Models As Universes

Brice Halimi

Abstract

Kreisel’s set-theoretic problem is the problem as to whether any logical consequence of ZFC is ensured to be true. Kreisel and Boolos both proposed an answer, taking “truth” to mean truth in the background set-theoretic universe. This article advocates another answer, which lies at the level of models of set theory, so that “truth” remains the usual semantic notion. The article is divided into three parts. It first analyzes Kreisel’s set-theoretic problem and proposes one way in which any model of set theory can be compared to a background universe and shown to contain “internal models.” It then defines logical consequence w.r.t. a model of ZFC, solves the model-scaled version of Kreisel’s set-theoretic problem and presents various further results bearing on internal models. Finally, internal models are presented as accessible worlds, leading to an “internal modal logic” in which internal reflection corresponds to modal reflexivity, and resplendency corresponds to modal axiom 4.

Keywords: logical validity · truth · informal rigour · Kreisel’s problem · logical consequence of ZFC · models of set theory · modal logic · recursively saturated structures

Georg Kreisel and George Boolos both raised the following problem: Given the language $L$ of first-order set theory, how can one be sure that any logically valid $L$-sentence is true? By itself, first-order logic does not seem to guarantee it at all. This problem is the “Valid Hence True” problem (hereafter VHT problem). Kreisel’s answer is positive and appeals to the completeness theorem for first-order logic. Boolos provides two answers, which resort to the reflection principle and to the completeness theorem, respectively. In both cases, Boolos proves logical validity to guarantee simple truth, and the proof relies on nontrivial reasons, but there is no reason after all why the truth being a consequence of logical validity should be immediate and obvious.

The VHT problem has been set up by Kreisel, and by Boolos as well, at the level of the background set-theoretic universe: Is any $L$-sentence that is logically valid (i.e., true in any structure contained in the universe)¹ true in the universe? But that way of setting up the VHT problem lays itself open to the following attack: Logical validity w.r.t. the universe makes perfect sense, but truth in the universe cannot be defined explicitly. (Indeed, Tarski’s semantics defines truth from recursive clauses giving the satisfaction conditions for complex formulas in terms of the satisfaction conditions for simpler formulas. If the variables of the object language range over the universe itself and thus are not restricted to a set, the implicit definition of satisfaction and truth cannot be converted into an explicit definition.) Shifting to the “model-scaled view,” i.e., the semantical view that considers only $L$-structures or models and excludes the background universe as a relevant object of study, the problem is reversed. Indeed, it makes perfect sense to say that an $L$-sentence is true in some $L$-structure, but it seems to make no sense at all to say that
that sentence is logically valid w.r.t. some $L$-structure. The predicament can be summarized in the following way:

<table>
<thead>
<tr>
<th>Kreisel-Boolos View</th>
<th>Model-Scaled View</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical Validity</td>
<td>OK</td>
</tr>
<tr>
<td>Truth</td>
<td>?</td>
</tr>
</tbody>
</table>

Ways of framing the VHT problem

In this paper, a close yet different problem will be addressed: *Is any logical consequence of ZFC (any first-order sentence that is true in every model of ZFC) ensured to be true?* This second problem will be referred to as “Kreisel’s Set-Theoretic problem” (hereafter KST problem). It presupposes a minimal commitment to ZFC, since any commitment to a theory of sets that is incompatible with ZFC would force a trivial negative answer to it.

As will presently appear, Kreisel’s and Boolos’ respective answers to the VHT problem can be adapted and completed so as to provide answers to the KST problem. In fact, the KST problem is a natural complement of the VHT problem. Remarkably enough, Kreisel and Boolos both formulate the VHT problem specifically about sentences of the language of first-order set theory. So they could and even should have raised the KST problem as well, because it makes as much sense, within the language of first-order set theory, to ask about logical consequences of set theory as to focus on logical truths. At any rate, both problems share a common ground. Actually, it could seem that the same predicament that afflicts the VHT problem will also afflict the KST problem: That truth in the universe is not accounted for by the usual semantics for models of ZFC, whereas logical consequence w.r.t. a model of ZFC seems difficult to conceive of.

My claim is that both assumptions are false, that truth in the universe can be formalized and that one can make good sense of logical consequence w.r.t. a given set model of ZFC. The latter point is not true of an $L$-structure in general. On that score, the VHT problem and the KST problem are not analogous. As a result, the second option about the KST problem, which consists in formulating it at the level of set models rather than at that of the background universe, is free from the main objection that plagued the VHT problem. I will argue that Kreisel’s and especially Boolos’ modified answers do provide a treatment of truth in the universe, but that, all things considered, the model-scaled option is better suited to examining and solving the KST problem than the one developed (albeit from two different perspectives) by both Kreisel and Boolos. Instead of thinking of the universe, at the onset, as being a kind of monster model, there is a natural way of presenting any model of ZFC as being a universe, so that the KST problem can be analyzed in a very precise manner, with the tools supplied by model theory. Furthermore, that option leads to further, more fine-grained results and, as will be seen, still allows giving an answer to the KST problem.

1 **Can the universe be conceived as a model?**

As it is well known, set theory has a very particular place within model theory, since the models of any given formal theory $T$ turn out to be set-theoretic structures, that is, members of a background set-theoretic universe. This holds in particular when the theory $T$ in point is the formal theory ZFC itself and, in that case, it is necessary to bear in mind the systematic replication between, on the one hand, such or such model of ZFC and, on the other hand, the background universe from which this model has been extracted, as any model of any formal theory.
To put it another way, models of a formal theory are members of a universe of sets which in turn can be seen as being itself a model, of course not of a formal theory, but rather of the informal set theory that one presupposes when doing mathematics. Since this universe may thus be described as the intended model of some set metatheory, any model of formal set theory is by principle very akin to it in some ways and constitutes, so to speak, a background universe in its own right. Hence, even though the distinction between a model of ZFC and the universe is perfectly clear, a connection remains, which in fact can be read both ways: The “true” universe can be conceived of, by extension, as a kind of “monster model,” just as any model can be seen as a kind of universe. The second way of looking at things will be explored soon. But, as already said, the first one has naturally given rise to the following question: What is the connection between truth in the “big” model, and truth in all the “small” models?

1.1 Kreisel

In an article,\(^2\) following which developed a whole current of reflection on the model-theoretic validity, Georg Kreisel introduces the following notions. For any first-order formula \(\alpha\),

\[
\text{Val} \alpha \overset{\text{def}}{=} \alpha \text{ is true in all structures whatsoever;}
\]

\[
V \alpha \overset{\text{def}}{=} \alpha \text{ is true in all structures whose domain is a member of the cumulative hierarchy (this corresponds to logical validity as it is taken in this paper, namely as truth in every structure).}
\]

If \(\alpha\) is a first-order formula with the symbol \(E\) of binary relation as its only nonlogical symbol,

\[
\alpha_\in \overset{\text{def}}{=} \alpha \text{ is true when the quantifiers in } \alpha \text{ range over all sets and } E \text{ is replaced by the “real” membership relation.}
\]

These notions come up quite naturally as soon as the logical validity of the sentence \(\alpha\) is understood as \(\alpha\)’s being “always true.” Indeed, Val \(\alpha\) is the direct expression of that idea; V \(\alpha\) is the mathematical enregimentation of it, which corresponds to the usual model-theoretic notion of truth in all the set structures; Finally, \(\alpha_\in\) means \(\alpha\)’s “universal truth,” not as truth in all the models of the set-theoretic universe, but as truth in the universe itself.

Now, it is also quite natural to expect that the different formulations of logical validity amount to the same thing. As to the question to know what relationship there is between V \(\alpha\) and Val \(\alpha\) for a sentence \(\alpha\) of the language of first-order logic, Kreisel begins by answering:

If \(\alpha\) is logically valid, then \(\alpha_\in\), i.e., (in symbols): Val \(\alpha \rightarrow \alpha_\in\). But one certainly does not conclude immediately: V \(\alpha \rightarrow \alpha_\in\); for \(\alpha_\in\) requires that \(\alpha\) be true in the structure consisting of all sets (with the membership relation); its universe is not a set at all. So V \(\alpha\) (\(\alpha\) is true in each set-theoretic structure) does not allow us to conclude \(\alpha_\in\) ‘immediately’ […].\(^3\)

Kreisel’s main objective is to settle, for first-order logic, the coextensivity of the intuitive notion of validity, Val, and of its set-theoretic counterpart (i.e., in the limits of the cumulative hierarchy), V. The equivalence of V with provability, that is, the (nontrivial) completeness theorem for first-order logic, is what makes the conclusion of Kreisel’s “squeezing argument” possible. Indeed, by virtue of the latter theorem, V \(\alpha\) implies D\(\alpha\) (the derivability of \(\alpha\) by means of the rules of first-order classical logic), and D\(\alpha \rightarrow \text{Val} \alpha\) may be accepted as a basic property of Val, so that V \(\alpha \rightarrow \text{Val} \alpha\), and finally V \(\alpha \leftrightarrow \text{Val} \alpha\), follow. On this account, logical validity of \(\alpha\) (V \(\alpha\)) entails Val \(\alpha\) and thus, by universal instantiation, the truth of \(\alpha\) as interpreted in the set-theoretic universe (\(\alpha_\in\)), which allows one to settle the VHT problem positively.

That solution can be carried over to the case of the KST problem: For any sentence \(\phi\) of \(L\), let ‘ZFC \(\vdash^+\phi\)’ be the relation that obtains when \(\phi\) is true in any set or class structure that “models” ZFC (this is the equivalent of Kreisel’s Val \(\phi\)). Then, ZFC \(\vdash \phi\) entails ZFC \(\vdash^+\phi\) (by completeness), which entails ZFC \(\vdash^+\phi_\in\), which entails in turn \(\phi_\in\): Problem solved. But there
are two difficulties which hinder that solution. The first one is that, while it may be a “basic” requirement of the notion of intuitive validity that derivability in pure first-order logic implies intuitive validity, it is not a basic requirement of intuitive validity any more that derivability in ZFC implies intuitive validity. Truths of pure first-order logic have a compelling character of their own, that makes it difficult not to consider them as true in whatever structure or universe of discourse, were it a class model or the background universe itself. But nothing of that kind occurs in the case of ZFC. There are various set theories after all and, for two such theories $T_1$ and $T_2$, nothing prevents models of $T_1$ from being objects of an informal model of $T_2$. For instance, one could want to study models of ZFC in a universe seen as a realization of Quine’s NF system. Being a logical consequence of ZFC, on that score, falls short of being logically valid and does not guarantee truth in the universe, unless the universe is postulated to be a model of ZFC.

There is in this respect a second difficulty, already present in Kreisel’s original solution, about the very status of truth in the universe. What appears clearly in the context of his argument, indeed, is that Kreisel is led to speak about the universe of all the sets as a structure, and to wonder if a logically valid sentence (or, in the modified version, a logical consequence of ZFC) will be true in this a little bit special model that is the universe of all sets. Now there are at least two reasons to question the very possibility to refer to truth of a sentence in the set-theoretic universe. The first reason comes from the “iterative conception” of set, that is, roughly, the conception according to which sets are all (but only) the objects reached by iterating the power set operation starting from the empty set. Echoing Zermelo’s position, William Tait raises the following point:

Contemporary set theorists frequently write informally as if $M_\Omega$ [the universe of all sets] were a model of set theory and, indeed, treat it as if it were a set except that for some mysterious reason it is not an element of the universe of sets. From their point of view, there is no difficulty with the notion of truth in $M_\Omega$ nor with the notion of a higher-order object, say a second-order class $A$: truth in $M_\Omega$ is just truth in a model and $A$ is just a subset of $M_\Omega$. When, as in the case of $M_\Omega$ itself, it is not, then it is called a proper class. But giving it a name does not really eliminate the mystery of why, when we treat it in all respects as a set, we nevertheless reject it as a set. [...] I think that, internal to the iterative conception, there is an explanation of why $M_\Omega$ cannot be regarded as a well-defined totality. But, accepting this point of view, the notion of truth in $M_\Omega$ requires explanation [...].

So the idea is clear: On pain of paradoxes, there is no universe of all sets forming a model of the axioms of set theory. The universe can be regarded only as a potential totality and, as a consequence, truth in the universe should not be regarded as determined for every sentence. Admittedly, the notion of proper class cannot be reduced to the idea that the universe can be regarded only as a potential totality. But, even though one is not willing to endorse the iterative conception, there is a more basic reason why one should not take truth in the universe for granted. Indeed, even though the universe is considered as a completed totality, truth in the universe cannot be handled exactly in the same way as truth in a given model, since, as a matter of principle, no formal semantics can underpin both kinds of truth, unless the universe is taken to be an actual model and plunged with all other models into some further background universe—but then, precisely, it would cease to be the universe.

As opposed to the two difficulties that affect Kreisel’s solution, there is in fact a structure in which all the sentences of the language of ZFC are ensured to have formalized truth conditions and in which all the sentences derivable in ZFC are ensured to be true: Namely, a model of ZFC. That will be the starting point of the solution proposed in this paper, the main problem of which
being to justify to see such a model as a genuine universe.

Admittedly, the transposition of Kreisel’s answer can be sharpened, so as not to simply presuppose the availability of the notion of “truth in V.” Indeed, owing to Mostowski theorem,\(^5\) one has that “If \(\phi\) is derivable (in pure first-order logic), then \(\phi\)” is derivable in Peano arithmetic, and thus in ZFC. In other words:

\[
\text{ZFC} \vdash \forall \phi \ (\phi \rightarrow \phi).
\]

But the completeness theorem for first-order logic is a theorem of ZFC:

\[
\text{ZFC} \vdash \forall \phi \ (\phi \rightarrow \phi).\]

So one gets:

\[
\text{ZFC} \vdash \forall \phi \ (\phi \rightarrow \phi).
\]

Hence the following schema: “If \(\phi\) is true in every model, then \(\phi\).” Requiring the notion of truth to lift from the schema to the corresponding generalization, one reaches the conclusion: Every sentence true in every model is true. (If this semantic ascent is a source of complaint, it is not a complaint against Kreisel’s answer, but a prohibition against the very possibility of phrasing his question.)

However, that result cannot be transposed to the case at stake because truth in every model of ZFC is less than truth in every model whatsoever. Thus the completeness theorem does not allow one to derive the following schema: “If \(\phi\) is true in every model of ZFC, then \(\phi\).” In fact, Löb’s theorem proves that, precisely, such a conditional can be derived only if \(\phi\) is already a theorem of ZFC:

\[
\text{ZFC} \vdash \forall \phi \ (\text{ZFC} \rightarrow \phi) \rightarrow \phi \implies \text{ZFC} \vdash \phi.
\]

Hence the conditional will miss out all the sentences that are not theorems of ZFC. Another solution has to be found.

