TEASING DEFINITIONAL EQUIVALENCE AND BI-INTERPRETABILITY APART

JASON CHEN & TOBY MEADOWS

Abstract. In a recent paper, Enayat and Łełyk [2024] show that second order arithmetic and countable set theory are not definitionally equivalent. It is well known that these theories are bi-interpretable. Thus, we have a pair of natural theories that illustrate a meaningful difference between definitional equivalence and bi-interpretability. This is particularly interesting given that Visser and Friedman [2014] have shown that a wide class of natural foundational theories in mathematics are such that if they are bi-interpretable, then they are also definitionally equivalent. The proof offered by Enayat and Łełyk makes use of an inaccessible cardinal. In this short note, we show that the failure of bi-interpretability can be established in Peano Arithmetic merely supposing that one of our target theories are consistent.

We begin by recalling some basic definitions and set up our notation. A more precise and detailed discussion can be found in [Visser, 2006, Visser and Friedman, 2014, Halvorson, 2019, Button and Walsh, 2018] or [Meadows, 2023]. Let T and S be theories articulated in the language $\mathcal{L}_T$ and $\mathcal{L}_S$ respectively. Suppose that they are mutually interpretable as witnessed by translations $t : \mathcal{L}_S \to \mathcal{L}_T$ and $s : \mathcal{L}_T \to \mathcal{L}_S$ giving rise to functions $t : mod(T) \to mod(S)$ and $s : mod(S) \to mod(T)$ where $mod(T)$ and $mod(S)$ are the classes of models satisfying T and S respectively.[1] We say that T and S are *definitionally equivalent* if:

(1) $\mathcal{A} = s \circ t(\mathcal{A})$ for all models $\mathcal{A}$ of T; and
(2) $\mathcal{B} = t \circ s(\mathcal{B})$ for all models $\mathcal{B}$ of S.

On the other hand, we say that T and S are *bi-interpretable* if, in addition to t and s witnessing mutual interpretability, there are functions η and ν, uniformly definable over T and S respectively, such that:

(1) $\eta^{\mathcal{A}} : \mathcal{A} \cong s \circ t(\mathcal{A})$ for all models $\mathcal{A}$ of T; and
(2) $\nu^{\mathcal{B}} : \mathcal{B} \cong t \circ s(\mathcal{B})$ for all models $\mathcal{B}$ of S.

Informally, we have definitional equivalence when we have translations that allows us to go back and forth to exactly where we started. Bi-interpretability, by contrast, is weaker in that we only return to an isomorphic structure where the relevant isomorphism is definable. For most mathematical purposes bi-interpretability seems to be compelling evidence that the theories in question can be regarded as informally equivalent.

However, it is not difficult to find a toy example that pulls these equivalence relations apart. Let $\mathcal{L}_T$ be the empty language and let T be the theory that says there are infinitely many objects. More specifically, we let T consist of the sentences $\exists_{\geq n} x\ x = x$ for all $n \in \omega$. Let $\mathcal{L}_S$ be the language

We're grateful to Ali Enayat for looking over our sketches of these proofs and providing helpful insights and encouragement.

[1] We permit translations that make use of multiple dimensions and quotients.

consisting of a single constant symbol c. And let S consist of the same sentences as T. Thus, S says nothing at all about c. We now sketch a quick proof of the following:

Proposition 1. *T and S are bi-interpretable, but not definitionally equivalent.*

Proof. (Bi-interpretability) The essential idea is to move from a model of S to a model of T by discarding the object denoted by c; and in the other direction, we move from a model of T to models of S by using a three-dimensional quotient interpretation t to "add" a new object that will can be denoted by c. More specifically, given a model $\mathcal{A} = \langle A \rangle$ of T, we let our domain be A^3 and then define an interpretation for $\doteq$ and c. Given $\bar{x} = \langle x_1, x_2, x_3 \rangle$, $\bar{y} = \langle y_1, y_2, y_3 \rangle$ from A^3 we let

$$\bar{x} \doteq \bar{y} \;\Leftrightarrow\; (x_2 = x_3 \wedge y_2 = y_3 \wedge x_1 = y_1) \vee (x_2 \neq x_3 \wedge y_2 \neq y_3)$$

and

$$\bar{x} = c \;\Leftrightarrow\; x_2 \neq x_3.$$

Putting all of this together, we let $t(\mathcal{A}) = \langle A^3, c^{\mathcal{A}}, \doteq \rangle$. In the other direction, we start with a model $\mathcal{B} = \langle B, b \rangle$ of S and provide an interpretation s by defining a new domain using the formula

$$\delta_s(x) := (x \neq c).$$

We then let $s(\mathcal{B}) = \langle \delta_s^{\mathcal{B}} \rangle$ where $\delta_s^{\mathcal{B}} = \{x \in B \mid \mathcal{B} \models \delta_s(x)\}$. Clearly, t and s witness mutual interpretability. Finally, we define the required isomorphisms. Given a model $\mathcal{A}$ of T, we have $s \circ t(\mathcal{A}) = \langle E, \doteq^{s o t(\mathcal{A})} \rangle$ where $E = \{\bar{x} \in A^3 \mid \mathcal{A} \models x_2 = x_3\}$ and for $\bar{x}, \bar{y} \in E$, $\doteq^{s o t(\mathcal{A})} = \{\langle \bar{x}, \bar{y} \rangle \in (A^3)^2 \mid \mathcal{A} \models x_1 = y_1\}$. Thus, we may let our isomorphism be defined by the formula $\eta(x, \bar{y})$ be $x = y_1 = y_2 = y_3$. The definition of ν uses similar techniques and we leave it to the reader.

