universal set and $\Theta(z)$ says that z omits some ordinal, then the object

$$A = \{\, z \mid z = V \ \vee\ \mathrm{ord}(z) \ \vee\ \big(V \in z \wedge (z \in z \oplus \Theta(z))\big) \,\}$$

satisfies $A \in A \leftrightarrow A \notin A$, and the same argument refutes every formula that is asymmetric in this way. We then derive the asymmetry of the ordinals, up to exchanging the two relations, from mixed extensionality and comprehension without the regularity axiom; this is the only use of classical logic. Both parts are formalized in Agda and in Lean 4; a variant of the proof that uses no classical logic is formalized as additional code. The code is at github.com/BedrockInstitute/DESTInconsistency.

1 The theory

Language. First-order logic with $=$ and two membership relations $\in$ and ε. For a side s (one of $\in$, ε) we write $x \in_s y$, let s^* be the other side, and call $F_s(a) = \{x : x \in_s a\}$ the *s-face* of a. An object a is *regular* if $F_\in(a) = F_\varepsilon(a)$.

Uniform formulas. A formula is uniform if its only membership symbol is $\in$. Its *s-reading* φ^s interprets that symbol as $\in_s$. Thus $\varphi^\in = \varphi$, and φ^ε is what Holmes calls the dual φ^* of φ; we use the same star for the other side.

ME (mixed extensionality). If $F_\in(x) = F_\varepsilon(y)$, then $x = y$.

COMP (comprehension). For every uniform formula $\varphi(x, \bar{y})$ and all regular $\bar{y}$ there is $A = \{x \mid \varphi(x, \bar{y})\}$ with

$$F_\varepsilon(A) = \{x : \varphi^\in(x, \bar{y})\}, \qquad F_\in(A) = \{x : \varphi^\varepsilon(x, \bar{y})\},$$

that is, $F_s(A) = \{x : \varphi^{s^*}(x, \bar{y})\}$ for both sides s: the s-face of A is given by the s^*-reading of φ. Comprehension is thus *crossed*. We write $\{x \mid \cdots\}$ for objects given by COMP and $\{x : \cdots\}$ for classes, such as faces, and we leave the parameters $\bar{y}$ implicit where they play no role.

REG (regularity). If one face of x is included in the same face of some y whose members on both sides are all regular, then x is regular.

DEST is ME + COMP + REG [H05, §2], [HFL, §8.1]. DEST$^-$ is ME + COMP.

On the formulation. In [H05], Axiom 2.8 places "there is $\{x \mid \varphi\}$" before the quantifier over the parameters. We read it, as Holmes does in his proofs, as "for all regular parameters there is A"; read literally, the axiom is already inconsistent. Kisielewicz [K98], weakening his theory of [K89], folds REG into COMP through a condition on parameters: they must be partially contained in a suitable regular object, where x is partially contained in y if $F_s(x) \subseteq F_s(y)$ for some side s.* Holmes [H04, §2] states that this is equivalent to the formulation above, which is the one recorded as open in [HFL, §8.1] and [SEP, §10].

2 Asymmetry and the results

Being an ordinal is expressed by a parameter-free uniform formula ord. Holmes's definition [H05, §5] has the clause $\beta \in S$, where the parameter S is $\{x \mid x \text{ has a singleton}\}$; replacing that clause by "β has a singleton" gives such a formula, and Section 4 gives a variant. Using REG, Holmes showed that, with the relations suitably named, every ε-ordinal is an $\in$-ordinal, and that the object $\mathrm{Ord} = \{x \mid \mathrm{ord}(x)\}$ is an $\in$-ordinal but not an ε-ordinal [H05, Theorem 5.1 and the discussion after it]. So Ord is an ordinal on one side only. We isolate what is needed.

Definition. [Agda · Lean] Let $P(x)$ be a uniform formula with regular parameters, and let b, c be the two sides. P is *asymmetric from c to b* if

(i) $P^c(x)$ implies $P^b(x)$;
(ii) if $P^b(x)$, then no b-member y of x has $y \in_b y$;
(iii) some W with $F_b(W) = \{x : P^c(x)\}$ and $F_c(W) = \{x : P^b(x)\}$ has $P^b(W)$.

By COMP, objects W as in (iii) exist; (iii) asks that one of them satisfy P^b. With this naming, Holmes's ordinal formula is asymmetric from ε to $\in$: (i) and (iii) are his two facts, and (ii) holds because no ordinal has a self-membered member.