### 1.2 Boolos

Let’s now turn to another well-known paper, from George Boolos, urging to consider the background set-theoretic universe as a model in which any first-order sentence should be evaluated. In “Nominalist Platonism,” Boolos remarks indeed that, oddly enough, logical validity of a sentence of L does not guarantee its truth, i.e., that it holds as interpreted in the whole universe of sets. Boolos is of course fully aware of the distinction that has to be drawn between the universe and the domains that it provides us with. Still, for logical validity to be synonymous with logical truth, it is required that, at least, validity implies truth, that is, truth in the intended model that the universe constitutes implicitly. Boolos’ problem amounts to the VHT problem and should not be confused with the KST problem. Indeed, the question asked by Boolos is not as to whether an L-sentence true in every model of ZFC is true in the background universe, but as to whether a first-order sentence true in every model whatsoever is true in the background universe:

\[
[\ldots] \text{suppose that some sentence } G \text{ of the language of set theory is logically valid, true in all models. What guarantee have we that } G \text{ is true, that is, true when its variables are taken as ranging over all the sets there are and } \in \text{ as applying to (arbitrary) } x, y \text{ if and only if } x \text{ is in } y?^6
\]

Boolos suggests two ways out of the difficulty: The completeness theorem and the reflection principle. Both can be used as such only as far as the VHT problem, not the KST problem, goes. The first way out comes close to Kreisel’s solution. In fact, dissatisfied with the need of the completeness theorem to prove what should be obvious, namely that validity entails truth, Boolos puts forward a new notion of validity, which he calls “supervalidity,” of which truth in the universe is an obvious consequence.\(^7\) The supervalidity of a sentence of set theory (of first
or second-order), as expressed by a monadic second-order sentence whose quantifiers are to be interpreted plurally, constitutes an apparent strengthening of logical validity, but supervalidity, in the first-order case, turns out to be extensionally equivalent with logical validity. Could Boolos’ introduction of supervalidity be transposed to the case of the KST problem? As Boolos himself acknowledges, there is no clear way of extending logical consequence into some notion of superconsequence, as logical validity has been extended into supervalidity. Defining the notion of being a superconsequence of ZFC does not seem to be an easy option. That path would require anyway shifting to second-order set theory or to some extension of it, and proving that the supposed second-order notion of being a superconsequence of ZFC and the first-order notion of being a logical consequence of ZFC collapse. That is why Boolos’ first suggestion will not be pursued beyond Kreisel’s modified answer.

Boolos’ second suggestion has thus to be embraced as the main one. It relies on the reflection principle, namely on the following schematic theorem of ZFC: For any formula \( \varphi(x_1, \ldots, x_k) \) of \( L \),

\[
ZFC \vdash \exists \beta (\text{Ord}(\beta) \land \forall x_1 \in V_\beta \ldots \forall x_k \in V_\beta (\varphi(x_1, \ldots, x_k) \leftrightarrow \varphi^\beta(x_1, \ldots, x_k)),
\]

where \( \varphi^\beta \) refers to the relativization of \( \varphi \) to \( V_\beta \). \(^{10}\) As a consequence, if some sentence \( \phi \) is false (in the universe), then \( \neg \phi \) is true, hence true in some \( V_\alpha \), and so \( \phi \) cannot be true in all models of ZFC. However, the reflection principle is about finite conjunctions of formulas only and does not ensure levels \( V_\alpha \) of the cumulative hierarchy that model ZFC (ZFC cannot prove its own consistency). So the falsity of a sentence \( \phi \) does not entail by reflection the existence of a model of \( \text{ZFC} + \neg \phi \). Hence Boolos’ second solution does not work either in the case of the KST problem.

Nevertheless, it can be extended, through the addition of a satisfaction predicate \( \text{Sat}(u, v) \) and of a truth predicate \( \text{Tr}(u) \) to the language of \( \text{ZFC} \), since a truth predicate is the natural device to refer to infinitely many sentences. Let’s show in some detail how that extension can be worked out.

Let \( V \) be some fixed countable set, whose elements are taken as codes for the variable symbols of \( L \). Besides, let \( \gamma \in \gamma, \gamma' \in \gamma, \gamma \land \gamma' \) and \( \gamma \lor \gamma' \) be five fixed sets, taken as a code for the signature of \( L \). The following set \( F \) is then defined by induction: \( F_0 = (\{ \gamma \} \times V^2) \cup (\{ \gamma' \} \times V^2) \), \( F_{n+1} = F_n \cup (\{ \gamma \} \times F_n) \cup (\{ \gamma' \} \times F_n) \cup (\{ \gamma \lor \gamma' \} \times V) \times F_n \). An element \( (\gamma'v_1, v_2) \) of \( \{ \gamma' \} \times V^2 \) is written \( 'v_1' \in 'v_2' \), and the same obvious convention is adopted for all elements of \( F \). It is thus not difficult to see that, for any formula \( \phi \) of \( L \), there is a unique element \( \gamma \phi \) of \( F \) corresponding to \( \phi \). Now, let \( \text{Form}(x) \) be defined as: \( x \in F \). For any \( x \in F \), the set \( \text{fv}(x) \) of the free variables of \( x \) is readily defined by induction. Then \( \text{Sent}(x) \) is defined as \( (\text{Form}(x) \land \text{fv}(x) = \emptyset) \). Furthermore, a recursively enumerable formula \( \text{Ax} \) can be built so that, for any \( x, \text{Ax}(x) \) if \( x = \gamma \phi \) for some axiom \( \phi \) of \( \text{ZFC} \). Besides, \( \text{Assign}(y) \) is defined as “\( y \) is a map with domain \( V \)”.

One can then add a 2-place symbol \( \text{Sat}(x, y) \) to \( L \), so that \( \text{Sat}(\gamma \phi, s) \) should hold when \( s \) is an assignment for the variables of \( L \) which satisfies \( \phi \) in \( V \). The formulas \( \phi \) for which \( \text{Sat}(\gamma \phi, s) \) could hold should be the original formulas of \( L \) not containing ‘Sat’, so that no paradox arises. The axioms for \( \text{Sat} \) and \( \text{Tr} \) are:

- \( \forall x \forall y (\text{Sat}(x, y) \rightarrow \text{Form}(x) \land \text{Assign}(y)); \)
- the usual inductive clauses for satisfaction: \( \text{Sat}(v_1 \gamma v_2, s) \leftrightarrow s(v_1) \in s(v_2), \text{Sat}(v_1 \gamma = v_2, s) \leftrightarrow s(v_1) = s(v_2), \text{Sat}(\gamma \neg u, s) \leftrightarrow \neg \text{Sat}(u, s), \text{Sat}(u \lor u', s) \leftrightarrow (\text{Sat}(u, s) \lor \text{Sat}(u', s)) \) and \( \text{Sat}(\gamma \exists v)u, s) \leftrightarrow \exists x \text{Sat}(u, s[x/s(v)]); \)
- \( \text{Tr}(u) \leftrightarrow (\text{Sent}(u) \land \forall y (\text{Assign}(y) \rightarrow \text{Sat}(u, y))). \)

Let \( S \) be the conjunction of all these axioms, and let \( \text{ZFCS} = \text{ZFC} + S \) be the resulting system in \( L^+ = L \cup \{ \text{Sat}, \text{Tr} \} \), where the replacement axiom and the separation axiom are extended to
include formulas in which ‘Sat’ or ‘Tr’ occurs. Besides, it is well-known that semantic notions about $L$ can be formalized within $L$.\footnote{Informal Trivialized by Löb’s theorem} This formalization readily extends to $L^+$. In particular, there is a formula $\Sigma(A, u, s)$ in $L^+$ to the effect that $A$ is a set structure for $L^+$, $u$ is $\tau \phi^n$ for some formula $\phi$ of $L^+$ and $\phi$ holds in $A$ under the assignment $s$. Accordingly, there is a formula $\Theta(A, u) = \tau A \models \sigma^n$ in $L^+$ to the effect that $A$ is a structure for $L^+$, $u$ is $\tau \sigma^n$ for some sentence $\sigma$ of $L^+$, and $A \models \sigma$. One thus deals with two truth predicates, Tr and $\Theta$. It is noteworthy that ZFCS $\vdash \forall u (\text{Ax}(u) \rightarrow \text{Tr}(u))$ and ZFCS $\vdash \forall A \forall u (\Theta(A, \tau \text{Tr}(u)) \rightarrow \Theta(A, u))$.

Now, the proof of the reflection principle for ZFC extends readily to ZFCS. It is indeed possible, given a formula $\psi(v_0, \ldots, v_n)$ of $L^+$, to build in $L^+$ a formula $\delta = F(s)$ expressing that $\delta$ is the least ordinal such that $V_\delta$ contains a witness $x$ for $v_0$ in $\psi$ under the assignment $s$ (namely a set $x$ for which $\psi$ is satisfied in $V_\delta$ when $x$ is assigned to $v_0$ and $s$ is used for the assignment to the remaining free variables of $\psi$), and 0 when there is no such witness $x$ at all. Then, for a given $\alpha$, the closure $\beta$ of $\alpha$ under $F$ can be defined in $L^+$ along the lines of the proof for ZFC, and for such $\beta$, $\forall_{v_1 \ldots} \forall_{v_n}(\psi \leftrightarrow \psi^V_{\beta})$ is provable in ZFCS. (Here, in the context of ZFCS, ‘$V_\beta$’ actually refers to the expansion $(V_\beta, \in, \text{Tr} \cap V_\beta)$ of $V_\beta$ to $L^+$.) In particular, for any given sentence $\phi$ of $L^+$ that is true in $V$, taking $\psi := (\phi \land \forall u (\text{Ax}(u) \rightarrow \text{Tr}(u)))$, one gets: ZFCS $\vdash \exists \beta(\forall u (\text{Ax}(u) \rightarrow \text{Tr}(u)) \land \phi)^V_{\beta}$. But\footnote{Point is that the proof above, of the existence of the required model $V_\beta$, is given in ZFCS while it cannot be given in ZFC, since it needs recourse to the predicate Tr. The theory ZFCS is significantly stronger than ZFC since, as just shown, it proves Con(ZFC); It is in fact conjectured to be equivalent to Morse-Kelley set theory (see below for the detail of the latter theory). That is the main shortcoming of Boolos’ modified solution. Indeed, one should argue just from within ZFC, in keeping with the spirit of reflection theorems, which are theorems of ZFC.} ZFCS $\vdash \exists \beta \exists V_\beta (\forall u (\text{Ax}(u) \rightarrow \text{Tr}(u)) \land \phi)^V_{\beta}$, and so ZFCS $\vdash \exists \beta(\forall u (\text{Ax}(u) \rightarrow \Theta(V_\beta, u)) \land (V_\beta, \text{Tr}))$. In other words, one has: ZFCS $\vdash \text{Tr}(\phi^n) \rightarrow \exists \beta \exists V_\beta \vdash ZFC + \phi^n$. Now, suppose that $\phi$ is not true. Then ZFCS proves that $\neg \phi$ is true and thus that $V_\beta \vdash ZFC + \neg \phi$ for some $\beta$, and so $\phi$ is not a logical consequence of ZFC. By contraposition, ZFCS proves any logical consequence of ZFC to be true (in the sense of ‘Tr’, which has been defined in $L^+$ but is not definable in $L$, owing to Tarski’s theorem on the undefinability of truth).

The above inductive characterization of the satisfaction relation Sat could be turned into an explicit definition, but such an operation, that would require second-order machinery, is not needed. In fact, Boolos’ paper precisely aims at providing, in terms of plural quantification, an equivalent of a second-order definition of the set of all true sentences of $L$.\footnote{The KST problem: Is any logical consequence of ZFC true?} The main point is that the proof above, of the existence of the required model $V_\beta$, is given in ZFCS while it cannot be given in ZFC, since it needs recourse to the predicate Tr. The theory ZFCS is significantly stronger than ZFC since, as just shown, it proves Con(ZFC); It is in fact conjectured to be equivalent to Morse-Kelley set theory (see below for the detail of the latter theory). That is the main shortcoming of Boolos’ modified solution. Indeed, one should argue just from within ZFC, in keeping with the spirit of reflection theorems, which are theorems of ZFC.

Here is a summary of the treatments of the KST problem drawn from Kreisel’s and Boolos’ original treatments of the VHT problem:

<table>
<thead>
<tr>
<th>Kreisel’s View</th>
<th>Modified</th>
<th>Boolos’ View</th>
<th>Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth in the universe</td>
<td>Informal</td>
<td>Formalized through a satisfaction predicate added to the language of ZFC</td>
<td></td>
</tr>
<tr>
<td>Answer to the KST problem</td>
<td>Trivialized by Löb’s theorem</td>
<td>Requires to shift to ZFCS, a proper extension of ZFC</td>
<td></td>
</tr>
</tbody>
</table>
orem trivializes the answer given to the KST problem, as set out above; Or one defines it by
truth conditions for semantic predicates added to the language, as Boolos does, but then this
option requires shifting to the significantly stronger theory ZFCS — for which a new version of
the KST problem will have to be faced and will again require one to resort to some even stronger
theory, and so on, which triggers an infinite regress. Indeed, Boolos’ modified second solution
relies on the fact that $\text{ZFC} \vdash \tau \phi \rightarrow \text{Tr}(\phi)$, or equivalently (by completeness) that
$\text{ZFCS} \models \tau \phi \rightarrow \text{Tr}(\phi)$. But then the question naturally arises as to whether such a
logical consequence of ZFCS is true itself. Even if that question is deemed not to be as crucial
an issue as the original one, because it does not pertain as directly to ZFC, it certainly has to be
faced as soon as one shifts from ZFC to ZFCS. The answer to the KST problem has just been
pushed back up a level.

The path advocated in this paper consists in defining truth through a formal semantics (as
opposed to Kreisel’s modified solution) and in answering the question raised about ZFC while
sticking to ZFC itself (as opposed to Boolos’ modified solution). The natural way to go to work
semantically within ZFC is to frame the KST problem at the level of models of ZFC, so that
any definition of truth in the universe becomes unnecessary. Obviously, the counterpart of that
option is the need to define what it means for a sentence of $L$ to be, relatively to some model of
ZFC, a logical consequence of ZFC. Instead of asking whether any logical consequence of ZFC is
true (in the universe), the question becomes whether any $M$-logical consequence of ZFC is true
in $M$, for some model $M$ of ZFC, and whether any sentence that is an $M$-logical consequence
for any $M$ is true in any $M$ (the notion of “$M$-logical consequence of ZFC” will be defined
soon).