(Not definitionally equivalence) Suppose toward a contradiction we have interpretation as giving rise to functors $t : mod(T) \to mod(S)$ and $s : mod(S) \to mod(T)$ witnessing the definitional equivalence of T and S. If we start with a model $\mathcal{B} = \langle B, b \rangle$ of S, we have just two choices for our interpretation:

(1) we can remove – as we did above – b from the domain to obtain $\langle B \backslash b \rangle$; or
(2) we can simply forget the denotation of b and retain the domain to give $\langle B \rangle$.

If we take option (1), we cannot have definitional equivalence since we have moved to a proper subset of the original domain, which cannot be recovered. So we are stuck with option (2) and we have $t(\mathcal{B}) = \langle B \rangle$. Now if we are to have $s \circ t(\mathcal{B}) = \mathcal{B}$, we must have $s(\langle B \rangle) = \langle B, x \rangle$ for some $x \in B$. But this would mean that we could define an element of an infinite set in the empty language. This is plainly impossible as can be established with a simple automorphism argument. □

So much for toy theories. If we move to theories with more serious foundational credentials, things get more interesting. First we recall a few definitions.[2] We say that an interpretation is *one-dimensional*, if the formula defining the domain of the interpretation has one free variable and thus, defines a subset of domain of models of the interpreting theory. Say that an interpretation is *identity preserving* if it translates the identity predicate to itself.[3] We say that a one-dimensional interpretation is *unrelativized*

[2]These can be found in [Visser and Friedman, 2014] but we include them to make things a little more self-contained.
[3]So the translation t in the proof of Proposition 1 is multi-dimensional since it uses ordered triples. Moreover, we see that it is not identity-preserving since, in order to form the quotient, we make an explicit definition of the identity relation, denoted $\doteq$, on those triples.

if it does not restrict the domain.[4] We say that a one-dimensional interpretation is *direct* if it is unrelativized and preserves identity. A theory is said to be *sequential* if it directly interprets adjunctive set theory, which is the theory saying that: there is an empty set; and for any sets x, y there is some z containing exactly y and the members of x. Such a theory is capable of doing some basic coding. Any serious foundational theory is obviously sequential.

Theorem 2. *[Visser and Friedman, 2014] Let T be a sequential theory.*[5] *Suppose that T and S are bi-interpretable as witnessed by one-dimensional, identity preserving interpretations. Then T and S are definitionally equivalent.*

This theorem is particularly helpful. For example, it is well-known that: Peano arithmetic is bi-interpretable with a finite version of ZFC;[6] and ZFC is bi-interpretable with ZFC with foundation removed and replaced by Aczel's anti-foundation axiom. Each of these theories is sequential and the interpretations linking them can be arranged to be identity preserving. Thus, we see that both pairs are actually examples of definitionally equivalent theories. However, a crucial element in these argument is the ability of these theories to eliminate the use of quotient interpretations by finding representatives for the equivalence classes. In arithmetic, we just need to pick the least element; in set theory, we use Scott's trick, whereby we take the set of elements of the equivalence class of minimal rank. But not all theories can perform this kind of trick. In set theory, we seem to need some form of reflection in order to eliminate quotients.[7]

This idea provides a lead toward a natural pair of theories that are bi-interpretable but not definitionally equivalent.[8] In particular, ZFC^- (ZFC without the powerset axiom)[9] cannot perform Scott's trick.[10] And indeed, it is in this area that Enayat and Łełyk find their example. Let ZFC_{count} be ZFC^- with an axiom stating that every set is countable.[11] Let SOA be the theory of second order arithmetic with full comprehension and choice for all definable sets of reals indexed by naturals.[12] First we note that:

---

[4]The interpretation s in Proposition 1 is relativized since we use the formula δ_s to remove c from the domain of the interpretation.

[5]Visser and Friedman actually use a weaker class of theories called *conceptual theories* in their paper, but this will not affect the discussion in this paper.

[6]Some care is required in the axiomatization. We assume that we are using set induction rather than foundation. Or we can add an axiom stating that every set is contained in a transitive set. See [Kaye and Wong, 2007].

[7]In model theory, this is known as *eliminating imaginaries*. See Section 4.4 of [Hodges, 1997].

[8]We note that Visser and Friedman [2014] do provide a pair of sequential theories that are bi-interpretable but not definitionally equivalent. The example is interesting, however, the theories are not in common use and might be thought of as being contrived for the purposes of the result. As such, we regard them as an *unnatural* pair of theories.

[9]We should also use Collection rather than Replacement. See [Gitman et al., 2016].

[10]It is worth noting that ZFC^- can be augmented to a theory that can eliminate imaginaries by, for example, adding the assumption that $V = L$ which ensures that the universe has a definable well-ordering.

[11]Note that the C in ZFC_{count} is redundant, since the $_{count}$ axiom ensures that every set is well-ordered by a bijection with ω.