Theorem 1 (Diagonal)**.** [Agda · Lean] *Under* COMP*, no formula is asymmetric. The proof uses neither* ME *nor classical logic.*

Theorem 2. [Agda · Lean] DEST^- *is inconsistent in classical logic. Hence* DEST *is inconsistent.*

Theorem 1 and the asymmetry of the ordinals together give a contradiction. The asymmetry can be proved in two ways, and either one suffices. First, Holmes's two facts, proved with REG, provide it, so Theorem 1 already refutes full DEST. Second, the proof of Theorem 2 does without REG: ME and COMP alone make the ordinals asymmetric from one side to the other, and Theorem 1 gives the contradiction. A variant of the proof, with a different ordinal formula, uses no classical logic; it is given only as code, in `agda/Intuitionistic.agda` and `lean/Intuitionistic.lean`, and Section 6 says what passing from it to first-order intuitionistic logic uses. The rest of the paper is arranged as follows. Section 3 proves Theorem 1; Section 4 collects facts about regular objects and ordinals; Section 5 derives from them the asymmetry of the ordinals and completes the proof of Theorem 2.

* If the condition asks for a regular object with regular members, it matters which parameters the proofs use. The proof of Theorem 2 uses only parameters that are regular objects with regular members, with two exceptions. One is the universal set V, which can be eliminated: replace $z = V$ by $\forall y\,(y \in z)$ and $V \in z$ by $\exists v\,(\forall y\,(y \in v) \wedge v \in z)$; by ME, V is the only object whose s-face is everything. The other is the object h in the proof of Lemma 4, which is regular but need not have regular members. Replacing "has a singleton" in (O4) and in the object D of Proposition 2 by "has a singleton that has a singleton" removes this exception too: the proof of Lemma 4 then uses as parameter a singleton of h, which is regular and has the regular object h as its only member. So under this reading, too, the theory is inconsistent in classical logic (`kisielewicz-strict` in `agda/Supplement.agda`).

3 Comprehension refutes asymmetry

The last step of the proof of Theorem 1 is propositional. We state it as a lemma, since the proof of Lemma 3 uses it too.

Lemma 1 (Exclusive or)**.** [Agda · Lean] *Let $\varphi(z)$ be a uniform formula and A an object such that, for both sides s,*

$$A \in_s A \leftrightarrow \big(A \in_{s^*} A \oplus \varphi^{s^*}(A)\big),$$

where $\oplus$ is exclusive or: $z \in z \oplus \varphi(z)$ stands for $(z \in z \wedge \neg\varphi(z)) \vee (z \notin z \wedge \varphi(z))$, and the right side is its s^*-reading at A. Then there is no side s with $\varphi^s(A)$ and $\neg\varphi^{s^*}(A)$.

Proof. Suppose $\varphi^s(A)$ and $\neg\varphi^{s^*}(A)$. Then the equivalences for s and for s^* become $A \in_s A \leftrightarrow A \in_{s^*} A$ and $A \in_{s^*} A \leftrightarrow A \notin_s A$. If $A \in_s A$, the two give $A \notin_s A$; so $A \notin_s A$. Then $A \in_{s^*} A$, and then $A \in_s A$, a contradiction. □

Proof of Theorem 1. Let P be asymmetric from c to b, with W as in (iii). Let $V = \{x \mid x = x\}$. Both faces of V are everything, so V is regular and $V \in_s V$. Let $\Theta(z)$ be $\exists u\,(P(u) \wedge u \notin z)$, and let

$$A = \{\, z \mid z = V \;\vee\; P(z) \;\vee\; \big(V \in z \wedge (z \in z \oplus \Theta(z))\big) \,\},$$

with parameters V and those of P. So $z \in_b A$ iff the c-reading of the condition holds at z, and $z \in_c A$ iff the b-reading does.