Boolos and Kreisel considered two kinds of truth, truth$_1$ in a set structure and truth$_2$ in the
background universe, and asked about the connection between truth$_1$ in all structures and truth$_2$
in the universe. That way of dealing with the KST problem raises the difficulties that have just
been pointed out. Those difficulties are old ones: The “predicament” according to which one
presupposes a meaningful set theory to deal with models of the set-theoretic axioms, is, according
to Ignacio Jané, one of the main reasons why Skolem complained about axiomatized set
theory.¹⁴ The analysis presented here will consist, contrary to Kreisel and Boolos, in formu-
lating the KST problem in such a way that the notion of truth that occurs in the definition of
being a logical consequence of ZFC, as truth in any structure for the language, is the same as
that about which it is asked whether or not it is ensured by being a logical consequence of ZFC.
It is indeed clearer to deal with only one kind of truth, as clearly laid down by the usual rules of
Tarskian semantics. Above all, it allows one to answer a question about ZFC while remaining
within the limits of ZFC. As what follows will show, it is in fact possible to turn any model of
ZFC into a universe, and thus to do only with models and model-theoretic truth.

Before turning to that point, a caveat is in order about the relativism that could be suspected
to underpin the present perspective. Speaking of models of ZFC as different “universes” and
describing truth and logical consequence as relative to some “universe” does not preclude the
existence of some background absolute universe whose objects include all the models of ZFC.
Moreover, as will presently appear, the distinction between standard and nonstandard models
will be critical, yet it can established only from the external vantage point of some universe.
However, the recognition of a single absolute background universe does not commit one to a
full theory of absolute truth in the set-theoretic universe. Admittedly, the fact that a given model
of ZFC satisfies such and such sentence, or that a given model is nonstandard, constitutes a truth
in the background universe; Thus, model theory of ZFC requires to make sense of the truth of
semantic statements such as “$M \vDash \theta$,” “the model $M$ is nonstandard,” “For any model $M$ of
ZFC, $M \vDash \theta$,” or “For any sentence $\theta$ of the theory $T$, $M \vDash \theta$.” But a general formal theory of
truth in the background universe is not necessary to that end.
On the contrary, taken at the level of the background universe, as the VHT problem is construed by Kreisel and Boolos, the KST problem does not admit of any clear answer if the truth conditions in the universe of any sentence of the language have not first been formally defined (were it explicitly or implicitly). But then working within ZFC is not enough: Truth formalized through a semantic predicate does guarantee truth in some model of ZFC (and thus truth in all models of ZFC does guarantee truth), but on the condition of shifting to a proper extension of ZFC. On the contrary, confining oneself to the model theory of ZFC avoids any recourse to a more robust background theory. Admittedly, the KST problem then ceases to be addressed as it was originally formulated, to be replaced by its model-relativization (truth becoming truth in some model of ZFC). But if one insists upon addressing the original KST problem, then Kreisel’s and Boolos’ modified solutions show that it becomes impossible to provide a fully satisfying answer. To get a fully satisfying analysis of the KST problem, keeping within ZFC, one has to allow the model-scaled construal of the problem to supersede its original formulation.

As just said, this option does not preclude the vantage point of view of the background universe: This is quite the opposite, simply because saying that a sentence is true in a model of ZFC remains a fact in the background universe. But that does not require formalizing truth in the universe either. On the other hand, the model-scaled construal of the KST problem is actually compatible with the “multiverse view” in set theory, but does not force its endorsement either. The multiverse view holds that there are a multitude of set-theoretic universes, each of which embodies a concept of set and a set-theoretic truth of its own, so that set-theoretic truth is irreducibly relative to the universe in which one happens to stand. That view may be rejected if some background universe is presumed as absolute. Nevertheless, the multiverse view can be adopted as a theoretic framework to harness the resources of model theory (a subtheory of ZFC) and to formalize the notion of set-theoretic point of view as embodied by models of ZFC. So the present perspective is methodologically akin to the multiverse view, yet not philosophically committed to any multiverse realism.

It simply takes literally the idea, expressed by Joel David Hamkins as underpinning the multiverse view, that set theory has become model theory of set theory. It is not about giving up the idea of truth as ultimately truth-in-the-background-universe (the universe is not counted among the variable possible universes of discourse in which explicitly evaluate sentences of the language of ZFC), but, essentially, about making it possible to analyze the KST problem about ZFC within the limits of ZFC itself. Moreover, such an approach reaches further and more calibrated results, which provide more control over the treatment of the problem, while still permitting a definite answer. That is why it will be argued that looking at models of ZFC as universes allows one to set as well as to settle the KST problem in a new and more satisfying way than Kreisel and Boolos did in the case of the VHT problem. But how it is possible to define logical consequence of ZFC w.r.t. each model of ZFC remains to be explained. What follows below is the explanation.

### 1.3 Internal models

As we shall see, any model of ZFC can be shown to contain, as an element of its domain, in a sense to be precised, another model of ZFC. In a way, this should come as no surprise. Indeed, the formal set theory ZFC is strong enough to express the essentials of mathematics, and so any model of ZFC should be able to achieve the representation of any mathematical reality, a model of ZFC included. This “replication” phenomenon has now to be set out precisely.

Let $V$ be the background set-theoretic universe and $\in$ the membership relation between the objects of $V$. At the risk of confusion, ‘$\in$’ will also denote the membership symbol in the first-order language $L$ of ZFC. A (set) model $M$ of ZFC is a set $|M|$ (i.e., an object of the universe) endowed with a relation $\in_M$ such that $M = (|M|, \in_M)$ satisfies the axioms of ZFC.
but whose interpretation of membership has not to coincide with the restriction of $\in$ to $M$. In case $\in_M \simeq \in \upharpoonright |M|$, $M$ is said to be a standard model.\footnote{17}

As recalled about Boolos’ modified view, the main notions in the metatheory of ZFC can be formalized within $L$, through some usual Gödel numbering.\footnote{18} The code of any formula $\phi$ of $L$ consists then in a sequence $\Gamma \phi \upharpoonright \gamma$ of numerals and gives rise in any model $M$ of ZFC to an interpretation $\Gamma \phi^M$, where each numeral of the sequence is interpreted by the corresponding integer of $M$. A finite sequence (in the sense of $M$) of integers of $M$ such as $\Gamma \phi^M$ is called an $M$-formula. It is then possible to define in $L$ the predicate ‘$\text{For}(x)$’ to the effect that $x$ encodes the construction of a formula of $L$, and the relation ‘Dem$(y, x)$’ to the effect that $y$ encodes a ZFC-proof of the item encoded by $x$. An $M$-formula is an object $x$ in $M$ such that $M \models \text{For}(x)$, and an $M$-proof is an object $y$ in $M$ such that $M \models \exists x (\text{For}(x) \land \text{Dem}(y, x))$. Moreover, any statement $S$ falling to the basic semantics of $L$, such as $'N \models \phi[s]'$, can likewise be coded into a sentence $\Gamma S \upharpoonright \gamma$ of $L$, here $\Gamma N \models \phi[s]$.\footnote{19}

A given model $M$ is called $\omega$-standard if $\in_M$ is transitive and well-orders all the finite ordinals of $M$. Then, for any $a \in |M|$, $M \models "a$ is a finite ordinal” implies that there is no sequence $(a_i)_{i \in \omega}$ of elements of $|M|$ such that $\forall i \in \omega, a_{i+1} \in_M a_i$ with $a_0 = a$. (This is in particular the case if $M$ is standard.) The integers of a $\omega$-standard model of ZFC are then isomorphic to the genuine integers of the universe. On the contrary, a model $M$ of ZFC is called non-$\omega$-standard if it admits a nonstandard integer, that is, if there is in $M$ an infinite set (from the point of view of the universe) that yet $M$ recognizes as being a finite ordinal. The existence of non-$\omega$-standard models is a direct consequence of the compactness theorem for first-order logic.

If $M$ is $\omega$-standard, the $M$-formulas (resp. the $M$-proofs) are in a 1-1 correspondence with the genuine formulas (resp. the proofs) of ZFC. If not, some $M$-formulas and $M$-proofs fail to correspond to any formula or proof of ZFC. Indeed, for any formula $\phi$ of $L$, there is a unique element $x = \Gamma \phi \upharpoonright \gamma$ of $F$ corresponding to $\phi$ and thus a unique corresponding $M$-formula $x^M$, but the converse is not true in case of an $M$-formula whose length is a nonstandard integer. At any rate, the extension of the provable formulas grows as one switches from the universe $V$ to models of ZFC. It is always possible that an element $N$ of $M$ which (as a set) is not a model of ZFC, is still recognized by $M$ as being such.

Now, let $M$ be a given model of ZFC and $N$ an element of $|M|$ such that $M \models \Gamma N$ is a structure for $L$. This implies that there exists $|N|, E^N \in |M|$ such that $M \models N = \langle |N|, E^N \rangle \land E^N \subseteq |N| \times |N|$. One then defines $|N_M| := \{x \in |M| : M \models x \in |N|\}$ and $E^N_M := \{(x, y) \in |N_M| \times |N_M| : M \models (x, y) \in E^N\}$. The structure $N_M := \langle |N_M|, E^N_M \rangle$ is called the replica of $N$ in $M$. In case $M = \langle |M|, \in_M \rangle$ is a transitive $\in$-model of ZFC (i.e., $\in_M = \in \upharpoonright |M|$ and, $\forall x \in |M|, x \subseteq M$), the replica $N_M$ of any $N$ in $M$ is isomorphic to $N$.

**Lemma 1.1.** For any sentence $\phi$ of $L$ and any model $M$ of ZFC, one has:

\[ M \models \Gamma N \models \phi \Leftrightarrow M \models \phi^M. \]

**Proof.** The proof is by induction on $\phi$, which is possible since, by hypothesis, $\phi$ is a genuine formula of $L$, so that its interpretation in $M$ is not a $M$-pseudoformula (of infinite complexity). Besides, $'N_M \models \phi'$ and $'M \models \Gamma N \models \phi^M'$ verify exactly the same recursive clauses. This suffices to conclude. (Let’s remark that even if $N$ fails to be a structure for $L$, $N_M$ will be such a structure anyway.) \qed

**Theorem 1.2** ([Suzuki and Wilmers, 1971], [Schlipf, 1978]). Let $M$ be a model of ZFC. Then there exists $N \in |M|$ such that $N_M \models ZFC$ (but not necessarily: $M \models \Gamma N \models ZFC^\tau$).

**Proof.** Let’s distinguish two cases.
First case. $M$ is $\omega$-standard. In this case, all the formulas of $L$ and all the proofs within $\text{ZFC}$ may be coded into elements of $M$ in a transparent way. Since by hypothesis $M$ is $\omega$-standard, there cannot be any $M$-proof which would not code a real proof. This holds in particular for all $\text{ZFC}$-proofs. Now let’s exploit the assumption that there exists a model $M$ of $\text{ZFC}$. The very existence of $M$ implies that $\text{ZFC}$ is consistent. Hence $V \models \text{Con}(\text{ZFC})$. This means that there is no proof in $\text{ZFC}$ of ‘0 = 1’. For the reason which has just been mentioned, there are no more proofs according to $M$ than there are in reality. Therefore $M$ does not acknowledge any proof of ‘0 = 1’ either, that is, $M \nvDash \text{Con}(\text{ZFC})$. Since the (formalized version of) the completeness theorem for first-order logic is a theorem of $\text{ZFC}$, it can be deduced that $M \models \text{there exists a model of } \text{ZFC}$. 

So there is $N \in |M|$ such that $M \models \forall n \exists \alpha \forall V \alpha \models A_0 \land A_1 \land \ldots \land A_n \land \ldots$. Now, the formula ‘$\exists \alpha \forall V \alpha \models A_0 \land A_1 \land \ldots \land A_n$’ is a formula $\chi(n)$ of $L$. One has that $M \models \exists n \chi(n)$, thus $M \models \{ n \in \omega : \neg \chi(n) \} \neq \emptyset$. Besides, $M \models \text{“Every nonempty subset of } \omega \text{ has a least element”}$ (since it is also a theorem of $\text{ZFC}$). So there exists $n_0 \in \omega^M$ such that $M \models \forall n < n_0 \, \chi(n)$. On the other hand, $M \models A_0 \land \ldots \land A_{n_0}$ for each integer $n$, so, owing to the reflection principle (also true in $M$), $M \models \chi(n)$ for each standard integer $n \in \omega$. Hence $n_0$ is necessarily a nonstandard integer of $M$. But $M \models \chi(n_0 - 1)$, in other words $M \models \exists \alpha \forall V \alpha \models A_0 \land \ldots \land A_{n_0 - 1}$, and since $n_0 - 1$ is also nonstandard, there finally exists an ordinal $\alpha$ of $M$ such that $(V^M_\alpha)_M \models \text{ZFC}$ (the notation $(V^M_\alpha)_M$ here refers to the replica in $M$ of the rank $(V_\alpha)^M$ of the cumulative hierarchy internal to $M$). This does not mean that $M \models \forall V \alpha \models \text{ZFC}$. As a matter of fact, from the point of view of $M$, $V^M_\alpha$ only satisfies a finite number of axioms of $\text{ZFC}$: It is only from an external point of view that $n_0$ turns out to be infinite. Its nonstandard nature causes $M$ to think that the actual model $N$ of $\text{ZFC}$ that it contains fails to satisfy some part of $\text{ZFC}$. In comparison to lemma 1.1, it appears that $M$ thinks of every axiom of $\text{ZFC}$ that $N$ satisfies it, and yet does not think that $N$ is a model of $\text{ZFC}$.

Gathering the standard case and the nonstandard one gives the following result: Any model $M$ of $\text{ZFC}$ contains an element whose replica in $M$ is a model of $\text{ZFC}$.

The upshot of this result is that any model $M$ contains an internal model $N$ in which all the axioms of $\text{ZFC}$ are true, even though the statement that all the axioms of $\text{ZFC}$ are true in $N$ may be false in $M$. This is what happens when the axioms of $\text{ZFC}$ that are true in $N$ are indexed in $M$ by a nonstandard integer: The number of those axioms is infinite, yet finite as viewed from within $M$.