[12]We adopt the logicians' reals in the paper and say that $\mathbb{R} = \mathcal{P}(\omega)$. The choice schema then says that for any formula $\varphi(n, Y)$ in the language of SOA, if for all n there is some Y such that $\varphi(n, Y)$, then there is some Z such that for all n, $\varphi(n, (Z)_n)$, where $(Z_n) = \{i \in \omega \mid (i, n) \in Z\}$. See [Enayat and Łełyk, 2024] and [Simpson, 1999] for more details. In the latter, this theory is denoted as $\Sigma^1_\infty - AC_0$ noting that the definable choice principles end up implying the comprehension principles. See Section VII.6 and Lemma VII.6.6 in [Simpson, 1999] for more information. We note that these choice principles are required since there are models of second order arithmetic with full comprehension where definable choice fails for a Σ^1_3-set: see Remarks VII.6.3 in [Simpson, 1999] and Example 15.57 in [Jech, 2003]. Finally, we also note that SOA is also often notated as Z_2 while PA is often denoted as Z_1 [Hilbert and Bernays, 1934]. However, it is not always clear whether Z_2 is intended to include the definable choice principles we incorporate in SOA.

Theorem 3. *[Mostowski, 1961][13] ZFC_{count} is bi-interpretable with SOA.*

Just to give the idea, we first note that since PA and finite ZFC are definitionally equivalent,[14] we can, without harm, think of the number domain of a model of SOA as being V_ω. To move from a model of ZFC_{count} to a model of SOA we just forget all the sets with rank $> \omega$. This direction is trivial. In the other direction, we start with a model of SOA and simulate the effect of hereditarily countable sets using well-founded, extensional relations R on ω with top elements that collapse to be such sets.[15] Since there are many such relations that collapse to a particular hereditarily countable set, the sets are not in bijection with their representatives. This is addressed by a quotient interpretation that defines a natural notion of identity on these relations. The required definable isomorphisms are then given by: sending sets in a model of ZFC_{count} to those reals that collapse to code them; and sending numbers and sets in models of SOA to those reals coding sets that are appropriately isomorphic to them.[16] Given that a quotient is required, this gives us some reason to doubt that these theories are definitionally equivalent. Enayat and Łełyk have confirmed this intuition.

Theorem 4. *[Enayat and Łełyk, 2024] ZFC_{count} is not definitionally equivalent with SOA, provided that there is an inaccessible cardinal or there is an ω-model of the theory extending ZFC by an inaccessible cardinal.*[17]

Unlike the toy example above, this provides a concrete example of a pair of well-understood and very commonly used theories that are bi-interpretable but not definitionally equivalent. We shall now prove this claim without the inaccessible cardinal. Moreover, we'll do this twice. First, with a simple ZFC proof; and second, with an indirect proof in PA. The first proof exploits the fact that certain models of SOA cannot define well-orderings of length $\geq \omega_1$. The second proof exploits the fact that certain models of ZFC_{count} cannot define linear orders of their domain.

---

[13]The attribution for this result is a little convoluted. One reason for this is that while the bones of the proof seem to have been around since the late 1950s, the definition of bi-interpretability wasn't formally isolated until [Ahlbrandt and Ziegler, 1986], although similar ideas were in circulation in the 1970s: see, for example, [Osius, 1974]. We follow convention and attribute the result to Mostowski, although [Mostowski, 1961] only appears to contain the easy direction of the proof that delivers a model of SOA from ZFC_{count} essentially by truncation: see Theorem 7.15 in [Mostowski, 1979]. Given that the key trick in the other, more difficult direction involves Mostowski's famous collapse function, this convention still seems apropos. A detailed proof that SOA can interpret ZFC_{count} using trees is provided in Theorem 5.5 of [Apt and Marek, 1974] where they attribute the result to [Kreisel, 1968] and [Zbierski, 1971]. More subtle results on these interpretations can be found in Section VII.3 of [Simpson, 1999] and, more recently, [Kanovei and Lyubetsky, 2025].

[14]This follows from Corollary 5.5 in [Visser and Friedman, 2014] and the fact that PA and finite ZFC are bi-interpretable via identity preserving interpretations. Note also that by finite ZFC, we mean a theory that uses the Set Induction schema rather than Foundation; or alternatively, we could include an axiom saying that every set is contained in a transitive set. See [Kaye and Wong, 2007] for more details. Also note that we could just use ω and its subsets from a model of ZFC_{count} to deliver a model of SOA.

[15]A detailed description of this kind of construction can be found in Chapter VII.3 of [Simpson, 1999].

[16]Note that these isomorphisms are only "functions" relative to the defined notion of identity in the quotient interpretation. In particular, two reals are deemed identical if they collapse to the same set.

[17]We note that the statement of this theorem in [Enayat and Łełyk, 2024] makes use of the weaker assumption that ZFC plus an inaccessible cardinal is consistent. However, the proof given there seems to naturally work by using one of the assumptions above. If we collapse an inaccessible cardinal, then the resultant version of $H(\omega_1)$ gives us the model of ZFC_{count} we want and the result can then be pulled back to the ground universe. But if we merely start with a model $\mathcal{M}$ of ZFC plus an inaccessible and collapse, then the argument of the proof just shows that $\mathcal{M}$ thinks SOA and ZFC_{count} aren't definitionally equivalent. We have no reason to think $\mathcal{M}$ is correct about this unless $\mathcal{M}$ is, say, an ω-model. It is not so easy to say which of these assumptions is preferable. While the first assumption has lower consistency strength than the second, the second – unlike the first – can be accommodated by theories with less ontological overheads than ZFC.