1. $P^b(V)$ fails by (ii), since $V \in_b V \in_b V$. $P^c(W)$ fails: otherwise $W \in_b W$ by (iii), against (ii) for W.
2. **$\Theta^b(A)$ holds**, because $W \notin_b A$. Each disjunct fails in the c-reading at W: $W \neq V$, since $P^b(W)$ and not $P^b(V)$; $P^c(W)$ fails; and $V \in_c W$ would give $P^b(V)$ by (iii).
3. **$\Theta^c(A)$ fails.** If $P^c(u)$, then $P^b(u)$ by (i), so $u \in_c A$.
4. **At $z = A$ only the last disjunct can hold.** $V \in_b A$ and $V \in_c A$ by the first disjunct. $A \neq V$, since $W \in_b V$ and $W \notin_b A$. $P^b(A)$ fails by (ii), since $V \in_b A$ and $V \in_b V$; so $P^c(A)$ fails by (i).
5. Hence $A \in_s A \leftrightarrow (A \in_{s^*} A \oplus \Theta^{s^*}(A))$ for both sides s. By steps 2 and 3, $\Theta^b(A)$ holds and $\Theta^c(A)$ fails, against Lemma 1 with $\varphi = \Theta$.

Neither this proof nor that of Lemma 1 uses ME or classical logic. □

Remark (Why the diagonal works). Russell's instance $R = \{x \mid x \notin x\}$ gives

$$R \,\varepsilon\, R \leftrightarrow R \notin R, \qquad R \in R \leftrightarrow \neg(R \,\varepsilon\, R).$$

Together they give $R \in R \leftrightarrow \neg\neg(R \in R)$, which is no contradiction: both equivalences hold when exactly one of $R \in R$ and $R \,\varepsilon\, R$ is true, so R is merely irregular. For the diagonal object A, $\Theta^b(A)$ holds and $\Theta^c(A)$ fails, so the two equivalences in the proof of Lemma 1 contain a single negation and give $A \in_b A \leftrightarrow A \notin_b A$.

4 Regular objects and ordinals

This section and the next prove Theorem 2; they are the only places that use ME and classical logic. Three facts about regular objects come first.

Lemma 2 (Rigidity)**.** [Agda · Lean] *If g is regular and $F_s(z) = F_s(g)$, then $z = g$.*

Proof. $F_s(z) = F_s(g) = F_{s^*}(g)$, so $z = g$ by ME. □

Lemma 3 (Flip [H05, Lemma 2.11])**.** [Agda · Lean] *Let σ be a uniform sentence with regular parameters. Then $\sigma^{\in} \leftrightarrow \sigma^{\varepsilon}$. In particular, the two readings of a uniform formula with regular parameters agree at every regular object.*

Proof. Let $A = \{x \mid x \in x \oplus \sigma\}$. By COMP, $A \in_s A \leftrightarrow (A \in_{s^*} A \oplus \sigma^{s^*})$ for both sides s. By Lemma 1 with $\varphi = \sigma$, there is no side s with σ^s and $\neg\sigma^{s^*}$; hence, by classical logic, $\sigma^{\in} \leftrightarrow \sigma^{\varepsilon}$. □

Lemma 4 (Singletons [H05, Lemma 2.10])**.** [Agda · Lean] *Call h an s-singleton of y if $F_s(h) = \{y\}$. Every regular object has singletons on both sides, and every object with singletons on both sides is regular.*

Proof. For the first claim, take $\{x \mid x = r\}$ for regular r. For the second, let h be an s-singleton of y and h' an s^*-singleton of y. ME gives $h = h'$ (in one order or the other), so both faces of h are $\{y\}$ and h is regular. Let $k = \{x \mid \exists u\,(x \in u \wedge u \in h)\}$. Then $F_{s^*}(k) = F_s(y)$, so $y = k$ by ME, and therefore $F_{s^*}(y) = F_{s^*}(k) = F_s(y)$. □

Definition. [Agda · Lean] $\mathrm{ord}(a)$ is the uniform formula, without parameters, that is the conjunction of:

(O1) $z \in y \wedge y \in a \to z \in a$;

(O2) every member of a satisfies (O1);

(O3) no member of a is a member of itself;

(O4) every member y of a has a singleton, i.e. $\exists h\,\forall v\,(v \in h \leftrightarrow v = y)$;

(O5) for members y, z of a: $y \in z \vee y = z \vee z \in y$;

(O6) for every C and every $x \in a$ with $x \notin C$, there is z with $z = x$ or $z \in x$, such that $z \notin C$ and every member of z is in C.

An *s-ordinal* is an object satisfying ord^s. In ZF this defines the von Neumann ordinals.

Lemma 5 (Ordinals)**.** [Agda · Lean]

(1) The s-members of an s-ordinal are s-ordinals.

(2) No s-ordinal is an s-member of itself.