2 Models conceived as universes

We are now in a position to think of any model of $\text{ZFC}$ as being a particular universe. Indeed, we shall call internal model of $\text{ZFC}$ any model of $\text{ZFC}$ of the form $N_M$, where $M$ is a model of $\text{ZFC}$ that is given beforehand. The previous result ensures that any model $M$ of $\text{ZFC}$ has internal models. Hence it becomes possible to define logical consequence from $\text{ZFC}$ w.r.t. any given model $M$ of $\text{ZFC}$, and thus to tackle the KST problem at the level of models of $\text{ZFC}$. Before pursuing this, it should be explained how seeing a model of $\text{ZFC}$ as a universe is in line with a natural way of looking at models of set theory.
2.1 The idea of interpretational point of view

Within the range of all models of ZFC, two models ought to be singled out as seemingly resisting the existence of internal models: Shepherdson’s minimal model $M_0$ of ZFC, on the one hand, and any model $M^*$ of ZFC + $\neg$Con(ZFC), on the other hand. (In virtue of Gödel’s second incompleteness theorem, Con(ZFC) cannot be proved within ZFC itself, so that ZFC + $\neg$Con(ZFC) is consistent and therefore has a model.) In the first case, there are indeed models of ZFC internal to $M_0$ (sets within $M_0$ which are isomorphic with their replica in $M_0$, and models of ZFC), but those models are all nonstandard, and $M_0$ faithfully recognizes that they are both models of ZFC and nonstandard. In the second case, $M^*$ is doomed to be non-$\omega$-standard, because otherwise ZFC would be able to derive its own inconsistency. The model $M^*$ cannot acknowledge the existence of any model of ZFC, since it precisely asserts that there cannot be any, nevertheless it contains elements $N^*$ such that each replica $N^*_M$ is actually a model of ZFC. It states that any of its internal structures $N^*$ satisfies at most a finite number of the axioms of ZFC (or that there is some finite ordinal $n$ such that the replacement schema restricted to $\Sigma_n$-formulas is inconsistent), but this number is nonstandard, so that in fact, viewed from outside of $M^*$, $N^*_M$ satisfies the whole theory ZFC.

Such a conclusion obviously involves the absolute point of view of the real universe. But, as already said, the presupposition of the background universe is integral to the perspective developed in this paper, simply as the semantic counterpart of the fact that the analysis is kept within the limits of ZFC: One has to deal only with objects of the background universe, as one deals only with statements which are derivable in ZFC. The notion of point of view itself corresponds to an actual set-theoretic operation, namely $(M, N \in |M|) \mapsto N_M$.

The examples of $M_0$ and of $M^*$ justify considering, more generally, any model of set theory not only as a structure, that is, as a domain to evaluate formal sentences, but also as a point of view, that is, as a structure constituting a background universe on its own, as including models of formal theories and establishing a specific satisfaction relation between them and formulas.

A sentence such as ‘$\exists f \forall x f(x) = x$’ may serve as a first example to understand the difference between these two levels: This sentence is true about any model, and as such true from the point of view of the universe, but it cannot be true in the universe itself, because the map $f$ purporting to exist cannot be but a proper class. Another example, due to Vann McGee, is provided by the quantifier $\exists^{AI}$, where ‘$(\exists^{AI} x) (\phi(x))$’ means “the individuals satisfying ‘$\phi(x)$’ are too many to form a set.” Then ‘$(\exists^{AI} x) (x = x)$’ is true in the universe, and still false in any set structure for the language. Both examples substantiate the principle of a distinction between what holds in the universe and what holds from its point of view.

The same distinction can be made about models of set theory. Viewing models as “points of view” is not in the least contrary to standard set theory, but catches up with a well-established tradition dating back to Skolem’s paradox. The concept of point of view has notably been brought up by Ignacio Jané in his paper about Skolem. As Jané points out, Skolem himself speaks of set-theoretic notions (membership, being a binary relation, being a function, and so forth) “in the sense of the axiomatization,” which is to be understood as “in the sense of the model” that is taken to interpret the axioms. In particular, any member $a$ of a model $M$ of ZFC gives rise to the set $a^* = \{ x \in |M| : x \in_M a \}$. The set $a^*$ (in $V$) is nothing but $a$ as seen from the point of view of $M$, even though $a^*$ does not necessarily belong to $M$. The relativity phenomenon in which Skolem’s paradox is grounded is “the discrepancy between $M$’s assessment of $a$ and $a$’s (or rather, $a^*$’s) true status.”

In the present setting, any model $M$ of ZFC can be identified as a structure with the set of all sentences of the language of ZFC which turn out to be true when the quantifiers that they contain are restricted to the domain of $M$, whereas the point of view of $M$ consists in the re-
interpretation of all model-theoretic notions which that model builds up within its domain. Of course anything pertaining to $M$ is entirely determined by the extension of $M$’s interpretation of the membership relation. So identifying the structure of $M$ with that interpretation would result in reducing its point of view to its structure. What constitutes the right determination of the “point of view” embodied by a model is open to discussion, but has to remain distinct from its structure. To set things down, let’s say that “point of view” embodied by a model is open to discussion, but has to remain distinct from its structure. So identifying the structure of $M$ course anything pertaining to interpretation of all model-theoretic notions which that model builds up within its domain. Of $ZFC$ $ZFC$ $\text{Con}(ZFC)$ of $ZFC$ $\phi$ of all the conditions of the form ‘$N \models \Gamma$’ that are realized, where $\Gamma$ is any set of sentences. It could be argued that no sentence $\phi$ can be said to hold from the point of view of a model $M^*$ of $ZFC + \neg \text{Con}(ZFC)$, even if $\phi$ is true in $M^*$: According to $M^*$, there is no model of $ZFC$ in which $\phi$ is true, because there is no model of $ZFC$ at all. But that is not right, since there are internal models $N_M$, of $M^*$ (in which $\phi$ may be true), even though $M^*$ fails to recognize any $N \in |M^*|$ as a model of $ZFC$.

To sum up, while Kreisel and Boolos referred to the universe as being by extension a kind of model, it appears that it is also possible to look at any model of $ZFC$ as being a surrogate universe which contains models of $ZFC$ or, more precisely, from the point of view of which other structures appear to be models of $ZFC$.

### 2.2 Depth of logical consequence

A last clarification, which relies precisely on the notion of point of view, should be useful to dispel any appearance of paradox that the preceding result may arouse. As a matter of fact, it could seem that the previous theorem can again be applied to any model internal to the original model $M$ of $ZFC$, and again, so that the axiom of foundation is eventually violated. In fact, this is not the case, because the internal model $N_M$ does not coincide with the element $N$ of the domain of $M$. More precisely, one knows, by the previous result, that there is $M_1 \in |M|$ such that $M_1 := (M_1)_M \models ZFC$. So the next internal model will be internal to $M_1$, not to $M_0$. It will be a model $(M_2)_{M_1}$ with $M_2$ belonging to $M_1$, but not necessarily to $M_1$, so that any infinite descending $\in$-chain $\ldots |M_2| \in |M_1| \in |M|$ is avoided in the end. The structure $M_1$ is a model of $ZFC$ so long as one endorses the point of view of $M_1$ that it is a model of $ZFC$ (i.e., $M \models \Gamma(M_1) \equiv ZFC \cap \Gamma$). Otherwise put, $M_1$ as seen from $M$ is a model of $ZFC$, according to the universe, but $M_1$ as seen from the universe is not necessarily, according to $M$, a model of $ZFC$. It is not true that each model $M$ of $ZFC$ contains models of $ZFC$: It is rather that any model contains objects which, viewed from its point of view, are models of $ZFC$.

At this point, it is quite natural to put forth the idea of validity depth.

**Definition 2.1.** An $L$-sentence $\phi$ is a 2-logical consequence of $ZFC$ iff, for any $M \models ZFC$ and any $N \in |M|$, $N_M \models ZFC$ implies $N_M \models \phi$.

Actually, 2-logical consequences and logical consequences of $ZFC$ turn out to collapse.

**Theorem 2.2** ([Ressayre, 1983], 3.3). Let $N$ be a model of $ZFC$. Then there is a model $M \equiv N$ such that $\exists \Omega \in |M| \forall a \in \Omega, M \models a \models \Omega_M$, where ‘$M \models a \models \Omega_M$’ means that there is an isomorphism $f : M \rightarrow \Omega_M$ such that $\forall x \in |M| \cap a, f(x) = x$.

**Corollary 2.3.** Let $\phi$ be a sentence of $L$. Then $\phi$ is a 2-logical consequence of $ZFC$ iff it is a logical consequence of $ZFC$.

**Proof.** Any logical consequence of $ZFC$ is by definition a 2-logical consequence of $ZFC$. Conversely, suppose $\phi$ is a 2-logical consequence of $ZFC$, and let $N$ be any model of $ZFC$. By the
theorem 2.2, there exists $M \equiv N$, $\Omega \in |M|$ and $a \in \Omega$ such that $M \models_\Omega \Omega_M$. So, in particular, one has $M \equiv \Omega_M$. Since by hypothesis $\Omega_M \models \phi$, one finally gets $N \models \phi$, and this holds for any $N \models \text{ZFC}$.

An alternative definition would consist in identifying a 2-logical consequence* of ZFC with an $L$-sentence which is true in any model of ZFC belonging to (the domain of) any model of ZFC. But in fact the previous equivalence would remain true. Indeed, any logical consequence of ZFC is a 2-logical consequence* of ZFC. Conversely, suppose $\phi$ is a 2-logical consequence* of ZFC, and let $M$ be a model of ZFC. The first-order theory $\text{Th}(M)$ of $M$ in $L$ is consistent, and so admits a countable recursively saturated model $M'$. By a theorem of Schlipf, $M'$ belongs to some model $M$ of ZFC, and so by hypothesis $M' \models \phi$, so $\phi \in \text{Th}(M)$, i.e., $M \models \phi$. Hence $\phi$ is true in any model of ZFC.

### 2.3 Logical consequence and internal logical consequence

The framework that has been set out so far naturally leads to an examination of the KST problem at the level of models of ZFC. On the model-theoretic conception of logical validity, the question of whether a particular inference is valid seems indeed to depend on facts concerning the background universe of set theory (namely, on what models happen to exist). It is then a natural move to consider this issue by looking at one of its first-order analogs, that is, to examine the situation when a model of set theory replaces the universe and to explore how the resulting consequence relation changes (or does not change) as one moves from one model of set theory to another. In particular, since being a logical consequence of ZFC amounts to being true in every model of ZFC (viewed as a structure), a natural question is: What would it mean for a sentence to be a logical consequence of ZFC from the point of view of every model of ZFC?

**Definition 2.4** ($M$-logical consequence of ZFC). Let $\phi$ be an $L$-sentence and $M$ be a model of ZFC. Then $\phi$ is called an $M$-logical consequence of ZFC, written $\text{ZFC} \models_M \phi$, iff for every $N \in |M|$, $N_M \models \text{ZFC}$ implies $N_M \models \phi$.

**Definition 2.5** (internal logical consequence of ZFC). Let $\phi$ be an $L$-sentence. Then $\phi$ is called an internal logical consequence of ZFC, written $\text{ZFC} \models^i \phi$, iff $\text{ZFC} \models_M \phi$ for any model $M$ of ZFC.

The intuitive meaning of $\text{ZFC} \models_M \phi$ is that $\phi$ would be a logical consequence of ZFC were $M$ the background universe. The intuitive meaning of $\text{ZFC} \models^i \phi$, then, is that $\phi$ is a logical consequence of ZFC from all the possible points of view—the notion of a point of view being understood in the sense of the index status of $M$ in $\text{ZFC} \models_M \phi$, i.e., with respect to the local prism of the universe that constitutes such and such model $M$ of ZFC. Now, a few things naturally deserve to be studied.

The first one bears on the relation between $\text{ZFC} \models_M \phi$ and $M \models \phi$. One may take as an example the natural model of ZFC that is $\text{V}_{\theta}$, where $\theta$ is the first strongly inaccessible ordinal. In this case one knows, by a result of Montague and Vaught, there is an ordinal $\theta^* < \theta$ such that $\langle \text{V}_{\theta^*}, \in \rangle \equiv \langle \text{V}_\theta, \in \rangle$, with $\langle \text{V}_{\theta^*}, \in \rangle \in \text{V}_\theta$. Let’s suppose $\text{ZFC} \models_{\text{V}_\theta} \phi$. Then, by definition, $\langle \text{V}_{\theta^*}, \in \rangle \models_{\text{V}_\theta} \phi$. But $\text{V}_{\theta}$ is a transitive $\in$-structure, and so $\langle \text{V}_{\theta^*}, \in \rangle_{\text{V}_\theta} \models \phi$. Consequently, $\text{V}_{\theta^*} \models \phi$, and finally $\text{V}_{\theta} \models \phi$. Hence $\text{ZFC} \models_{\text{V}_{\theta}} \phi$ implies $\phi$, for any $L$-sentence $\phi$.

Now let’s call a cardinal $\gamma$ a universe cardinal iff $\text{V}_{\gamma} \models \text{ZFC}$, and let $\gamma_0$ be the least universe cardinal. The weak axiom of universes is the sentence WAU of $L$ saying that “there are unboundedly many universe cardinals.” It is a standard result that for any inaccessible cardinal $\kappa$ one gets: $\text{V}_{\kappa} \models \text{ZFC} + \text{WAU}$. But of course, by minimality, $\text{V}_{\gamma_0} \not\models \text{WAU}$, so that in fact (because...
\( \gamma_0 < \kappa \) \( \forall \gamma \in V_\kappa \) and \( \forall \gamma \forall \gamma_0 \not\in \text{WAU} \), resulting in \( \forall \gamma \forall \gamma_0 \Downarrow \text{ZFC + \text{WAU}} \). Consequently, \( M \models \phi \) does not entail \( \text{ZFC} \models M \phi \).

The converse general question, as to whether \( \text{ZFC} \models M \phi \) entails \( M \models \phi \), receives a positive answer.

**Theorem 2.6.** Let \( \phi \) be an \( L \)-sentence and \( M \) be a model of \( \text{ZFC} \) such that \( \text{ZFC} \models M \phi \). Then \( M \models \phi \).

**Proof.** Let’s suppose that \( M \models \lnot \phi \). This proves that \( \text{ZFC + \lnot \phi} \) is consistent. The proof of the theorem 1.2 can then be rewritten, with \( \text{ZFC + \lnot \phi} \) replacing \( \text{ZFC} \). One thus concludes that there exists \( N \in |M| \) such that \( N \models \text{ZFC + \lnot \phi} \), hence that \( \text{ZFC} \not\models M \phi \). Accordingly, \( \text{ZFC} \models M \phi \) implies \( M \models \phi \). \( \Box \)

As noticed, the two conditions \( \text{ZFC} \models M \phi \) and \( M \models \phi \) are not equivalent in general. This fact is still compatible with \( \text{ZFC} \models M \phi \) for each \( M \) iff \( M \models \phi \) for each \( M \), as what follows proves.