1. Simple proof in ZFC

We start by recalling two well-known facts about forcing and use them to prove a lemma from which our main claim quickly follows.

Fact 5. *[Lévy, 1965] Let $\mathbb{P}$ be weakly homogeneous and G be $\mathbb{P}$-generic over V. Suppose $A \subseteq V$ and A is definable in $V[G]$ by a formula using parameters from V.*[18] *Then $A \in V$.*

Fact 6. *(Laver-Woodin)*[19] *If V is a generic extension of some inner model W, there is a formula defining W in V using a parameter from W. More specifically, there is a formula $\varphi(x, y)$ and $r \in W$ such that for all x*

$$x \in W \;\Leftrightarrow\; \varphi(x, r)^V.$$

The following definition is the focus of our lemma below.

Definition 7. Let $\Gamma \subseteq \mathcal{P}(\mathbb{R})$ be a point-class. We say that Γ has the *short well-ordering property* if every well-ordering $R \in \Gamma$ has order type $< \omega_1$.

The following lemma provides the combinatorial content for our theorem.

Lemma 8. *There is a generic extension $V[G]$ of V that thinks $OD \cap \mathcal{P}(\mathbb{R})$ has the short well-ordering property.*

Proof. We assume CH holds in V; if necessary, just collapse the ordinals below $2^{\aleph_0}$. Now let G be $Col(\omega, \{\omega_1\})$-generic[20] over V and work in $V[G]$. Suppose toward a contradiction that R is an OD -well-ordering of reals that has order-type $\geq \omega_1$. Then since we've collapsed V's continuum, we see that $field(R) \cap (V[G]\backslash V) \neq \emptyset$. And since R is a well-ordering, we may define the R-least element x of $V[G]\backslash V$. This definition only makes use of the ordinals used in the definition of R and a parameter from V given by Fact 6 that allows us to define V in $V[G]$. Fact 5, then tells us that $x \in V$, which is a contradiction. □

This next lemma shows that the combinatorial lemma is enough for a goal.

Lemma 9. *Suppose every ordinal definable well-ordering of a set of reals is shorter than ω_1. Then SOA is not definitionally equivalent with ZFC_{count}.*

Proof. Let S be the first theory and let T be the second. Suppose toward a contradiction that we have interpretations

$$t : mod(T) \leftrightarrow mod(S) : s$$

[18]See the beginning of Section 2 in [Woodin et al., 2012] for a nice sketch of the idea behind the proof of this. If such a set were definable in $V[G]$, then the homogeneity of $\mathbb{P}$ ensures that it can already be defined in V using the forcing relation.

[19]See [Reitz, 2007] or [Woodin, 2012] for a proof of this.

[20]Here, I'm following the definition given in Chapter 10 [Kanamori, 2003]. Thus, $Col(\omega, S)$ for $S \subseteq Ord$ is intended to collapse every ordinal in S to be countable in a generic extension. It is worth noting this since some authors will write $Col(\omega, \omega_1)$ to denote what we are calling $Col(\omega, \{\omega_1\})$.

witnessing their definitional equivalence. Note that $M = \langle H_{\omega_1}, \in\rangle \models T$. Then we see that $t(M) = \langle N_0, N_1, \sigma\rangle \models S$.[21] Now if $s \circ t(M) = M$, we must be able to define a well-ordering of length ω_1 in $t(M)$. We show that this is impossible.

Note first that the naturals N_0 of $t(M)$ must be well-founded since $\omega = \omega^{sot(M)}$ is definable in $t(M)$ and if its naturals were ill-founded, then $t(M)$ couldn't define a well-ordering of type ω. This means that we can collapse $t(M)$ to form a model $N^* = \langle \omega, N_1^*, \sigma^*\rangle \cong t(M)$ where $N_1^* \subseteq \mathcal{P}(\omega)$. Now if $t(M)$ can define a well-ordering of length ω_1, then so can N^*. Moreover, since $N_1^* \subseteq \mathcal{P}(\omega)$ we may assume that any such well-ordering defined over N^* is a well-ordering of a set of reals. But then since N_1^* is definable from H_{ω_1} and t, we see that such a well-ordering is definable in the parameter, ω_1, and thus, ordinal definable. This contradicts our initial assumption. □

Remark. The proof above also shows that ZFC_{count} is not a retract (as in "half" of a bi-interpretation) of SOA by one-dimensional interpretations. It can also be shown that ZFC_{count} is solid[22] and from this it can be shown that the model $t(M)$ in the proof above is actually the standard model of second order arithmetic; i.e., $N_1^* = \mathcal{P}(\omega)$.

Finally, we put the two lemmas together to get the desired result.