(3) If x is an s-ordinal and $x \notin_s C$, there is z with $z = x$ or $z \in_s x$, such that $z \notin_s C$ and $F_s(z) \subseteq F_s(C)$.

(4) A regular s-ordinal is an s^*-ordinal.

Proof. (1) Let y be a member of a. By (O1) of a, the members of y are members of a; so (O1) for y is (O2) of a, and (O2)–(O6) for y follow from those of a. (2) By (O3). (3) If $F_s(x) \subseteq F_s(C)$, take $z = x$. Otherwise apply (O6) of x to a member of x that is not in C, and use (O1) of x. (4) By Lemma 3. □

Lemma 6 (Initial segments)**.** [Agda · Lean] *Let α be an s-ordinal and γ a regular object with $F_s(\gamma) \subseteq F_s(\alpha)$ such that $z \in_s y \in_s \gamma$ implies $z \in_s \gamma$. Then $\gamma = \alpha$ or $\gamma \in_s \alpha$.*

Proof. If $F_s(\alpha) \subseteq F_s(\gamma)$, then $\gamma = \alpha$ by Lemma 2. Otherwise pick $x \in_s \alpha$ with $x \notin_s \gamma$, and apply (O6) of α with $C = \gamma$. This gives $z \in_s \alpha$ with $z \notin_s \gamma$ and $F_s(z) \subseteq F_s(\gamma)$. Conversely, let $y \in_s \gamma$. Then $y \in_s \alpha$, so by (O5) $y \in_s z$, or $y = z$, or $z \in_s y$; the last two would put z in γ. So $F_s(z) = F_s(\gamma)$, hence $z = \gamma$ by Lemma 2, and $\gamma \in_s \alpha$. □

Lemma 7 (Comparability)**.** [Agda · Lean] *Regular s-ordinals α, β satisfy $\alpha \in_s \beta$ or $\alpha = \beta$ or $\beta \in_s \alpha$.*

Proof. Let $\gamma = \{x \mid x \in \alpha \wedge x \in \beta\}$. Its s-face and s^*-face are $F_{s^*}(\alpha) \cap F_{s^*}(\beta)$ and $F_s(\alpha) \cap F_s(\beta)$, which are equal because α and β are regular. So γ is regular, and it satisfies the condition of Lemma 6 by (O1) of α and β. By Lemma 6, $\gamma = \alpha$ or $\gamma \in_s \alpha$, and $\gamma = \beta$ or $\gamma \in_s \beta$. If $\gamma \in_s \alpha$ and $\gamma \in_s \beta$, then $\gamma \in_s \gamma$, against (O3) for α. The three remaining combinations give the claim. □

5 The ordinals are asymmetric

Proposition 1 (Regular ordinals on one side)**.** [Agda · Lean] *If every c-ordinal is regular, then* ord *is asymmetric from c to c^*.*

Proof. Let $b = c^*$.

(i) A c-ordinal is regular, hence a b-ordinal by Lemma 5(4).

(ii) This is (O3).

(iii) Let $W = \{x \mid \mathrm{ord}(x)\}$, so $F_b(W)$ is the class of c-ordinals and $F_c(W)$ the class of b-ordinals. Then W is a b-ordinal. Its b-members are regular c-ordinals, hence regular b-ordinals. Clause by clause:

- (O1): if $z \in_b y \in_b W$, then $z \in_c y$ since y is regular; so z is a c-ordinal by Lemma 5(1), and $z \in_b W$;
- (O2), (O3): the members are b-ordinals; use Lemma 5(2);
- (O4): Lemma 4;
- (O5): Lemma 7;
- (O6): apply Lemma 5(3) to the member x. □

Proposition 2 (Irregular ordinals on at most one side)**.** [Agda · Lean] *The two sides do not both have irregular ordinals.*

Proof. This proposition is from [H05, after Theorem 5.1], where the proof uses REG; ours does not. Suppose they do.

