Notre Dame Journal of Formal Logic
Volume 58, Number 1, 2017

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 with respect to 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**.

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

Received June 6, 2011; accepted March 9, 2014
First published online December 14, 2016
2010 Mathematics Subject Classification: Primary 03C62, 03A05; Secondary 03C55, 03C70, 03B45
Keywords: logical validity, truth, informal rigour, Kreisel, Boolos, logical consequence of ZFC, models of set theory, modal logic, recursively saturated structures
 10.1215/00294527-3716058

(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 with respect to 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.) By shifting to the *model-scaled view*, that is, 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 with respect to some L-structure. The predicament can be summarized as shown in Table 1.

In this paper, a close yet different problem will be addressed: *Is any logical consequence of* ZFC *(any sentence of the first-order language of ZFC 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's 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 with respect to a model of ZFC makes no sense.

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 with respect to 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's 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

Table 1 Ways of framing the VHT problem.

	Kreisel–Boolos view	Model-scaled view
Logical validity	OK	?
Truth	?	OK

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 an answer to the KST problem.

1 Can the Universe Be Conceived as a Model?

As 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 under consideration is the formal theory ZFC itself, and in that case it is necessary to bear in mind the systematic replication between, on one hand, some model of ZFC and, on the other, 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 set-theoretic 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 big 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 stated, 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 (i.e., set) models?

1.1 Kreisel In an article,[2] following which developed a whole current of reflection on model-theoretic validity, Georg Kreisel introduced the following notions. For any first-order formula α,

- $\mathrm{Val}\,\alpha := \alpha$ is true in all structures whatsoever;
- $\mathrm{V}\alpha := \alpha$ is true in all structures whose domain is a member of the cumulative hierarchy.

If α is a first-order formula with the symbol E of binary relation as its only nonlogical symbol,

- $\alpha_\in := \alpha$ is true when the quantifiers in α range over all sets and E is replaced by the "real" membership relation.

These notions come up quite naturally as soon as the logical validity of the sentence α is understood as α's being "always true." Indeed, $\mathrm{Val}\,\alpha$ is the direct expression of that idea; $\mathrm{V}\alpha$ is the set-theoretic enregimentation of it and corresponds to logical validity as it is taken in this paper, namely, as truth in every set structure (the usual model-theoretic notion); finally, $\alpha_\in$ means α'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 of what relationship there is between Vα and Val α for a sentence α of the language of first-order logic, Kreisel begins by answering:

> If α is logically valid, then $\alpha_\in$, *i.e.*, (in symbols): Val $\alpha \to \alpha_\in$. But one certainly does not conclude *immediately*: V$\alpha \to \alpha_\in$; for $\alpha_\in$ requires that α be true in the structure consisting of all sets (with the membership relation); its universe is not a set at all. So Vα (α 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, Vα implies Dα (the derivability of α by means of the rules of first-order classical logic), and D$\alpha \to$ Val α may be accepted as a basic property of Val, so that V$\alpha \to$ Val α and finally V$\alpha \leftrightarrow$ Val α follow. On this account, logical validity of α (Vα) entails Val α and thus, by universal instantiation, the truth of α 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 ϕ of L, let "ZFC $\models^+ \phi$" be the relation that obtains when ϕ is true in any set or class structure that models ZFC. (This is the equivalent of Kreisel's Val ϕ.) Then, ZFC $\models \phi$ entails ZFC $\vdash \phi$ (by completeness), which entails ZFC $\models^+ \phi$, which entails in turn $\phi_\in$. Hence, the problem is apparently 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, which makes it difficult not to consider them as true in whatever structure or universe of discourse, whether it be 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 "New Foundations" system. Being a logical consequence of ZFC, on that score, falls short of ensuring 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 *slightly special* model that is the universe of all sets. Now there are at least two reasons to question the very possibility to refer to the 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_Ω [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_Ω nor with the notion of a higher-order object, say a second-order class A: truth in M_Ω is just truth in a model and A is just a subset of M_Ω. When, as in the case of M_Ω 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_Ω cannot be regarded as a well-defined totality. But, accepting this point of view, the notion of truth in M_Ω requires explanation [...].[4]

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 being to justify viewing 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's theorem,[5] one has that "if ϕ is derivable (in pure first-order logic), then ϕ" is derivable in Peano arithmetic and, thus, in ZFC. In other words,

$$\mathrm{ZFC} \vdash \ulcorner \vdash \phi \urcorner \rightarrow \phi.$$

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

$$\mathrm{ZFC} \vdash \ulcorner \models \phi \urcorner \rightarrow \ulcorner \vdash \phi \urcorner.$$

So one gets:

$$\mathrm{ZFC} \vdash \ulcorner \models \phi \urcorner \rightarrow \phi.$$

Hence, one has the following schema: "if ϕ is true in every model, then ϕ." 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, this 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 ϕ is true in every model of ZFC, then ϕ." In fact, Löb's theorem proves that, precisely, such a conditional can be derived only if ϕ is already a theorem of ZFC:

$$\text{ZFC} \vdash \ulcorner \vDash_{\text{ZFC}} \phi \urcorner \to \phi \quad \text{implies ZFC} \vdash \phi.$$

Another solution has to be found.

1.2 Boolos We now turn to another well-known paper, from George Boolos, urging us 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, the logical validity of a sentence of L does not guarantee its truth, that is, 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, logical validity implies truth, that is, truth in the intended model that the universe constitutes implicitly. Boolos's 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:

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

Boolos suggests two ways out of the difficulty: the completeness theorem and the reflection principle. Both can be used 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 to logical validity. Could Boolos's 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.[8] That path would require shifting to second-order set theory or to some extension of it[9] 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's first suggestion will not be pursued beyond Kreisel's modified answer.