Theorem 10. *SOA is not definitionally equivalent with ZFC_{count}.*

Proof. Use Proposition 8 to move to a generic extension $V[G]$ of the universe in which every ordinal definable well-ordering of a set of reals is shorter than ω_1. Then Lemma 9, tells us that the stated theories are not definitionally equivalent in $V[G]$. The statement that these theories are definitionally equivalent is arithmetic,[23] thus, if it is true in $V[G]$ it is also true in V. □

2. Proof in PA

As in the previous section, we start by proving a more general lemma from which the main claim follows.[24]

Lemma 11. *If we add a Cohen real to the universe V, then there is no ordinal definable relation S on $\mathcal{P}(\mathbb{R})$ that is connected and asymmetric.*

Proof. We are essentially using a simplification of the proof that the axiom of choice fails in the second Cohen model as delivered in Theorem 5.19 of [Jech, 2008]. The plan is to describe, in a generic extension, a pair P of sets of reals that has no ordinal definable element. To see that this suffices, suppose there was an ordinal definable relation S on $\mathcal{P}(\mathbb{R})$ that is connected and asymmetric. Then we may obtain an ordinal definable element of the pair P by taking the S-least element of the P.

[21] Here N_0 is the number domain, N_1 is the set domain, and σ is the interpretation of the non-logical vocabulary.

[22] See [Enayat, 2016] for the definition of solidity and a proof that this theory is solid.

[23] In particular, definitional equivalence can be articulated as a Σ^0_3 statement: see Fact 14 below. To establish this, observe that T and S are definitionally equivalent if *there are natural numbers* coding computable translations t and s such that: *every natural number* coding a sentence φ such that either *there is no natural number* coding a proof witnessing $S \vdash \varphi$ or *there is a natural number* coding a proof witnessing that $T \vdash t(\varphi)$; and *every natural number* coding a formula $\psi(\bar{x})$ in the language of T is such that *there is a natural number* coding a proof that $T \vdash \forall\bar{x}(\psi(\bar{x}) \leftrightarrow t\circ s(\psi)(\bar{x}))$; and a pair of similar clauses for sentences $\psi \in T$ and formulae $\varphi(\bar{y})$ in the language of S.

[24] For a related result establishing that there are models of ZFC with no definable global linear ordering with Cohen forcing, see Theorem 3.1 in [Enayat, 2004].

We start by defining a forcing $\mathbb{P} = \langle P, \leq \rangle$ where P consists of partial functions

$$p : (2 \times \omega) \times \omega \rightharpoondown 2$$

ordered by reverse inclusion. This is intended to deliver us a pair of sets each containing infinitely many Cohen reals. But do note that, in the codes, this is a minor variation of the usual forcing to add a Cohen real. Let G be $\mathbb{P}$-generic over V. We then define some useful names and the objects they denote in $V[G]$. For $n, i, j \in \omega$ and $e \in 2$, we let

- $\dot{x}_{e,i} = \{\langle \check{j}, p \rangle \mid p(n, e, i, j) = 1\}$;
- $x_{e,i} = \{j \in \omega \mid \exists p \in G\ p(n, e, i, j) = 1\}$ be a real;
- $\dot{X}_e = \{\langle \dot{x}_{n,e,i}, 1 \rangle \mid i \in \omega\}$;
- $X_e = \{x_{n,e,i} \mid i \in \omega\}$ be a countable set of reals;
- $\dot{P} = \{\langle \dot{X}_{n,0}, 1 \rangle, \langle \dot{X}_{n,1}, 1 \rangle\}$; and
- $P = \{X_{n,0}, X_{n,1}\}$ be a pair of countable sets of reals

This gives us the set P that we want. P contains exactly two sets X_0 and X_1. X_0 contains a set of Cohen reals $x_{0,i}$ for all $i \in \omega$; and similarly, for $X_{n,1}$. Note that a simple density argument reveals that $\Vdash x_{e,i} \neq x_{e^*,i^*}$ whenever $\langle e, i \rangle \neq \langle e^*, i^* \rangle$. Thus, $\Vdash \dot{X}_0 \neq \dot{X}_1$. Next observe that any permutation of $2 \times \omega$ delivers an automorphism of $\mathbb{P}$ that can be extended to a map from $V^{\mathbb{P}}$ to itself in a natural way. Moreover, for any such automorphism, we have

$$p \Vdash \varphi(\dot{y}_0, ..., \dot{y}_n) \Leftrightarrow \pi(p) \Vdash \varphi(\pi \dot{y}_0, ..., \pi \dot{y}_n).$$

We now claim that there is no ordinal definable set denoted by a term $t = t(\bar{\alpha})$ such that $t \in P$. To see this, suppose toward a contradiction that there is some $p_0 \in G$ that forces that there is such a t. Note that since t is ordinal definable, it is not affected by automorphisms of $\mathbb{P}$. We may then fix some $p \leq p_0$ with $p \in G$ that decides the value of t. For definiteness, suppose that $t = X_0$, so we have

$$p \Vdash t = \dot{X}_0.$$

Now it can then be seen that there is an automorphism π of $\mathbb{P}$ such that:

- $\pi(p) \not\perp p$;
- $\pi \dot{P} = \dot{P}$; and
- $\pi \dot{X}_0 = \dot{X}_1$.