1. **Cut.** If p is an irregular s-ordinal and t a regular s-ordinal whose s-members are regular, then $t \in_s p$. Indeed, a regular s-ordinal z with $F_s(z) \subseteq F_s(p)$ satisfies $z = p$ or $z \in_s p$ by Lemma 6 (its condition is (O1) of z), and $z = p$ would make p regular; so $z \in_s p$. If some $x \in_s t$ is not in p, apply (O6) of t with $C = p$. This gives $z \in_s t$ with $z \notin_s p$ and $F_s(z) \subseteq F_s(p)$, where z is a regular s-ordinal, which is impossible. So $F_s(t) \subseteq F_s(p)$, and $t \in_s p$.
2. **Descent.** If there is an irregular s-ordinal α, then there is an irregular s-ordinal p whose s-members are all regular.
   - Let $D = \{x \mid x \text{ has a singleton}\}$. Then $F_s(D)$ is the class of objects with an s^*-singleton.
   - The s-members of α have s-singletons by (O4). So, by Lemma 4, an s-member of α lies in $F_s(D)$ iff it is regular.
   - If all s-members of α are regular, take $p = \alpha$. Otherwise some irregular $x \in_s \alpha$ satisfies $x \notin_s D$. Apply (O6) of α with $C = D$ to get $z \in_s \alpha$ with $z \notin_s D$ and $F_s(z) \subseteq F_s(D)$.
   - The s-members of z are in α and in D, hence regular. And z is irregular, since otherwise $z \in_s D$. By Lemma 5(1), z is an s-ordinal; take $p = z$.
3. **Conclusion.** By Descent, take such a p on side $\in$ and such a q on side ε.

- Each $\in$-member x of p is a regular $\in$-ordinal, hence an ε-ordinal by Lemma 5(4). Its members on either side are its $\in$-members, since x is regular; these lie in p by (O1), so they are regular. By Cut, $x \,\varepsilon\, q$.
- Symmetrically, every ε-member of q is an $\in$-member of p.
- So $F_\in(p) = F_\varepsilon(q)$, and $p = q$ by ME. Then $F_\varepsilon(p) = F_\in(p)$, so p is regular, a contradiction. □

Proof of Theorem 2. By Proposition 2 and classical logic, some side c has only regular ordinals. By Proposition 1, ord is asymmetric from c to c^*, which contradicts Theorem 1. □

Remark (The instances of COMP used). The proof of Theorem 2 uses only eight instances of COMP, so ME together with these eight instances is already inconsistent in classical logic. Here they are, with the object each one gives and where it is used:

- $x = x$: the universal set V, in Section 3;
- $x = r$ with r regular: a singleton of r, in the proof of Lemma 4;
- $\exists u\,(x \in u \wedge u \in h)$: the object k in the proof of Lemma 4;
- $x \in x \oplus \mathrm{ord}(r)$: the object A in the proof of Lemma 3, as used for Lemma 5(4);
- $x \in \alpha \wedge x \in \beta$: the object γ in the proof of Lemma 7;
- "x has a singleton": the object D in the proof of Proposition 2;
- $\mathrm{ord}(x)$: the object W in the proof of Proposition 1;
- the diagonal formula of Section 3 for $P = \mathrm{ord}$, with parameter V: the object A there.

Each has at most two parameters and quantifier rank at most 5, so the contradiction already lies in a small fragment of comprehension. Section 7 says what this means for the search for models.

Remark (Where ME is used). The proof of Theorem 2 applies ME to two objects that are not known to be regular only in the second claim of Lemma 4 and in the step $p = q$ of Proposition 2. Indeed, in classical logic COMP is already inconsistent with ME restricted to pairs x, y one of which is regular, ME restricted to pairs x, y such that every $\in$-member of x is regular, and the second claim of Lemma 4 (`me-localized`).

6 Verification

The code is in the repository `github.com/BedrockInstitute/DESTInconsistency`. The links [Agda · Lean] after the heads of definitions, theorems, lemmas, and propositions point to the corresponding code there.

- **Agda** (2.8.0.2). `agda/DESTInconsistency.agda` is a single file with no imports, whose `OPTIONS` pragma sets `--safe --without-K`. Sections 0–1 of the file (logic, syntax, readings, axioms, asymmetry) and the statements in its Section 4 are the trusted part. The theorems, quoted from the file, are

```
diagonal : (M : Set) (E : Rel M) → Comp M E →
           {k : ℕ} (P : Fm (suc k)) (ρ : Fin k → M) →
           ((i : Fin k) → Regular E (ρ i)) →
           {b c : Side} → Opp b c → ¬ Asymmetric E P ρ b c

inconsistent : (M : Set) (E : Rel M) → DEST⁻ M E → DNE → ⊥

DEST-inconsistent : (M : Set) (E : Rel M) → DEST M E → DNE → ⊥
```