**Corollary 2.7.** Let \( \phi \) be an \( L \)-sentence. Then: \( \text{ZFC} \models \lnot \phi \) iff \( \text{ZFC} \models i \phi \).

**Proof.** By generalization over \( M \), the previous theorem guarantees that \( \text{ZFC} \models i \phi \) implies \( \text{ZFC} \models \phi \). Conversely, suppose that \( \phi \) is a logical consequence of \( \text{ZFC} \). Then, in particular, \( N \models \phi \) for any internal model \( N \) of \( \text{ZFC} \), so, by definition, \( \text{ZFC} \models M \phi \), and this holds for any model \( M \) of \( \text{ZFC} \).

We are now in a position to get back to the KST problem. Set at the level of models of \( \text{ZFC} \), that problem consists in the two following questions: Is any \( L \)-sentence true in \( M \) if an \( M \)-logical consequence of \( \text{ZFC} \)? Is any \( L \)-sentence true in all models of \( \text{ZFC} \) if an internal logical consequence of \( \text{ZFC} \)? Both answers are positive, as mentioned below, along with Kreisel’s answer and Boolos’ answer:

<table>
<thead>
<tr>
<th>Kreisel’s Modified View</th>
<th>Boolos’ Modified View</th>
<th>Model-Scaled View</th>
<th>Generalization to every ( M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi ) is a logical consequence of ( \text{ZFC} )</td>
<td>( \text{ZFC} \models i \phi )</td>
<td>( \text{ZFC} \models \phi )</td>
<td>( \text{ZFC} \models i \phi )</td>
</tr>
<tr>
<td>( \phi ) is true</td>
<td>( \phi ) is formally true</td>
<td>( \phi ) is true in the extension of the truth predicate added to ( L )</td>
<td>( \phi ) is true in every ( M )</td>
</tr>
<tr>
<td>Answer to the KST question</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes (equivalence)</td>
</tr>
</tbody>
</table>

Answers to the KST problem

The last column of the table above is but the generalization to every \( M \) of the model-scaled view relativized to some model \( M \) of \( \text{ZFC} \) (as expressed by the previous column). Truth in every model of \( \text{ZFC} \), and thus derivability in \( \text{ZFC} \), does not purport to capture what it means for a sentence of the language \( L \) of \( \text{ZFC} \) to be true, but only generalizes truth in \( M \) in the same way as \( \text{ZFC} \models i \phi \) generalizes \( M \)-logical consequence to any \( M \). Thus, the bottom right-most box actually answers, rather than the KST problem properly speaking, a variant of it, namely the generalization to any model of the model-relativization of the KST problem. Remarkably, the model-relativization of the KST problem (“Model-Scaled View”) gets itself a definite positive answer, that does not depend on the model \( M \) under consideration.


2.4 Elementary internal models

So the framework of internal models allows one to address and settle the KST problem, just as Kreisel and Boolos did. But it has two further advantages. Firstly, it relies only on the regular notion of truth in a structure and does not exceed the resources of ZFC itself: It resorts neither to some informal notion of truth (as in Kreisel), nor to some formal theory stronger than ZFC (as in Boolos), so that the treatment of the KST problem remains completely autonomous and yet reaches a clear and sharp solution. Moreover, as emphasized, this solution is uniform, unaltered by the model-relativization that makes it possible, so that the problem does not get scattered and receives a single answer. Secondly, this relativization to some model $M$ ("Is any $L$-sentence true in $M$ if an $M$-logical consequence of ZFC?") lends itself to a more detailed consideration, depending on what is assumed about $M$, and thus leads to more fine-grained results.

Indeed, the theorem 2.7 supports the conclusion that nothing new will ever come out of the consideration of all models of ZFC whatsoever, and that one should focus on the local consideration of the internal models of a single model of ZFC. Each model $M$ can be seen as giving rise to a specific logic $\mathcal{L}_M$, simply defined by the following: For each recursively enumerable set of sentences $T$ in $\mathcal{L}$ and each sentence $\phi$ of $\mathcal{L}$, $T \models_M \phi$ iff, for any internal $L$-structure $N_M$ in $M$, if $N_M \models T$ then $N_M \models \phi$. But (along the lines of the theorem 1.2) one has the basic results that if $M \models \phi$, then $ZFC \not\models_M \neg \phi$. This motivates us to consider not only a single base model of ZFC, but more specifically the relationship that there can be between a single base model $M$ and its internal models $N_M$.

A first thing that can be noticed is that $M$ cannot have any control of its internal models, to the extent that the class of all models internal to $M$ is not definable over $M$. Indeed, $n$ is a nonstandard integer of $M$ (if there is any) if and only if whenever $M \models \neg \forall N \models$ the first $n$ axioms of $\text{ZFC}^{-1}$, $N_M$ is a model of ZFC. Consequently, the notion of internal model cannot correspond to a definable class, because it would mean that $M$ would be able to define its nonstandard integers. So it is not possible from the point of view of $M$ itself to quantify over all its internal models, which precludes the definition, in $M$, of being an $M$-logical consequence.

Under these conditions, the first natural question hinges on the existence of an internal model $N_M$ being elementarily equivalent to $M$. Indeed, we know that, for each $L$-sentence $\phi$ such that $M \models \phi$, there exists some internal model $N_M$ of $M$ in which $\phi$ is true as well. In which cases is it possible to assume the internal model $N_M$ to be the same for all sentences $\phi$ true in $M$? In other words, which are the models $M$ whose semantic reflection is uniform? The next result answers that question. For the sake of its formulation, formulas $\phi$ of $L$ are now coded by numerals $\overline{n}(\phi)$ rather than by finite sequences of numerals. For any integer $n$, one notes $s(n)$ for the formula, if any, coded by $n$.

**Theorem 2.8.** Let $M$ be a model of ZFC. One defines the standard system of $M$ as being the set of the standard truncatures of all real numbers of $M$:

$\text{St}(M) = \{\text{st}(A) : A \in |M|, M \models A \subseteq \omega\},$

where $\text{st}(A) = \{n \in \mathbb{N} : M \models \overline{n} \in A\}$. Then there is $N \in |M|$ such that $N_M \equiv M$ iff $\text{Th}(M) \in \text{St}(M)$.

**Proof.** Suppose that there exists $N \in |M|$ such that $N_M \equiv M$. Then $\text{Th}(M) = \text{Th}(N_M)$. Now, $N_M \models \sigma$ iff $M \models \text{st}(\sigma)$ for any sentence $\sigma$ of $L$. Besides, one can write a formula $\text{Sat}(N, x)$ in two variables $N, x$ formalizing the statement that $N$ is a structure for $L$ and that $x$ codes a sentence of $L$ true in $N$. Therefore, $A_N = \{\alpha \in \omega^M : M \models \text{Sat}(N, x)[x = \alpha]\}$ is an element of $M$ (because $M$ satisfies the comprehension schema) and, by construction, an object that $M$ reckons to be a set of finite integers. But then $\text{Th}(N_M) = \text{st}(A_N)$ belongs to $\text{St}(M)$, and so does $\text{Th}(M)$.
Conversely, suppose that \( \text{Th}(M) \) is in the standard system of \( M: \text{Th}(M) = \{ n \in \mathbb{N} : M \models s(n) \} = \text{st}(B) \) for some \( B \in |M| \) with \( M \models B \subseteq \omega \). Hence \( n \in \text{Th}(M) \) iff \( n \in \mathbb{N} \) and \( M \models \pi \in B \). One can assume that, for any \( x \in B, M \models \forall x \) is the code of some sentence of \( L^n \). Besides, there exists a function \( g \), definable in \( L \), such that, for any finite family \( F \) of formulas of \( L, n(\langle \phi \in F \rangle) = g(\langle n(\phi) : \phi \in F \rangle) \), where the integers \( n(\phi) \) are ordered increasingly. Then one states the following definition in \( M: M \models \forall x \in B \ j(x) = g(\langle y : y \in B, y \leq x \rangle) \). Now suppose \( M \models \forall x \in B \ \exists N \ \text{Sat}(N, j(x)) \). Then, by compactness (which holds in \( M \)), \( M \models \exists N \ \forall x \in B \ \text{Sat}(N, x) \). In particular, there exists \( N \in |M| \) such that, for any \( n \in \text{Th}(M), M \models \text{Sat}(N, n) \), and so \( N_M \equiv M \). On the other hand, if \( M \not\models \forall x \in B \ \exists N \ \text{Sat}(N, j(x)) \), one has that \( M \models \exists N \exists x_0 \ \forall x < x_0 \rightarrow \exists N \ \text{Sat}(N, j(x)) \) and \( \neg \exists N \ \text{Sat}(N, j(x)) \). For any \( n \in \text{Th}(M), M \models \bigwedge_{i \in \text{Th}(M), i \leq n} s(i) \), so by reflection, and because \( \{ \pi^n : n \in \text{Th}(M) \} \) is an initial segment of \( B, M \models \exists \alpha_n \ \text{Sat}(V_{\alpha_n}, s(n)) \), which means that \( M \models \exists \pi < x_0 \). Thus \( x_0 \) is nonstandard. But since \( M \models \exists N \ \text{Sat}(N, j(x_0 - 1)) \), once again there exists \( N \in |M| \) such that, for any \( n \in \text{Th}(M), M \models \text{Sat}(N, n) \), hence such that \( N_M \equiv M \). \( \square \)

The upshot of this result is that for any transitive \( \in \)-model \( |M|, \in \) of \( \text{ZFC} \), there is \( x \in |M| \) such that \( \langle x, \in \rangle \equiv \langle |M|, \in \rangle \) iff, for any \( M \)-definable subtheory \( S \) of \( \text{Th}(M) \), \( M \models \text{SM}(S) \), where \( \text{SM}(S) \) is a shorthand for the sentence (in the language of \( \text{ZFC} \)) to the effect that there is a standard model of \( S \). Following George Wilmers,\(^{28} \) such a model \( M \) is said to be “internally standard.”

The criterion given by the proposition 2.8 is nontrivial, since it really divides the spectrum of all models of set theory into two camps. Indeed, any full standard model of second-order set theory contains every real, hence in particular its own standard system. On the other hand, the theory of any pointwise definable model \( M \) of \( \text{ZFC} \) cannot be in \( M \)’s standard system. As a matter of fact, suppose that \( \text{Th}(M) \) is in \( \text{St}(M) \), which means that \( \text{Th}(M) \) is the standard part of some \( M \)-sequence \( s \in |M| \). Now let \( \phi(x) \) express “\( x \) codes a sentence whose negation belongs to \( s \).” In particular, \( M \models \phi(x) \) iff \( M \models \neg s(x) \). Since \( s \) is definable, \( \phi(x) \) is a formula without parameters. By Kleene’s fixed point theorem, there is a sentence \( \psi \) such that \( \text{ZFC} \vdash \psi \leftrightarrow \phi(n(\psi)) \). But then \( M \models \psi \) iff \( M \vdash \neg \psi \). Since it is a fact (established by Ali Enayat)\(^{29} \) that every countable model of \( \text{ZFC} \) has a pointwise definable model as a generic extension, models without elementary equivalent internal models do exist.\(^{30} \)

The natural step to take to strengthen the proposition 2.8 is to require that the internal model is an elementary substructure of the original one. To put things slightly differently: On which conditions could one get the existence of \( N \in |M| \) such that \( N = (V_\alpha)^M \) for some \( \alpha \in |M| \) with \( M \models \text{“} \alpha \text{ is a transfinite ordinal”} ? \) Obviously, such an internal elementary substructure cannot be found in the case of Shepherdson’s \( M_0 \). So suppose, for the sake of argument, that \( M \) is non-\( \omega \)-standard. The idea would be to establish, for some nonstandard integer \( n_0 \) of \( M \), that \( M \models \forall \varphi(\bar{x}) \in \Sigma_{n_0} \ \forall \bar{a} \in |N|^k (\varphi \iff \varphi(\bar{n}(\bar{a}))) \). But this cannot be obtained by any application of the reflection schema. In fact, the set of sentences true in \( (M, V_\alpha^M) \) (that is, in the expansion of \( M \) into an interpretation of \( L \cup \{ c_a : a \in V_\alpha^M \} \) is too big to be a set in \( M \). The best approximation of the existence of an internal elementary substructure \( N_M \prec M \) lies in the following result, whose proof goes along the same principles as that of the previous theorem:

**Theorem 2.9.** Let \( M \) be a model of \( \text{ZFC} \) and \( \alpha \) an ordinal of \( M \). Then there exists \( N \in |M| \) such that (i) \( V_\alpha^M \subseteq |N| \) and (ii) \( (N_M, V_\alpha^M) \equiv (M, V_\alpha^M) \) iff \( \exists s : (V_\alpha^M)^{\omega_M} \rightarrow \phi(\omega_M), s \in |M|, \) such that \( \forall \bar{a} \in V_\alpha^M \ \text{st}(s(\bar{a})) = \text{Th}(M, \bar{a}) \).

Again, the condition of this theorem really divides the models of \( \text{ZFC} \) into two camps. Indeed, the minimal model \( M_0 \) of \( \text{ZFC} \) is obviously in the negative camp. On the contrary, any
recursively saturated model of ZFC is in the positive camp (it is a standard result that any consistent first-order theory has a finite or countable recursively saturated model). As a matter of fact, suppose that $M$ is a recursively saturated model of ZFC. Then, for $n$ fixed, let $\phi_n(x)$ be the formula to the effect that $x$ is an ordinal and that any tuple of elements of $V_x$ exactly when it satisfies it in $M$. By reflection, $\phi_n(x)$ is realizable for any $n$. So, by recursive saturation, the set of all the $\phi_n$ is also realizable by some ordinal $\beta$ of $M$. For such ordinal $\beta$, one has that $(V_\beta)^M$ is an elementary substructure of $M$.

### 2.5 Stronger and weaker set theories

The previous results can be extended to set theories stronger than ZFC. In particular, this is the case for Morse-Kelley set theory (MK), which is a first-order two-sorted analog of second-order set theory. In fact, its objects are only classes, sets (denoted with lower case variables) being defined as those classes $x$ which belong to some class ($\exists X x \in X$). Its intended models are the $V_{\kappa+1}$’s for inaccessible cardinals $\kappa$. One of the distinctive features of MK is allowing the bound variables in the schematic formula appearing in the class comprehension schema to range over proper classes as well as sets: $\forall W_1 \ldots \forall W_n \exists Y (x \in Y \leftrightarrow \exists X (x \in X \land \phi(x, W_1, \ldots, W_n)))$ is an axiom for every formula $\phi(W_1, \ldots, W_n)$ in which $Y$ is not free. This in particular is the case for $Y = V$, where $V$ is the universal class defined by $\forall x (\exists X x \in X \rightarrow x \in V)$. The fact that the theory MK is strictly stronger than ZFC comes from the impredicativity of the comprehension schema. It is a proper (non-conservative) extension of ZFC. But it does not preclude the facts 1.2, 2.3, 2.7 and 2.8 from extending from ZFC to MK. In particular, every model of MK contains as an element of its domain a structure whose replica is a model of MK, and every model of MK contains an elementary equivalent internal model iff its full theory is an element of its standard system (defined in an analogous way as in the case of a model of ZFC).