We just describe the underlying permutation and leave the proof of these facts to the reader.[25] First, fix a sufficiently large $k \in \omega$ that for all $i \geq k$, p cannot decide $x_{0,i} \in X_0$ or $x_{1,i} \in X_1$. We continue by informally describing π by its behavior in the generic extension. First, we swap the interval $[0, k)$ of Cohen reals associated with X_0 with the interval $[k, 2k)$ associated with X_1. Then we swap the interval $[0, k)$ of Cohen reals associated with X_1 with the interval associated with $[k, 2k)$ in $X_{0,0}$. Above $2k$, we just swap the Cohen reals associated with X_0 with those of X_1. Much more formally, $\pi : 2 \times \omega \to 2 \times \omega$

[25]A definition of a very similar permutation π can be found at the end of the proof of Lemma 5.19 at the top of page 71 in [Jech, 2008].

is such that for all $e \in 2$ and $i \in \omega$

$$\pi(e,i) = \begin{cases} \langle e-1, i+k\rangle & \text{if } i \in [0,k) \\ \langle e-1, i-k\rangle & \text{if } i \in [k,2k) \\ \langle e-1, i\rangle & \text{if } i \in [k,\omega) \end{cases}$$

Recalling that t is unaffected by π, we see that

$$\pi(p) \Vdash t = \pi\dot{X}_0$$

and so

$$p \cup \pi(p) \Vdash t = \dot{X}_1$$

which is a contradiction, since $\Vdash \dot{X}_0 \neq \dot{X}_1$. □

The Lemma above is demonstrated using ZFC. However, it is not difficult to see that the proof can be adapted to the context of ZFC_{count} where we use ordinary definability rather than ordinal definability. Moreover, the proof works to show that there is no such definable relation on any domain (including the universe itself) that extends the set A delivered in the proof.

Corollary 12. *(ZFC_{count}) If we add a Cohen real to the universe, then there is no definable relation S on the extended universe that is connected and asymmetric. More precisely, for all formula $\varphi_S(x,y)$ of $\mathcal{L}_\in$, we have*

$$\Vdash_{Add(\omega,1)} \exists x \exists y(\varphi_S(x,y) \leftrightarrow \varphi_S(y,x)).$$

Before, we prove the result in PA, let us first give a quick proof in ZFC_{count} that may give a clearer picture of the underlying idea.

Theorem 13. *(ZFC_{count}) SOA is not definitionally equivalent with ZFC_{count}, if one of these theories is consistent.*

Before we give the proof, let us first discuss the statement of this theorem and how it differs from Theorems 4 and 10. The first thing to note is that Theorem 4 makes use of a consistency assumption that cannot be proved in ZFC. Theorem 10 arguably improves this by removing that assumption and just proving the result in ZFC. Theorem 13, however, also uses a consistency statement. One might think that this has taken us a step backwards, but this would be misguided. Here, we require the assumption since ZFC_{count} cannot prove the consistency of either SOA or ZFC_{count}, although it can prove their equiconsistency. Without the ability to prove the consistency of one of these theories, it will not be possible to deliver a model witnessing the failure definitional equivalence. Moreover, if one (and thus, both) of them are inconsistent they will be vacuously definitionally equivalent since they have no models. In contrast, the background assumptions of both Theorems 4 and 10 are sufficient to prove the consistency of both SOA and ZFC_{count}. Thus, Theorem 10 has also been improved by using a background theory that is insufficient to prove the consistency of its target theories.Of course, other theories weaker than ZFC_{count}, like KP, will also be able to prove the theorem above. Our final result below is proved in the arithmetic theory PA, which seems fitting since definitional equivalence can be articulated as an arithmetic statement.

Proof. We work informally in ZFC_{count}. Suppose toward a contradiction that we have interpretations

$$t : mod(ZFC_{count}) \leftrightarrow mod(SOA) : s$$

witnessing that SOA and ZFC^{-} are definitionally equivalent. Let $\mathcal{M}$ be a model of ZFC_{count} and without loss of generality suppose that it is countable and satisfies $V = L$. Now let c be a Cohen real over $\mathcal{M}$. Then $\mathcal{M}[c]$ satisfies ZFC_{count} and the statement that its universe is constructed from c; i.e., $V = L[c]$.[26] Using Lemma 12 in $\mathcal{M}[c]$, we see that $\mathcal{M}[c]$ cannot define a linear ordering of its domain. On the other hand, $t(\mathcal{M}[c])$, as a model of SOA, can easily define a linear order of its entire domain using, say, the lexicographic ordering of 2^{ω}. But since ZFC_{count} and SOA are definitionally equivalent, we see that $\mathcal{M}[c]$ and $t(\mathcal{M}[c])$ share the same domain and thus, any relation definable over $t(\mathcal{M}[c])$ is also definable $\mathcal{M}[c]$. This is a contradiction. □

We're almost ready for the final result, but it will be helpful to first give this alternative, syntactic characterization of definitional equivalence that is amenable to use in theories of arithmetic.

Fact 14. *Let T_0 and T_1 be theories articulated in $\mathcal{L}_0$ and $\mathcal{L}_1$ respectively. Then T_0 and T_1 are definitionally equivalent, if there are translations $t_0 : \mathcal{L}_1 \to \mathcal{L}_0$ and $t : \mathcal{L}_0 \to \mathcal{L}_1$ giving rise to direct interpretations such that for $i \in \{0, 1\}$:*[27]

(1) *For all sentences $\varphi \in \mathcal{L}_i$, if $T_i \vdash \varphi$, then $T_{i-1} \vdash t_{i-1}(\varphi)$; and*
(2) *For all formulae $\varphi(\bar{x}) \in \mathcal{L}_i$, $T_i \vdash \forall \bar{x}(\varphi(\bar{x}) \leftrightarrow t_i \circ t_{i-1}(\varphi)(\bar{x}))$.*

Finally, we are in position to demonstrate the main claim, by proving a lemma from which the result follows trivially.