`DNE` is double-negation elimination, i.e. classical logic; `diagonal` takes neither ME nor `DNE`. Check with

```
cd agda && agda --safe --no-libraries DESTInconsistency.agda
```

- **Lean 4** (v4.36.0, no libraries). `lean/DESTInconsistency.lean` has the same structure; its trusted part, Part I and the statements of Part III, is reproduced in Appendix A. Unlike in Agda, relations are `Prop`-valued and COMP yields ∃ rather than Σ; classical logic is the explicit hypothesis `∀ P : Prop, ¬¬P → P` of `inconsistent` and `DEST_inconsistent`. `#print axioms` reports that `diagonal`, `inconsistent`, and `DEST_inconsistent` depend on no axioms. Lean's `leanchecker` re-checks the compiled development, and the independent kernel nanoda 0.4.17 re-checks the 263 declarations these three theorems depend on, with no axioms permitted. `lean/Probes.lean` checks three things: ME plus Russell's instance alone has a one-point model; ME plus COMP without the regularity condition on parameters is inconsistent; and COMP is inconsistent by itself once each face uses the reading of its own side, $F_s(A) = \{x : \varphi^s(x)\}$. Check with

```
cd lean && lake build && lake env leanchecker DESTInconsistency
```

- **Additional code.** The files `agda/Intuitionistic.agda` and `lean/Intuitionistic.lean` give the variant without classical logic mentioned in Section 2; their trusted parts are those of the main files without classical logic. In the Lean file, `#print axioms` reports that the three theorems depend on no axioms, and nanoda re-checks the 267 declarations they depend on. `agda/Supplement.agda` proves the three theorems cited in Sections 1, 5, and 8, where their names, like the two file names in Section 2, link to the code; it imports nothing, and its trusted part is Sections 0–1 and the ordinal formula of the main Agda file, definitions used in the statements, and the statements. The Agda files are checked as the main one; `lake build` also builds the Lean file, and `leanchecker` re-checks it.
- **Blind formalizations.** Two isolated Claude Opus 5.5 instances formalized the argument in Agda from scratch, each seeing only its source: one the first draft of this paper, the other a one-page summary of the diagonal (`blind-test/diagonal-summary.zh.png`). Both type-check. The second found that the diagonal uses neither ME nor classical logic, which led to Theorem 1. A third instance, seeing only Sections 1–5 of the first version of this paper and a draft of the variant without classical logic (`blind-test/intuitionistic-input.md`), formalized it in Lean 4 from scratch (`blind-test/intuitionistic.lean`); its theorems depend on no axioms, and it found that the draft omitted the condition of Lemma 6 in Cut.
- **Scope.** `inconsistent` and `DEST_inconsistent` say that no structure satisfies the axioms. Passing to inconsistency of the first-order theory uses the completeness theorem, which is not formalized. In Lean, the ∃ of COMP can be used only to prove propositions, like an existential statement in a first-order derivation, whereas in Agda the Σ gives each witness as a function of the parameters. Passing from `lean/Intuitionistic.lean` to inconsistency of the first-order theory in intuitionistic logic uses the completeness of intuitionistic logic for Kripke models, which is not formalized, and the soundness of Lean's logic for these models, which we have not checked.

7 Retrospective

The result came out of an AI-assisted project, recorded in more than four hundred rounds, whose aim was to prove DEST consistent.

The facts were known early. Theorem 5.1 of [H05] and the discussion after it give the asymmetry of the ordinals, and being an ordinal is expressed by a parameter-free uniform

formula. The project treated the asymmetry as something a model must accommodate, and kept asking which models could do so. It seldom asked whether an instance of comprehension could use the asymmetry to decide whether a witness belongs to itself.

Two judgements kept it from that question.

- Attempts to single out the least irregular ordinal by a formula failed for a real reason: a formula true at an irregular ordinal is also true at some regular ordinal. The refutation needs only that some ordinal is omitted, not which one.
- The project proved correctly that if both sides use the same Boolean function to decide whether an object belongs to itself, no contradiction arises, and drew from it the rule that a one-bit diagonal cannot give a contradiction. The rule overlooked that a formula read on the two sides can act differently, because it consults the ordinals of its own side.