Finally, the last proposition still holds true when $M$ is a model of MK and $(V_\alpha)^M$ is replaced with $(V_{\alpha+1})^M$ (to allow for the fact that the second-order part of $(V_\alpha)^M$ is preserved).

Some results can also be found about set theories weaker than ZFC, in particular about admissible set theory, or Kripke-Platek set theory with urelements (KPU). As a theory, KPU represents a weakening of ZFC which is aimed at embedding recursion theory into model theory. This induces in particular the consideration of the linguistic resources internal to some admissible set $A$ (that is, to some model of KPU) and the statement of a completeness theorem (Barwise Completeness Theorem)\textsuperscript{31} with respect to the language internal to that admissible set. But KPU retains all that is necessary to enable a structure to give rise to the kind of reflective constructions that can be carried out within the models of ZFC. More specifically, the fragment $L_A$ defined by an admissible set $A$ is defined as the set of all formulas $\varphi$ of $L_{\infty, \omega}$ whose codes belong to $A$. One can then focus on results establishing (1) the conditions on which a sentence of $L_A$ is satisfied in a structure internal to $A$, or, in the reverse direction, (2) the conditions on which a sentence of $L_A$ is valid w.r.t. an admissible set $B$ to which $A$ is internal. Because admissible sets are supposed to be transitive $\in$-models of KPU, any admissible set $A \in B$ coincides with the corresponding internal model $A_B$.

As for question (1), there is a classical result to the effect that if two structures $M, N$ for $L$ are both internal to $A$ ($M, N \in |A|$) and $L_A$-elementary equivalent, then they are potentially isomorphic (and thus isomorphic in case $M$ and $N$ are furthermore supposed to be countable). Consequently, the set of sentences of $L_{\infty, \omega}$ belonging to $A$ represents a measure of the variety of countable structures internal to $A$, in the sense that if $M, N \in |A|$ are not isomorphic, then there exists a discriminating sentence $\phi \in L_A$ such that $M \models \phi$ and $N \models \neg \phi$.

As for question (2), we have the following fact.\textsuperscript{32}

**Proposition 2.10.** Let $A, B$ two admissible sets such that $A \in |B|$ and $B \models \text{"A is countable"}$. 

18
Then for any $\phi \in L_A$, one has that $B \models \gamma \phi$ is valid $\gamma$ iff $A \models \exists P \gamma P$ is a proof of $\phi \gamma$.

Proof. Suppose $B \models \gamma \phi$ is valid $\gamma$. Then (by Barwise completeness theorem, which can be derived within KPU), $B \models \exists P \gamma P$ is an infinitary proof of $\phi \gamma$, hence there really exists such an infinitary proof $P$ of $\phi$, hence (again owing to Barwise completeness theorem) $A \models \exists P \gamma P$ is a proof of $\phi \gamma$. Conversely, suppose $A \models \exists P \gamma P$ is a proof of $\phi \gamma$. Then $\phi$ is valid. Now the set $Sat(A)$ of all the valid sentences of $L_A$ is $\Sigma_1$ on $A$, hence an object of $B$ (because $\mathbb{YP}_A \subseteq B$). By absoluteness, $\phi \in Sat(A)$ implies that $B \models \phi \in Sat(A)$, and thus that $B \models \gamma \phi$ is valid $\gamma$. \hfill $\square$

**Corollary 2.11.** Let $A$ and $B$ be admissible sets, with the same hypotheses as above. Then, for any sentence $\phi$ of $L_A$, $\text{ZFC} \models_B \phi$ iff $\phi$ is valid.

One last thing ought to be mentioned to specify the kind of special result that admissible sets lend themselves to as far as internal models go. Let $A$ be an admissible set and $A' \in |A|$. According to Barwise completeness theorem, the set $E_A$ of valid $A$-sentences of $L_{\infty,\omega}$ is $\Sigma_A$ on $A$, which means that there exists a $\Sigma$-formula $F(x, a)$ with parameters in $A$ such that $\sigma \in E_A$ implies $A \models F(\exists \sigma, \bar{a})$. Besides, $L_{A'}$ is a $\Delta_A$-subset of $A$. Likewise, the main syntactic and semantical notions relative to $L_{A'}$ are $\Delta$ on $A$. In particular, there is a $\Delta_A$-formula $S(x, y)$ such that for any sentence $\sigma$ of $L_A$, $A \models S(A', \gamma \sigma)$ iff $A' \models \sigma$. Moreover, the members of $A$ which are themselves structures for $L_A$ form also a $\Delta_A$-subset of $A$; Let’s note $G_A(x)$ the corresponding $\Delta_A$-formula. Now let’s say that an admissible set $A$ is reflective if, for any sentence $\sigma$ of $L_A$, $\sigma$ is a valid sentence iff $A \models S(A', \gamma \sigma)$ for any structure $A' \in A$ for $L_A$, that is, iff $A \models \forall A' \left( G_A(A') \rightarrow S(A', \gamma \sigma) \right)$. As this last formula is $\Pi_A$, the set $E_A$ of all valid $A$-sentences is $\Delta_A$. One may partially characterize the admissible sets $A$ such that $E_A$ is $\Sigma_A$ and not $\Delta_A$. As a matter of fact, let’s put, for any admissible $A$: $A^+ = \bigcap \{B / A \in B, B \models \text{ZFC} \}$; $A^+$ is itself an admissible set. An admissible set of this form is called a next admissible set. For any next admissible set $A$, $E_A$ is provably not $\Delta_A$.$^{33}$ Consequently, no next admissible set is reflective.

### 3 Going modal

It is possible to show to some further extent the fruitfulness of the setting that we chose to settle the KST problem. That setting has important conceptual advantages, that we already mentioned. In particular, it allows one to bypass the notion of truth in the universe. But it also lends itself, on top of the fine-grained results already stated, to a modal twist that opens up further results of a new kind.

The idea of what a given model can see leads indeed to think of any internal model $N_M$ as being accessible from the point of view of $M$. This amounts to thinking of models of ZFC as possible worlds within some Kripke frame for modal logic,$^{34}$ and to defining an accessibility relation between models of ZFC by setting: $M'$ is accessible from $M$ iff $M'$ is (isomorphic to) some model of ZFC internal to $M$.

This has connections with the “modal logic of forcing” developed by Hamkins and Benedikt Löwe,$^{35}$ but the great difference with the latter is that the accessibility relation works downward instead of going upward and that the collection of all successors of a given model of ZFC is bound to be a set, not a proper class. In that perspective, the additional semantical clause is:

**Definition 3.1.** Let $M$ be a model of ZFC and $\phi$ a sentence of the language $L$ of ZFC. Then ‘$M \models \Box \phi$’ is a shorthand for the existence of some $N \in |M|$ such that $N_M \models \text{ZFC} + \phi$.

Defining $\Box \phi$ as $\neg \Box \neg \phi$ as usual, one has that $M \models \Box \phi$ is equivalent to $\text{ZFC} \models_M \phi$. 

19
Adding a modal operator directly to the language \( L \) of ZFC, however, is not an option, in particular because it would be awkward to extend the separation and replacement schemes to formulas involving modalities. This is where propositional modal logic comes into play. Its language is the language \( L' \) generated by the addition of \( \Box \) to the language of propositional logic. A formal system of modal logic in \( L' \) is said to be normal if the following holds: (i) all tautologies of propositional calculus; (ii) the axiom \( \text{K} = \Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q) \); (iii) the rule of uniform substitution (if \( \phi(p_1, \ldots, p_n) \) is a theorem, so is \( \phi[\beta_k/p_k] \) \((k = 1, \ldots, n)\); (iv) modus ponens; (v) the rule of necessitation (if \( \alpha \) is a theorem, so is \( \Box \alpha \)). Important modal axioms are: \( p \rightarrow \Box p \), or, equivalently, \( \Box p \rightarrow p \) \((\text{T})\), \( \Diamond \Box p \rightarrow \Diamond p \) \((4)\), \( \Box p \rightarrow \Box \Diamond p \) \((5)\), and \( \Box (\Box p \rightarrow p) \rightarrow \Box p \) \((\text{GL})\). S4 is the normal modal system that contains the axioms \text{T} and \text{4}; \text{S5} is \text{S4} + 5.

**Definition 3.2.** An interpretation \( i \) of \( L' \) into \( L \) is a map that assigns to each propositional letter \( p \) an arbitrary sentence of \( L \).

For any such interpretation \( i \) and any structure \( M \) for \( L \), it is possible to define inductively \( 'M \models i(A)' \) for every modal proposition \( A \) as follows:

- \( M \models i(\neg A) \) iff \( M \not\models i(A) \)
- \( M \models i((A \land B)) \) iff \( M \models i(A) \) and \( M \models i(B) \)
- \( M \models i(\Box A) \) iff \( \text{ZFC} \models M, i(A) \).

**Definition 3.3.** Given a formula \( A \) of \( L' \) and a model \( M \) of ZFC, \( A \) is modal-internally valid in \( M \), written \( M \models_{\text{int}} A \), iff for any interpretation \( i \) of \( L' \) into \( L \), \( M \models i(A) \).

A formula \( A \) of \( L' \) is a valid principle of internal modal logic if it is modal-internally valid in any model of ZFC. The set of all valid principles of internal modal logic is denoted by \( \text{IML} \).

**Proposition 3.4.** There is no formula \( P(x) \) of \( L' \) such that, for any sentence \( \phi \) of \( L \) and any model \( M \) of ZFC, \( M \models \Box \phi \) iff \( M \models P(\overline{n(\phi)}) \).

**Proof.** Suppose that there is such a formula. By virtue of the fixed point theorem, there is a sentence \( \phi \) such that \( \phi = \neg P(\overline{n(\phi)}) \). Now, suppose \( M \models \phi \). This implies \( M \models \Box \phi \), that is, by hypothesis, \( M \models P(\overline{n(\phi)}) \), i.e., \( M \models \neg \phi \). In consequence, \( M \models \phi \) for any \( M \models \text{ZFC} \), and so \( P(\overline{n(\phi)}) \in \text{ZFC} \). Thus, supposing that ZFC is consistent and that there is a model \( M \) of ZFC = \( \text{ZFC} + \neg \phi \), one gets that \( M \models P(\overline{n(\phi)}) \) and so, by hypothesis, that \( M \models \Box \phi \). Hence the existence of an internal model \( N_M \) of ZFC + \( \neg \phi + \phi \), which is excluded. As a result, on the assumption that ZFC is consistent, there is no such formula \( P(x) \), which proves that \( L(\Box) \) has an expressive power of its own in comparison to \( L \).36

The modal operator \( \Box \) introduced in the context of \( \text{IML} \) also has an expressive power of its own in comparison to usual systems of modal logic. Indeed, propositional modal logic cannot easily make sense of the possibility of a whole theory, instead of a single sentence. In Kripke semantics for modal logic, the clause \( \mathfrak{M}, w \models \Box T \) cannot be rendered by \( \forall \sigma \in T \lor \mathfrak{M}, w \models \Box \sigma \) (the conjunction of possibilities is not the possibility of the conjunction), and the easy way \( (\forall \mathfrak{M}, w \models \Box \bigwedge_{\sigma \in T} \sigma) \) is not available, unless one allows infinitary conjunctions. On the contrary, the definition of \( M \models \Box \phi \) in internal modal logic can readily be generalized into: \( M \models \Box T \) iff there is \( N \in |M| \) such that \( N_M \models T \), for any theory \( T \) extending ZFC. For a standard model \( M \) and \( T \in |M| \), one has that \( M \models \Box T \) iff there is \( N \in |M| \) such that \( M \models \neg N \models T \). But of course this is not the general case. In effect, \( M = M^* \) and \( T = \text{ZFC} \) give a counterexample. Thus, the meaning of the internal modal operators cannot be expressed in the language \( L \) of set theory.
Before turning to internal modal logic, an alternative presentation could come in useful to make it appear that in fact all models of ZFC are involved in making up a huge modal frame. Let’s say that a model $N$ of ZFC is accessible* from a given model $M$ iff, for any $L$-sentence $\phi$, $ZFC \models_M \phi$ implies $N \models \phi$. Then, let’s say that $\square^* \phi$ is true in $M$ iff $\phi$ is true in every model of ZFC accessible* from $M$.

**Lemma 3.5.** For any sentence $\phi$ of $L$ and any model $M$ of ZFC, $M \models \square^* \phi$ iff $M \models \square \phi$.

**Proof.** The forward implication is immediate. For the converse implication, it is sufficient to remark that ZFC $\models_M \phi$ implies that $N \models \phi$ for any model $N$ accessible* from $M$.

**Proposition 3.6.** IML is a normal modal logic.

**Proof.** The validity of all propositional tautologies and of the axiom $K$ is straightforward, as is the preservation of validity by *modus ponens*. As to the uniform substitution rule, suppose that $M \models i(\phi(p_1, \ldots, p_n))$, for any interpretation $i$ and any model $M$. There is no such interpretation as $i^* : p_k \mapsto i(\beta_k)$, but one can proceed by induction on $\phi$. The only case worth considering is $\phi = \square \psi$. So suppose $M \models i(\square \psi(p_1, \ldots, p_n))$, for any interpretation $i$ and any model $M$.

In other words, $N_M \models i(\psi(p_1, \ldots, p_n))$ for any internal model $N_M$ of any model $M$. Given a fixed model $M$, $\text{Th}(M)$ is obviously consistent, so, as in the proof of the theorem 2.6, the existence of $N \in \|M\|$ such that $N_M \models \text{Th}(M)$ can be concluded: Any model $M$ of ZFC is elementarily equivalent to one of its internal models. As a consequence of the hypothesis on $\square \psi$, $M \models i(\psi(p_1, \ldots, p_n))$ and, by induction hypothesis, $M \models i(\psi[\beta_k/p_k])$. This is true for any model $M$ and so, in particular, for any internal model $N_M$ of $M$. Thus $ZFC \models_M i(\psi[\beta_k/p_k])$, hence $M \models i(\square(\psi[\beta_k/p_k]))$, i.e., $M \models i(\phi[\beta_k/p_k])$, which is true again for any interpretation $i$ and any model $M$. Finally, as to the necessitation rule, suppose $M \models i(\alpha)$ for any model $M$, and thus in particular for any internal model $N_M$ of any model $M$. Then $M \models i(\square \alpha)$, and this holds for any interpretation $i$ and any model $M$.