Lemma 15. *(PA) ZFC_{count} cannot interpret SOA via a direct interpretation; i.e., an interpretation that preserves identity and the domain, if either ZFC_{count} or SOA is consistent.*

Proof. We proceed by contraposition working informally in PA. Thus, we suppose that we have a translation $t : \mathcal{L}_{SOA} \to \mathcal{L}_{\in}$ giving rise to a direct interpretation such that for all $\varphi \in \mathcal{L}_{SOA}$, if $SOA \vdash \varphi$, then $ZFC_{count} \vdash t(\varphi)$. And we aim to show that ZFC_{count} and SOA are inconsistent. First observe that using, say, the lexicographic ordering on subsets of ω, there is formula $\psi(x, y)$ such that

$$SOA \vdash \forall x \forall y(\psi(x, y) \leftrightarrow \neg\psi(y, x)).$$

Thus, we see that

$$ZFC_{count} \vdash \forall x \forall y(t(\psi)(x, y) \leftrightarrow \neg t(\psi)(y, x)).$$

However, by Corollary 12, we know that for all formulae $\chi(x, y)$ of $\mathcal{L}_{\in}$, ZFC_{count} proves that

$$\Vdash_{Add(\omega,1)} \exists x \exists y(\chi(x, y) \leftrightarrow \chi(y, x)). \tag{2.1}$$

Moreover, by standard arguments we know that ZFC_{count} proves the following:

[26] Note that since there is no guarantee that $\mathcal{M}$ is well-founded, we cannot define $\mathcal{M}[G]$ using the standard Val function that is familiar from [Kunen, 2006] or [Shoenfield, 1971]. Rather, we define membership and identity in the model using the forcing relation. So, for example, given $\mathbb{P}$-names $\dot{x}$ and $\dot{y}$, we let $\dot{x} \in^{\mathcal{M}[G[} \dot{y}$ iff there is some $p \in G$ such that $p \Vdash \dot{x} \in \dot{y}$. See [Corazza, 2007] or [Maddy and Meadows, 2020] for more details.

[27] See [Visser, 2006] for more discussion of this and similar results.

(1) $\Vdash_{Add(\omega,1)} \varphi$, for all axioms φ of ZFC_{count};
(2) $\Vdash_{Add(\omega,1)}$ is closed under proof in first order logic (i.e., if $\Gamma \vdash \varphi$ and $\Vdash_{Add(\omega,1)} \gamma$ for all $\gamma \in \Gamma$, then $\Vdash_{Add(\omega,1)} \varphi$); and
(3) $\nVdash_{Add(\omega,1)} \perp$.

Note that each of these claims are proved in ZFC_{count} not our background theory PA. As such, we can pluck these results directly from the textbooks.[28] Using the fact that PA proves internal Σ^0_1-completeness[29] we see that ZFC_{count} proves that $ZFC_{count} \vdash \forall x \forall y(t(\psi)(x,y) \leftrightarrow \neg t(\psi)(y,x))$ and thus, we may use (2) and (1) to show that that ZFC_{count} proves

$$\Vdash_{Add(\omega,1)} \forall x \forall y(t(\psi)(x,y) \leftrightarrow \neg t(\psi)(y,x)).$$

Moreover, using 2.1, we see that

$$\Vdash_{Add(\omega,1)} \exists x \exists y(t(\psi)(x,y) \leftrightarrow t(\psi)(y,x))$$

and so ZFC_{count} proves that $\Vdash_{Add(\omega,1)} \perp$. This implies that ZFC_{count} is inconsistent. And since the equiconsistency of ZFC_{count} and SOA is clearly provable in PRA, we see that SOA is also inconsistent as required. □

Then since interpretations witnessing definitional equivalence must be direct, we see that:

Corollary 16. *(PA) ZFC_{count} and SOA are not definitionally equivalent, if one of those theories is consistent.*

Thus, we have a proof of the arithmetic claim that ZFC_{count} and SOA are not definitionally equivalent conducted in a standard theory of arithmetic, PA. One might hope that to obtain the result above from a weaker theory like PRA. However, the use of internal Σ^0_1-completeness in the proof above seems to tell us that, at least for this strategy, some form of induction is required in our metatheory.[30] As such, we leave open the question of whether PRA can prove that ZFC_{count} and SOA are not definitionally equivalent.

References

Gisela Ahlbrandt and Martin Ziegler. Quasi finitely axiomatizable totally categorical theories. *Annals of Pure and Applied Logic*, 30(1):63–82, 1986.

K.R. Apt and W. Marek. Second order arithmetic and related topics. *Annals of Mathematical Logic*, 6(3):177–229, 1974.

J.L. Bell. *Set Theory: Boolean-Valued Models and Independence Proofs*. Clarendon Press, 2005.

Tim Button and Sean P. Walsh. *Philosophy and Model Theory*. Oxford, UK: Oxford University Press, 2018.

Paul Corazza. Forcing with non-wellfounded models. *The Australasian Journal of Logic*, 5, 2007.