Progress was measured by the wrong yardstick. Most rounds built models for larger and larger fragments of comprehension. This does not measure the distance to the full scheme: two formulas can each have witnesses in a structure where their conjunction has none. The human lead's recurring question, where the unifying step was, pointed at exactly this gap. The proof of Theorem 2 makes this concrete: it needs only eight instances of COMP, each with at most two parameters and quantifier rank at most 5 (see the remark after it). So any fragment of comprehension that has a model together with ME, however large, leaves out at least one of these eight small instances.

The impasse led to an external audit. The rounds were carried out mainly by Claude. By round 401 the attempts at a consistency proof had stalled. The human lead then brought in ChatGPT (Astra 6.2), which had not taken part in the project, as an external auditor. Claude wrote it a self-contained brief: the project's assessments of where the problem was stuck, each to be confirmed or rejected, and a request for a model of DEST, a contradiction, or a reduction of its consistency to a known problem.

How the pieces were found. The diagonal appeared once the question became whether some A can make "I omit an ordinal of my side" true on one side and false on the other. For full DEST every ordinal fact used was already in Holmes's paper. The audit first derived the asymmetry of the ordinals without REG and then found the diagonal; Theorem 1 came from the second blind formalization.

8 Outlook

The inconsistency does not end the study of double extension. The theory recorded as open fails, and the search for a consistent axiomatization becomes harder than before: the natural formulation is gone, and a replacement has to be found and then shown consistent. Theorem 1 says what a replacement that keeps crossed comprehension must avoid: no uniform formula may be asymmetric. A classical replacement that keeps COMP, and keeps Lemma 2 in the case where z is also regular, has irregular ordinals on both sides (`irregular-ordinals`); so in such a replacement neither face of Ord consists exactly of the regular ordinals, and for ord the claim through which Holmes interprets ZF [H05, Claim 4.1] fails there; so his interpretation of ZF does not carry over to such a replacement as it stands. The variant without classical logic in Section 2 indicates that passing to intuitionistic logic does not avoid the contradiction. Each attempt to avoid the contradiction may weaken the theory a little, but that need not make it pointless. In our view, a theory that is finally shown consistent is worth having if it

still interprets ZF, still has the basic irregular objects such as Russell's $\{x \mid x \notin x\}$, and has axioms that read naturally rather than as patches against known paradoxes.

Statement on the use of AI

Peter Zhang led the research: he posed the questions, set the direction, made the decisions, and takes responsibility for the content. The work was done with AI systems. ChatGPT (Astra 6.2, OpenAI) became the project's external auditor after the attempts at a consistency proof had stalled (Section 7); it derived the asymmetry of the ordinals without REG and found the diagonal argument. Claude (Anthropic) audited and simplified the argument, generalized it to Theorem 1, wrote the Agda and Lean formalizations, and drafted this paper; most of this was done by Claude Opus 5.5 in Claude Code, and the blind formalizations were written by isolated Claude Opus 5.5 instances. In this research, Claude Opus 5.5 also found the variant without classical logic mentioned in Section 2 and the three results in `agda/Supplement.agda`; independent reviews by other Claude Opus 5.5 instances and the third blind formalization checked the variant. Earlier rounds of the project, aimed at a consistency proof, used Claude and other AI systems.

References

[H04] M. R. Holmes, *Paradoxes in double extension set theories*, Studia Logica **77** (2004), 41–57.

[H05] M. R. Holmes, *The structure of the ordinals and the interpretation of ZF in double extension set theory*, Studia Logica **79** (2005), 357–372.

[HFL] M. R. Holmes, T. Forster, and T. Libert, *Alternative set theories*, Sets and Extensions in the Twentieth Century (D. M. Gabbay, A. Kanamori, and J. Woods, eds.), Handbook of the History of Logic, vol. 6, Elsevier/North-Holland, Amsterdam, 2012, pp. 559–632.

[K89] A. Kisielewicz, *Double extension set theory*, Rep. Math. Logic **23** (1989), 81–89.

[K98] A. Kisielewicz, *A very strong set theory?*, Studia Logica **61** (1998), 171–178.

[SEP] M. R. Holmes, *Alternative axiomatic set theories*, The Stanford Encyclopedia of Philosophy, plato.stanford.edu/entries/settheory-alternative.