One also has the following result.

**Proposition 3.7.** $T \in \text{IML}$.

**Proof.** Let $M$ and $i$ be any model of ZFC and any interpretation of $L'$ into $L$, respectively, and let’s suppose that $M \not\models i(p)$. Then it follows from the theorem 2.6 that $ZFC \not\models_M i(p)$. In other words, $M \models i(\square p)$ only if $M \models i(p)$, for any interpretation $i$. So $M \models_{\text{int}} T$, and this holds for any $M$.

It is noteworthy that the axiom $T$ is IML-valid despite the fact that the accessibility relation in play is not reflexive. (Indeed, it is obviously untrue that any model of ZFC is isomorphic to one of its own internal models.) This suffices to show the difference between IML and Kripke semantics. Actually, IML is specifically a way of encoding the existence of internal models in the guise of a $T$-style axiom, and thus of establishing a connection between set-theoretic reflection and modal reflexivity.

**Proposition 3.8.** GL $\notin \text{IML}$.

**Proof.** Let $M$ be a model of ZFC and let $i$ be some interpretation of $L'$ into $L$. By definition, $M \models i(\square(\square p \rightarrow p))$ iff $N_M \models i(\square p \rightarrow p)$ for any model $N_M$ of ZFC internal to $M$. But the theorem 1.2 and thus the axiom $T$ (looked at as a schematic set-theoretic fact) are themselves theorems of ZFC, hence IML-valid in any such $N_M$. So $M \models_{\text{int}} \square(\square p \rightarrow p)$, and so $M \models i(\square(\square p \rightarrow p))$, whatever $i$ may be. Still, $M \not\models i(\square p)$ does not hold as soon as $i(p) \notin \text{Th}(M)$, since in that case there exists an internal model $N_M$ of $M$ for which $N_M \not\models ZFC + \neg i(p)$, so that $ZFC \not\models_M i(p)$.
Definition 3.9. Given a class \( \mathcal{K} \) of models of ZFC, ‘\( M \vDash^\mathcal{K} i(A) \)’ is defined as in the Definition 3.2, except for the last clause, which is replaced with:

\[
M \vDash^\mathcal{K} i(\Box A) \text{ iff, for every internal model } N_M \text{ of } M \text{ in } \mathcal{K}, N_M \vDash^\mathcal{K} i(A).
\]

A \( \mathcal{K} \)-valid* principle of internal modal logic is a formula \( A \) of \( L' \) such that \( M \vDash^\mathcal{K} i(A) \) for any interpretation \( i \) of \( L' \) into \( L \) and any member \( M \) of \( \mathcal{K} \), which is written: \( \mathcal{K} \vDash^*_{\text{iml}} A \).

A natural question is: Which are the \( \mathcal{K} \)-valid* principles for well-known classes \( \mathcal{K} \) of models of ZFC?

Proposition 3.10. Let \( \mathcal{T} \) be the class of all transitive models of ZFC. One has: \( \mathcal{T} \vDash^*_{\text{iml}} S4 \).

Proof. Firstly (by the Propositions 3.6 and 3.7), \( \mathcal{T} \vDash^*_{\text{iml}} KT \). Furthermore, owing to Jensen-Karp theorem, if \( M, N \in \mathcal{T} \) with \( |N| < |M| \) \( (|M|) \) being the cardinality of the domain of \( M \), then, for any interpretation \( i \) of \( L' \) into \( L \), \( N \vDash i(p) \) implies \( M \vDash i(\Box p) \), even if \( N \not\vDash |M| \). Consequently, \( M \vDash i(\Box p) \) implies that \( N \vDash i(p) \) for any \( N \in \mathcal{T} \) such that \( |N| < |M| \). This holds in particular for any internal model \( N \) of \( M \) in \( \mathcal{T} \), and for any internal model of any internal model of \( M \) in \( \mathcal{T} \) as well. Hence, supposing \( M \vDash^\mathcal{T} i(\Box p) \), one has that \( N \vDash^\mathcal{T} i(\Box p) \) for any internal model \( N \) of \( M \) in \( \mathcal{T} \), and thus that \( M \vDash^\mathcal{T} i(\Box p) \). This holds for any interpretation \( i \) and any \( M \) in \( \mathcal{T} \), so \( \mathcal{T} \vDash^*_{\text{iml}} 4 \). \( \square \)

Proposition 3.11. Let \( \mathcal{S} \) be the class of all standard models of ZFC. One has: \( \mathcal{S} \nvDash^*_{\text{iml}} 5 \).

Proof. By minimality, the minimal model \( M_0 \) satisfies \( \neg \text{SM}(ZFC) \), where ‘\( \text{SM}(ZFC) \)’ is the sentence of \( L \) asserting the existence of a standard model of ZFC. Now, \( M_0 \) is isomorphic to \( L_0^\gamma \) for some standard model \( M \) of ZFC, where \( \gamma \) is the ordinal of all \( M \)-ordinals. By Mostowski collapsing lemma, \( M \) can be taken to be an \( \in \)-transitive model, so that \( M_0 \) is the real \( L_\gamma \). Since any internal model of \( M_0 \) is nonstandard, \( M_0 \vDash^\mathcal{S} \neg i(\Box p) \), whatever the interpretation \( i \) may be. As a result, if one considers the transitive \( \in \)-model \( M_2 = L_{\gamma+2} \), to which \( M_0 \) belongs, one has that \( (M_0)_M = M_0 \nvDash^\mathcal{S} i(\Box p) \), so that \( M_2 \nvDash^\mathcal{S} i(\Box p) \). Still, considering the internal (standard) model \( M_1 = L_{\gamma+1} = (M_1)_M \) of \( M_2 \), one has that \( M_1 \vDash \text{SM}(ZFC) \), since \( M_1 \vDash \neg \text{SM}(ZFC) \), so that \( M_2 \nvDash^\mathcal{S} i(\Box p) \) for any interpretation \( i \) assigning \( SM(ZFC) \) to \( p \). Hence \( M_2 \nvDash^\mathcal{S} i(5) \) for some interpretation \( i \), which entails that the axiom \( 5 \) is not \( \mathcal{S} \)-valid*.

It appears that the results above depend very much on limitations which are peculiar to the classes of models at stake. On the contrary, certain classes of models naturally stand out, namely those which are stable under internal models.

Definition 3.12. A class \( \mathcal{K} \) of models of ZFC is weakly downward stable if, for every \( M \in \mathcal{K} \), there exists \( N \in |M| \) such that \( N_M \in \mathcal{K} \). It is strongly downward stable if, for every \( M \in \mathcal{K} \) and every \( N \in |M| \), \( N_M \vDash ZFC \) implies \( N_M \in \mathcal{K} \).

A central advantage of any strongly downward stable class \( \mathcal{K} \) of models of ZFC is that \( \mathcal{K} \)-validity* can be expressed in a much more natural way and becomes homogeneous with modal-internally validity (in the sense of the Definitions 3.2 and 3.3).

Definition 3.13. A \( \mathcal{K} \)-valid principle of internal modal logic, for a given class \( \mathcal{K} \) of models of ZFC, is a formula \( A \) of \( L' \) that is modal-internally valid in any member of \( \mathcal{K} \), which is written: \( \mathcal{K} \vDash^*_{\text{iml}} A \).

Remark 3.14. For any strongly downward stable class \( \mathcal{K} \) of models of ZFC, \( \mathcal{K} \vDash^*_{\text{iml}} A \) iff \( \mathcal{K} \vDash_{\text{iml}} A \).
The class $\mathcal{M}$ of all models of ZFC is strongly stable for trivial reasons. But we know already that the only theory that is complete w.r.t. $\mathcal{M}$ is ZFC itself. So a better understanding of IML requires getting hold of other stable classes.

**Lemma 3.15.** $S$ and $T$ are not weakly downward stable.

**Proof.** This is due to the fact that the minimal model $M_0$ is a transitive, and thus standard, model of ZFC but that, by minimality, any internal model of $M_0$ has to be nonstandard, and thus not transitive.

**Lemma 3.16.** The class $\mathcal{R}$ of all countable recursively saturated models of ZFC and the class $\mathcal{N}$ of all non-$\omega$-standard models of ZFC are both strongly downward stable.

**Proof.** Let $M \in \mathcal{R}$ and $N \in |M|$ be such that $N_M \vDash ZFC$. Firstly, $N_M$ is obviously countable. Secondly, $M$, as any recursively saturated model of ZFC, is non-$\omega$-standard. Indeed, the type $p(x) = \{\text{Ord}(x)\} \cup \{x > n : n \in \mathbb{N}\} \cup \{x < \omega\}$ is a recursive type in $x$ that is finitely realized, and thus realized, in $M$. But any witness of $p(x)$ in $M$ is a nonstandard integer of $M$. Now, any internal model of a non-$\omega$-standard model of ZFC is recursively saturated. Indeed, let’s consider a non-$\omega$-standard model $M$ of ZFC, $N \in |M|$ such that $N_M \vDash ZFC$ and $p(x) = (\phi_n(x))_{n \in \mathbb{N}}$ a recursive type in $L$. Let’s suppose that $p(x)$ is finitely realizable in $N_M$: $N_M \vDash \exists x \wedge \bigwedge_{k < n} \phi_k(x)$ for any (standard) $n \in \mathbb{N}$. In other words, $M \vDash \forall N \vDash \exists x \bigwedge_{k < n} \phi_k(x)\,$, which can be written $M \vDash \Phi(N, n)$ with $\Phi \in L$. Now, by compactness, $M \vDash \Phi(N, \bar{c})$ for some nonstandard integer $c$ of $M$, which means that there is $b \in |N_M|$ such that, for any $k \in \mathbb{N}$ with $\bar{c}^M < c$, and thus for any $k \in \mathbb{N}$ whatsoever, $N_M \vDash \phi_k(x)[b]$. So $b$ realizes $p(x)$ in $N_M$, and so $N_M$ is recursively saturated. As a result, $\mathcal{R}$ is strongly downward stable. Besides, any internal model of a member of $\mathcal{N}$, being recursively saturated, is non-$\omega$-standard, and thus $\mathcal{N}$ is also strongly downward stable.

**Definition 3.17.** Given a first-order language $L_0$, an $L_0$-structure $M$ is said to be resplendent if, whenever $N \vDash \exists R \phi(R, \bar{m})$ with $M \prec N$ and $\bar{m} \in |M|^k$, $M \vDash \exists R \phi(R, \bar{m})$, where ‘$R$’ stands for a second-order variable.

In other words, $M$ is resplendent iff $M$ satisfies a $\Sigma^1_1$-formula $\exists R \phi(R, \bar{x})$ as soon as it satisfies all its first-order consequences $\{\psi(\bar{x}) \in L_0 : \phi(R, \bar{x}) \models \psi(\bar{x})\}$. Any resplendent structure is recursively saturated. Besides, Jon Barwise and Jean-Pierre Ressayre proved independently that any countable recursively saturated structure is resplendent. So in fact $\mathcal{R}$ coincides with the class of all countable resplendent (non-$\omega$-standard) models of ZFC. The class $\mathcal{R}$ has been studied by Victoria Gitman and Joel David Hamkins and proved to be a model of “the multiverse axioms.”

Gitman and Hamkins’ study also considers the class $\mathcal{R}$ for stability reasons and brings out very nicely how internally rich that class turns out to be. Still, it does not deal with modal issues and puts forward multiverse axioms which, by their very meaning, are upward oriented. I would like now to apply to the class $\mathcal{R}$ the downward modal point of view attached to IML.

**Theorem 3.18.** $\mathcal{R} \models_{\text{IML}} S4$.

**Proof.** Let $M \in \mathcal{R}$ and an interpretation $i$ be such that $M \vDash i(\Diamond \Diamond p)$. This means that there exists $N = (|N|, E) \in |M|$ such that $N_M \vDash \text{ZFC} + i(\Diamond \Diamond p)$. So there exists $\alpha = (\nu, \eta) \in |N_M|$ such that $\alpha_{N_M} \vDash \text{ZFC} + \phi$, where $\phi = i(p)$. Now, for any sentence $\theta$ of $L$, let’s consider the following sentence $\theta^*$ of $L(P^{(1)}, R^{(2)})$: $\gamma \wedge \exists x \forall y (P_y \leftrightarrow y \in x) \wedge \exists \forall y \forall z (R_y z \leftrightarrow (y, z) \in r) \wedge \exists \forall y \forall z (\gamma \wedge \exists \forall y \forall z (R_y z \leftrightarrow (y, z) \in r) \wedge \exists \forall y \forall z (\gamma \wedge \exists \forall y \forall z (R_y z \leftrightarrow (y, z) \in r)$.

Let $T_0$ be any finite fragment of ZFC + $\phi$, and $\sigma_0$ the conjunction of all the members of $T_0$. One has that $N_M \vDash \gamma \wedge \sigma_0$, so $(N_M, \nu_{N_M}, \eta_{N_M}) \vDash$
σ_0^*(P, R), and so \( M \models \neg \sigma \) \( \models \neg \nu \) and \( \eta \) are sets \( \wedge \neg \nu, \eta \) \( \models \sigma_0^* \). Hence (for the same reasons as in the proof of the Lemma 1.1) \( M \models \neg \nu \) and \( \eta \) are sets \( \wedge \neg \nu, \eta \) \( \models \sigma_0^* \), with \( \nu_N = \{ x \in |N| : N \models x \in \nu \} \) and \( \eta_N = \{ (x, y) \in |N| \times |N| : N \models (x, y) \in \eta \} \) (by comprehension, \( \nu_N \) and \( \eta_N \), as interpreted in \( M \), are indeed members of \( |M| \)). Thus \( \langle M, (\nu_N)_M, (\eta_N)_M \rangle = \langle M, \nu_{N_M}, \eta_{N_M} \rangle \models \sigma_0^*(P, R) \). \( \nu_{N_M} \) and \( \eta_{N_M} \) are indeed subsets of \( |M| \) and \( |M| \times |M| \), respectively, even though they are not necessarily members of \( |M| \). So \( M \) can be expanded to a model of any finite fragment of \( (ZFC + \phi)^* \). Now, by a result due to Jon Barwise,⁰ any resplendent \( L \)-structure \( M \), some elementary extension \( N \) of which can be expanded to a model of a recursive theory \( T^1 \) in \( L(R_1, \ldots, R_m) \), can itself be expanded to a model of \( T^1 \). Owing to \( M \)'s resplendency, it follows that \( M \) can be expanded to a model of the recursive theory \( (ZFC + \phi)^* \) in \( L(P, R) \). Thus there are \( A \subseteq |M| \) and \( B \subseteq |M| \times |M| \) such that \( \langle M, A, B \rangle \models (ZFC + \phi)^* \). This implies that there are \( a \) and \( b \) in \( |M| \) such that \( a_M = A, b_M = B \) and \( M \models \neg \sigma(a, b) \models \theta^N \) for any \( \theta \in ZFC + \phi \). So \( \langle a_M, b_M \rangle \models ZFC + \phi, \phi = \dot{i}(p) \), and so \( M \models \dot{i}(\diamond p) \).