[28] The classic location for these results is [Kunen, 2006]. However, note that this is not the approach taken in the main text. Rather, it is Approach 2 described in Section VII.9 on page 234 of that book. Similar results in the context of Boolean algebras can be found in [Bell, 2005], however, note that we cannot use the Boolean algebra approach here since we are working in ZFC_{count} and without powerset we cannot form the completion of $Add(\omega, 1)$. Nonetheless, the proofs are very similar.

[29] See, for example, (BLiii) on page 17 of [Lindström, 2003].

[30] We've tried a few other, less elegant, approaches and wherever we've looked some induction seems to be required.

Ali Enayat. On the Leibniz-Mycielski axiom in set theory. *Fundamenta Mathematicae*, 181(3):215–231, 2004.

Ali Enayat. Variations on a Visserian theme. In Jan van Eijck, Rosalie Iemhoff, and Joost J. Joosten, editors, *Liber Amicorum Alberti: a tribute to Albert Visser*, pages 99–110. College Publications, 2016.

Ali Enayat and Mateusz Łełyk. Categoricity-like properties in the first order realm. *Journal for the Philosophy of Mathematics*, 1:63–98, Sep. 2024.

Victoria Gitman, Joel David Hamkins, and Thomas A. Johnstone. What is the theory ZFC without power set? *Mathematical Logic Quarterly*, 62(4-5):391–406, 2016.

Hans Halvorson. *The Logic in Philosophy of Science*. Cambridge and New York: Cambridge University Press, 2019.

David Hilbert and Paul Bernays. *Grundlagen der Mathematik*. J. Springer, Berlin,, 1934.

Wilfrid Hodges. *A Shorter Model Theory*. CUP, Cambridge, 1997.

Thomas Jech. *Set Theory*. Springer, Heidelberg, 2003.

T.J. Jech. *The Axiom of Choice*. Dover Books on Mathematics Series. Dover Publications, 2008.

A. Kanamori. *The Higher Infinite: Large Cardinals in Set Theory from Their Beginnings*. Springer, 2003.

Vladimir Kanovei and Vassily Lyubetsky. Notes on the equiconsistency of ZFC without the power set axiom and 2nd order pa, 2025. URL `https://arxiv.org/abs/2507.11643`.

Richard Kaye and Tin Lok Wong. On Interpretations of Arithmetic and Set Theory. *Notre Dame Journal of Formal Logic*, 48(4):497 – 510, 2007.

G. Kreisel. A survey of proof theory. *Journal of Symbolic Logic*, 33(3):321–388, 1968. doi: 10.2307/2270324.

Kenneth Kunen. *Set Theory: an introduction to independence proofs*. Elsevier, Sydney, 2006.

Azriel Lévy. *Definability in axiomatic set theory. I*, pages 127–151. North-Holland Publ Co, 1965.

P. Lindström. *Aspects of Incompleteness: Lecture Notes in Logic 10*. Lecture notes in logic. Taylor & Francis, 2003.

Penelope Maddy and Toby Meadows. A reconstruction of Steel's multiverse project. *Bulletin of Symbolic Logic*, 26(2):118–169, 2020.

Toby Meadows. Beyond linguistic interpretation. *Review of Symbolic Logic*, 2023. doi: 10.1017/S1755020323000321.

A. Mostowski. Formal system of analysis based on an infinitistic rule of proof. In *Proceedings of the Symposium on Foundations of Mathematics, Warsaw 2-9 September 1959*, pages 141–166. Pergamon, 1961.

A. Mostowski. Formal system of analysis based on an infinitistic rule of proof. In Andrzej Mostowski, editor, *Foundational Studies Selected Works*, volume 93 of *Studies in Logic and the Foundations of Mathematics*, pages 390–415. Elsevier, 1979.

Gerhard Osius. Categorical set theory: A characterization of the category of sets. *Journal of Pure and Applied Algebra*, 4(1):79 – 119, 1974.

Jonas Reitz. The ground axiom. *Journal of Symbolic Logic*, 72(4):1299–1317, 2007.

J. R. Shoenfield. Unramified forcing. In *Axiomatic Set Theory*, volume XIII of *Proceedings of Symposium in Pure Mathematics*, pages 357–381. 1971.

Stephen G. Simpson. *Subsystems of Second Order Arithmetic*. Springer, Berlin, 1999.

Albert Visser. Categories of theories and interpretations. In Ali Enayat, Iraj Kalantari, and Mojtaba Moniri, editors, *Logic in Tehran. Proceedings of the workshop and conference on Logic, Algebra and Arithmetic, held October 18-22, 2003*, volume 26, pages 284–341. ASL, Wellesley, Mass., 2006.

Albert Visser and Harvey M. Friedman. When bi-interpretability implies synonymy. *Logic Group preprint series*, 320, 2014.

W. Hugh Woodin. *The Continuum Hypothesis, the Generic Multiverse of Sets, and the Ω Conjecture*. Cambridge University Press, 2012.

W. Hugh Woodin, Jacob Davis, and Daniel Rodríguez. *The HOD Dichotomy*, pages 397–419. London Mathematical Society Lecture Note Series. Cambridge University Press, 2012.

P. Zbierski. Models for higher order arithmetics. *Bulletin de l'Académie Polonaise des Sciences - Série des Sciences Mathématiques, Astronomiques et Physiques*, 19(7):557, 1971.