Appendix A Formal statements

```
namespace DEST

-- Part I. Trusted part
inductive Side where
  | mem
  | eps

inductive Opp : Side → Side → Prop where
  | me : Opp .mem .eps
  | em : Opp .eps .mem

def cons {α : Type} {n : Nat} (x : α) (ρ : Fin n → α) : Fin (n + 1) → α
  | ⟨0, _⟩ => x
  | ⟨i + 1, h⟩ => ρ ⟨i, Nat.lt_of_succ_lt_succ h⟩
```

```
def nil {α : Type} : Fin 0 → α
  | ⟨_, h⟩ => absurd h (Nat.not_lt_zero _)

inductive Fm : Nat → Type where
  | mem : Fin n → Fin n → Fm n
  | eq  : Fin n → Fin n → Fm n
  | not : Fm n → Fm n
  | and : Fm n → Fm n → Fm n
  | or  : Fm n → Fm n → Fm n
  | imp : Fm n → Fm n → Fm n
  | all : Fm (n + 1) → Fm n
  | ex  : Fm (n + 1) → Fm n

def Sat {M : Type} (E : Side → M → M → Prop) (s : Side) :
    {n : Nat} → Fm n → (Fin n → M) → Prop
  | _, .mem i j, ρ => E s (ρ i) (ρ j)
  | _, .eq i j,  ρ => ρ i = ρ j
  | _, .not φ,   ρ => ¬ Sat E s φ ρ
  | _, .and φ ψ, ρ => Sat E s φ ρ ∧ Sat E s ψ ρ
  | _, .or φ ψ,  ρ => Sat E s φ ρ ∨ Sat E s ψ ρ
  | _, .imp φ ψ, ρ => Sat E s φ ρ → Sat E s ψ ρ
  | _, .all φ,   ρ => ∀ x : M, Sat E s φ (cons x ρ)
  | _, .ex φ,    ρ => ∃ x : M, Sat E s φ (cons x ρ)

def Regular {M : Type} (E : Side → M → M → Prop) (a : M) : Prop :=
  ∀ x : M, E .mem x a ↔ E .eps x a

def Comp (M : Type) (E : Side → M → M → Prop) : Prop :=
  ∀ {k : Nat} (φ : Fm (k + 1)) (ρ : Fin k → M), (∀ i : Fin k, Regular E (ρ i)) →
    ∃ A : M, ∀ x : M,
      (E .eps x A ↔ Sat E .mem φ (cons x ρ)) ∧ (E .mem x A ↔ Sat E .eps φ (cons x ρ))

structure DESTminus (M : Type) (E : Side → M → M → Prop) : Prop where
  ME : ∀ a b : M, (∀ x : M, E .mem x a ↔ E .eps x b) → a = b
  COMP : Comp M E

structure DEST (M : Type) (E : Side → M → M → Prop) : Prop where
  core : DESTminus M E
  REG : ∀ a b : M,
    ((∀ x : M, E .mem x a → E .mem x b) ∨ (∀ x : M, E .eps x a → E .eps x b)) →
    (∀ y : M, (E .mem y b ∨ E .eps y b) → Regular E y) →
    Regular E a

def Asymmetric {M : Type} (E : Side → M → M → Prop) {k : Nat}
    (P : Fm (k + 1)) (ρ : Fin k → M) (b c : Side) : Prop :=
  (∀ x : M, Sat E c P (cons x ρ) → Sat E b P (cons x ρ)) ∧
  (∀ x y : M, Sat E b P (cons x ρ) → E b y x → ¬ E b y y) ∧
  ∃ W : M, (∀ x : M, E b x W ↔ Sat E c P (cons x ρ)) ∧
           (∀ x : M, E c x W ↔ Sat E b P (cons x ρ)) ∧ Sat E b P (cons W ρ)
```

```
-- Part III. The theorems, statements only (Part II, the proofs, is omitted)
theorem diagonal (M : Type) (E : Side → M → M → Prop) (C : Comp M E)
    {k : Nat} (P : Fm (k + 1)) (ρ : Fin k → M)
    (rρ : ∀ i : Fin k, Regular E (ρ i))
    {b c : Side} (o : Opp b c) : ¬ Asymmetric E P ρ b c

theorem inconsistent (M : Type) (E : Side → M → M → Prop)
    (dne : ∀ P : Prop, ¬¬P → P) (T : DESTminus M E) : False

theorem DEST_inconsistent (M : Type) (E : Side → M → M → Prop)
    (dne : ∀ P : Prop, ¬¬P → P) (T : DEST M E) : False
```