### Definition 3.19
A modal theory \( \Lambda \) is iML-complete w.r.t. a class \( K \) of models of \( ZFC \) if, for any formula \( A \) of \( L' \), \( \Lambda \models A \) iff \( K \models \dot{A} \).

### Theorem 3.20
\( S4 \) is iML-complete w.r.t. \( R \).

**Proof.** Owing to the preceding theorem, it only remains to show that if \( S4 \not\models A \), then there is an interpretation \( \dot{i} \) and \( M \in R \) such that \( M \not\models \dot{i}(A) \). One proceeds by induction on \( A \).

For a propositional variable \( p \), suppose that for any interpretation \( \dot{i} \) and any model \( N \) of \( ZFC \), \( N \models \dot{i}(p) \). This means that \( p \) is a tautology and thus that \( S4 \not\models \dot{p} \). So \( S4 \not\models \phi \) implies that there is a model \( N \) of \( ZFC \) such that \( N \not\models \dot{i}(p) \) for some interpretation \( \dot{i} \) of \( L' \). Now the conclusion results from a classical fact, already mentioned, namely that any complete first-order theory whose models are all infinite has a recursively saturated model: This holds in particular for the theory \( Th(N) \) of \( N \). For \( A = \neg B \), suppose that, for any interpretation \( \dot{i} \) of \( L' \) and any model \( M \) in \( R \), \( M \models \dot{i}(\neg B) \). As a consequence, for any fixed interpretation \( \dot{i} \), the theory \( ZFC + \dot{i}(B) \) is inconsistent: Otherwise, there is a model \( M \) of that theory and a countable recursively saturated model of \( Th(M) \), and thus a model \( M \in R \) of \( i(B) \). So, under the assumption that \( ZFC \) is consistent, \( ZFC \not\models i(B) \) for any interpretation \( i \) of \( L' \) into \( L \). As above, \( B \) is then an antilogy and thus \( A \) a theorem of \( S4 \). For \( A = (B \land C) \), \( S4 \not\models (B \land C) \) implies \( S4 \not\models B \) or \( S4 \not\models C \), let’s say \( S4 \not\models B \). But then (by induction hypothesis) there is an interpretation \( i \) and \( M \in R \) such that \( M \not\models i(B) \), thus such that \( M \not\models (i(B) \land i(C)) \), and so such that \( M \not\models i(A) \). Finally, let’s consider \( A = \neg B \), where \( B \) is any formula, the induction hypothesis being that if \( S4 \not\models B \), then there is an interpretation \( i \) and \( M \in R \) such that \( M \not\models i(B) \). Let’s suppose that \( S4 \not\models A \). Then (owing to the necessitation rule) \( S4 \not\models i(B) \), so by induction hypothesis there is an interpretation \( i \) and \( M \in R \) such that \( M \not\models i(B) \). Now, by [Schlipf, 1977] (Theorem III.2.6), there is a non-\( \omega \)-standard model \( U \) of \( ZFC \) and \( M' \subseteq |U| \) such that \( M \simeq M'|_U \). So \( U \models i(\neg B) \), hence \( U \models i(\neg B) \), thus \( U \not\models i(A) \) and, by [Gitman and Hamkins, 2010] (Corollary 8), \( U \) can be taken to be in \( R \), which allows one to conclude that there is an interpretation \( i \) and \( M \in R \) such that \( M \not\models i(A) \).

### Conclusion
Georg Kreisel and George Boolos tackled the VHT problem while focusing on the universe as being a kind of model of reference. Their answers can be modified so as to tackle the KST problem. This paper has proposed a new examination of the KST problem, based on the result that any model of set theory can be seen, as well, as a local universe, because it can be shown

---

⁰ Jon Barwise (1977) "Theories of Models."
to embrace internal models, so that not only truth in any given model of ZFC, but also logical consequence of ZFC w.r.t. any such model, make sense after all. The main thesis advocated here is that a model-scaled treatment of the KST problem has to be favored, because it does not resort to any informal notion of truth in the background universe, does not go beyond ZFC either and still reaches a fully definite (positive) answer. Accordingly, it can benefit from a model-theoretic analysis to give more fine-grained results, which have been drawn up in the second part of this paper.

Actually, the replication according to which any model of ZFC contains another internal one (although, as we have seen, so as not to give rise to any illfounded regression) is not the least adventitious, but on the contrary is part of the status of set theory as a basis for the whole of mathematics (including model theory of set theory itself). This peculiarity must be acknowledged as an essential feature of set theory, and therefore be dealt with from a philosophical point of view. In connection with the existence of internal models in any model of ZFC, a distinction has to be made between truth in a model and truth from the point of view of a model. The use of the seemingly vague and psychologistic notion of “point of view” is no accident, as it arises from the very framework of set theory, and stretches from Skolem’s paradox. Rather than considering it as an unavoidable awkwardness, as the revenge of Skolem’s paradox, I suggest taking it positively. This is not a mere metaphorical way of speaking, but a legitimate concept, whose content can be stated precisely; I have worked towards systematizing it as a semantical dimension per se and comparing it to the usual semantical concept of satisfaction and logical consequence.

Once the notion of being a logical consequence of ZFC from the point of view of a model $M$ of ZFC has been admitted and expressed through the notion of $M$-logical consequence as defined above, a natural question pertains to the connection between logical consequence of ZFC (in the classical sense) and $M$-logical consequence of ZFC for all models $M$ of ZFC (that is, internal logical consequence of ZFC). It has been established that the two properties are equivalent. That equivalence can be viewed as a point partly vindicating the robustness of model-theoretic definitions. Moreover, it is true that any model of ZFC of which all internal models of ZFC satisfy an $L$-sentence, satisfies itself that sentence. So the answer to the model-relativized version of the KST problem is positive and not itself model-relative: Being an $M$-logical consequence of ZFC ensures truth in $M$, for any $M$. Still, the question is amenable to further specification: The kind of kinship that may occur between a single model and one of its internal models varies according to criteria that can be brought out, and gives rise to results which have been set out. So the conception of a model of set theory as a surrogate universe, and accordingly of its internal models as models from the point of view of that surrogate universe, is a conception that can be detailed and developed fruitfully.

The connection between a model and its internal models can be studied in a modal framing as well. Indeed, it is quite natural to think of internal models as accessible worlds, and accordingly to conceive of truth in all internal models as interpreting a notion of necessity of some sort. That presentation, in the form of a modal system (internal modal logic), where modal reflexivity expresses set-theoretic reflection, suggests a new implementation of modal logic and casts new light on models of set theory. It leads to the singling out of classes of models of ZFC, in view of a natural stability condition, and allows the stating of a completeness result. The study of internal models of models of set theory (from different classes of models) holds out hope of further results, whether in modal terms or in purely set-theoretic ones. Those results come to what could be described, not as “set-theoretic geology,” namely the study of possible class models of ZFC of which the universe is a set forcing extension, but as the study of internal set models of ZFC, or “set-theoretic prospecting.”
Notes

1 In this paper, unless otherwise stated, “logically valid” will be taken to mean “true in every structure,” in a Tarskian way, rather than “true by virtue of logical form,” as a maybe more ordinary understanding has it.
2 [Kreisel, 1969], p. 89-91.
5 See [Mostowski, 1951].
6 [Boolos, 1998], p. 83.
7 See [Boolos, 1998], p. 84-85:

The formal definition of supervalidity is this: let $G$ be a sentence of the language of set theory. Select two monadic second-order variables $X$, $Y$. Replace all formulas $u \in v$ in $G$ by formulas $Y(u, v)$. Relativize all quantifiers $\forall v$ and $\exists v$ in the result to the formula $Xv$; that is, replace contexts $\forall v(\ldots)$ by $\forall v(Xv \rightarrow \ldots)$ and contexts $\exists v(\ldots)$ by $\exists v(Xv \land \ldots)$. Quantify universally with respect to $Y$. Take the result as the consequent of a conditional with antecedent $\exists x Xx$. Finally, quantify this conditional universally with respect to $X$. The result is the formalization of the assertion that $G$ is supervalid.

8 Nevertheless, the approach of [Rayo and Uzquiano, 1999] should be mentioned as an attempt, in the wave of Boolos’ plural interpretation, to provide an account of logical consequence for the language of second-order set theory.
9 See [Rayo and Uzquiano, 1999], p. 322.
10 Any class $\mathcal{C}$, considered as the predicate ‘$x \in \mathcal{C}$’, is definable in $L$.
11 See [Drake, 1974], p. 89-98.
13 Indeed, Boolos inductively defines a new predicate $\text{Sat}(R, s, F)$ to the effect that the ordered pairs of second-order variables and sets plurally referred to by $R$ and the assignment $s$ (that assigns a set to each first-order variable) satisfy the formula $F$. See [Boolos, 1998], p. 80-82.
14 See [Jané, 2001], p. 142-143.
15 See [Hamkins, 2012] for a thorough defense and illustration of that view.
16 The multiverse view is one of higher-order realism—Platonism about universes—and I defend it as a realist position asserting actual existence of the alternative set-theoretic universes into which our mathematical tools have allowed us to glimpse.” ([Hamkins, 2012])

17 Any set which happens to belong to the standard model $M$ can be considered in an equivalent way either as a member of $M$ or as a member of the universe $V$. Now if $M = \langle |M|, \in_M \rangle$ and $N = \langle |N|, \in_N \rangle$ are two models of $\text{ZFC}$, $M$ is said to be a substructure of $N$, $M \subseteq N$, if $|M| \subseteq |N|$ and $\forall x, y \in |M| (y \in_M x \rightarrow y \in_N x)$. For any $x \in |M|$, the extension of $x$ in $M$ is the set $x_M = \{y \in |M| : y \in_M x\}$. Hence, $M \subseteq N$ implies that $x_M \subseteq x_N$ for any $x \in |M|$. In case $M \subseteq N$ and $x_M = x_N$ for any $x \in |M|$, $N$ is said to be an end extension of $M$.
18 [Drake, 1974], p. 89-98. See also [Kunen, 1980], p. 38-42 and p. 143-146.
19 Another way of setting out this argument is the following: For any integer $n$, one has (see for example [Kunen, 1980], p. 133-141): $\text{ZFC} \vdash \text{Con}(n \text{ first axioms of } \text{ZFC})$. Now, let $M$ be an $\omega$-standard model of $\text{ZFC}$. For any $n$, if $M \models \omega$ ("$\omega$ is a finite integer"), then $\omega$ is (really) a finite integer. So $M \models \forall x ("x$ is a finite integer" $\rightarrow \text{Con}(x \text{ first axioms of } \text{ZFC}))$, therefore, by compacity (which is indeed a theorem of $\text{ZFC}$): $M \models \text{Con}(\text{ZFC})$.
20 [Jané, 2001], p. 136.
21 Since the Lemma 1.1 does not extend from the case of a single formula (or a finite set of formulas) to the case of an infinite theory $\Gamma$, it is necessary to distinguish, when $M$ is not $\omega$-standard, between $M \models \forall N \models \Gamma$ and $N_M \models \Gamma$.
22 See [Keisler, 1977], p. 70.
23 [Schlipf, 1977], Theorem III.2.6: Let $N$ be a countable recursively saturated model of $\text{ZFC}$. Then there is a non-$\omega$-standard model $M$ of $\text{ZFC}$ such that $N \subseteq |M|$. The definers cannot be "$M \models \forall N (\forall N \models \text{ZFC} \rightarrow \forall N \models \phi \text{)}", inasmuch as a model of $\text{ZFC}$ may not recognize any of its members as a model of $\text{ZFC}$.
24 See [Montague and Vauht, 1959], Theorem 6.8.
25 Applying the completeness theorem, one can deduce that there is proof of $\phi$ from the axioms of $\text{ZFC}$, which is encoded in $M$ by some $M$-proof. Hence $M \models \forall N \models \text{ZFC} \rightarrow \phi \text{), so that (by soundness) one has, for any } N \in |M|: M \models \forall N \models \text{ZFC} \rightarrow \phi \text{ only if } N \not\models \phi$. But, as already noted, this is not sufficient to conclude that $\text{ZFC} \not\models \phi$.
26 This means that there is a formula $\varphi_S(x)$ of $L$ such that $\phi \in S \text{ if } M \models \varphi_S(n(\bar{\phi}))$.
27 See [Wilmer, 1971].
28 See [Enayat, 2005], Remark 2.8.1.
In fact, it is not necessary to require pointwise definability: Any \( \omega_1 \)-well-founded DO model satisfies \( \text{Th}(M) \not\in \text{St}(M) \), where a DO model is a model of ZFC all of whose ordinals are first-order definable. And, since every well-founded model \( M \) of ZFC whose definable ordinals are not cofinal in \( \text{Ord}^M \) contains as a transitive element a DO model \( N \) of \( \text{Th}(M) \) ([Enayat, 2005], Theorem 2.12), one has that, unlike \( M \), \( N \) itself has no elementary equivalent internal model.


See [Barwise, 1975], Exercise 5.11.

See [Makkai, 1977], p. 263.

A Kripke frame allows one to provide a semantics for propositional modal logic. It consists of a set whose elements are called “possible worlds,” endowed with a binary relation between those worlds, the “accessibility relation.” In such a frame, a proposition is possible in a world if it is true in a world accessible from that world; It is necessary in a world if it is true in all worlds accessible from that world.

[Hamkins and Loewe, 2008].

As we have seen, \( M \models \Box \phi \) is equivalent to \( M \models \exists N \forall \gamma \neg N \models \text{ZFC} + \phi \equiv \) only in the case where \( M \) is \( \omega \)-standard.

[Boolos, 1993], p. 170-173.

See [Schlipf, 1977], Corollary III.1.8 and Theorem III.1.9, respectively.

[Gitman and Hamkins, 2010].

[Schlipf, 1977], Theorem III.1.7.

References