Boolos's 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,

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

where "Ord(β)" asserts that β is an ordinal and "φ^β" refers to the relativization of φ to V_β.[10] As a consequence, if some sentence ϕ is false (in the universe), then $\neg\phi$ is true and, hence, is true in some V_α, so ϕ 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_α of the cumulative hierarchy that model ZFC. (ZFC cannot prove its own consistency.) So the falsity of a sentence ϕ does not entail by reflection the existence of a model of ZFC + $\neg\phi$. Hence, Boolos's second solution does not transpose either to the case of the KST problem.

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

Let $\mathcal{V}$ be some fixed countable set whose elements are taken as codes for the variable symbols of L. In addition, let $\ulcorner\in\urcorner$, $\ulcorner=\urcorner$, $\ulcorner\neg\urcorner$, $\ulcorner\vee\urcorner$, and $\ulcorner\exists\urcorner$ be five fixed sets, taken as a code for the signature of L. The following set $\mathcal{F}$ is then defined by induction: $\mathcal{F}_0 = (\{\ulcorner\in\urcorner\} \times \mathcal{V}^2) \cup (\{\ulcorner=\urcorner\} \times \mathcal{V}^2)$, $\mathcal{F}_{n+1} = \mathcal{F}_n \cup (\{\ulcorner\neg\urcorner\} \times \mathcal{F}_n) \cup (\{\ulcorner\vee\urcorner\} \times \mathcal{F}_n^2) \cup ((\{\ulcorner\exists\urcorner\} \times \mathcal{V}) \times \mathcal{F}_n)$, $\mathcal{F} = \bigcup_{n\in\omega} \mathcal{F}_n$. An element $(\ulcorner\in\urcorner, v_1, v_2)$ of $\{\ulcorner\in\urcorner\} \times \mathcal{V}^2$ is written "$v_1\ulcorner\in\urcorner v_2$," and the same obvious convention is adopted for all elements of $\mathcal{F}$. It is thus not difficult to see that, for any formula ϕ of L, there is a unique element $\ulcorner\phi\urcorner$ of $\mathcal{F}$ corresponding to ϕ. Now, let $\mathrm{Form}(x)$ be defined as $x \in \mathcal{F}$. For any $x \in \mathcal{F}$, the set $\mathrm{fv}(x)$ of the free variables of x is readily defined by induction. Then $\mathrm{Sent}(x)$ is defined as $(\mathrm{Form}(x) \wedge \mathrm{fv}(x) = \emptyset)$. Furthermore, a recursively enumerable formula Ax can be built so that, for any x, $\mathrm{Ax}(x)$ if and only if $x = \ulcorner\phi\urcorner$ for some axiom ϕ of ZFC. In addition, $\mathrm{Assign}(y)$ is defined as "y is a map with domain $\mathcal{V}$." One can then add a 2-place symbol $\mathrm{Sat}(x, y)$ to L, so that $\mathrm{Sat}(\ulcorner\phi\urcorner, s)$ should hold when s is an assignment for the variables of L which satisfies ϕ in V. The formulas ϕ for which $\mathrm{Sat}(\ulcorner\phi\urcorner, s)$ could hold should be the original formulas of L not containing "Sat," so that no paradox arises. The axioms for Sat and Tr are:

- $\forall x \forall y(\mathrm{Sat}(x, y) \to \mathrm{Form}(x) \wedge \mathrm{Assign}(y))$;
- the usual inductive clauses for satisfaction: $\mathrm{Sat}(v_1\ulcorner\in\urcorner v_2, s) \leftrightarrow s(v_1) \in s(v_2)$, $\mathrm{Sat}(v_1\ulcorner=\urcorner v_2, s) \leftrightarrow s(v_1) = s(v_2)$, $\mathrm{Sat}(\ulcorner\neg\urcorner u, s) \leftrightarrow \neg\mathrm{Sat}(u, s)$, $\mathrm{Sat}(u\ulcorner\vee\urcorner u', s) \leftrightarrow (\mathrm{Sat}(u, s) \vee \mathrm{Sat}(u', s))$, and $\mathrm{Sat}((\ulcorner\exists\urcorner v)u, s) \leftrightarrow \exists x\, \mathrm{Sat}(u, s[x/s(v)])$;
- $\mathrm{Tr}(u) \leftrightarrow (\mathrm{Sent}(u) \wedge \forall y(\mathrm{Assign}(y) \to \mathrm{Sat}(u, y)))$.

Let S be the conjunction of all these axioms, and let ZFCS = ZFC + S be the resulting system in $L^+ = L \cup \{\mathrm{Sat}, \mathrm{Tr}\}$, 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.[11] This formalization readily extends to L^+. In particular, there is a formula $\Sigma(A, u, s)$ of L^+ to the effect that A is an L^+-structure, u is $\ulcorner\phi\urcorner$ for some formula ϕ of L^+, and ϕ holds in A under the assignment s. Accordingly, there is a formula $\Theta(A, u) = \ulcorner A \vDash \sigma\urcorner$ of L^+ to the effect that A is an L^+-structure, u is $\ulcorner\sigma\urcorner$ for some sentence σ of L^+, and $A \vDash \sigma$. One thus deals with two truth predicates, Tr and Θ. It is noteworthy that ZFCS $\vdash \forall u(\mathrm{Ax}(u) \to \mathrm{Tr}(u))$ and ZFCS $\vdash \forall A \forall u(\Theta(A, \ulcorner\mathrm{Tr}(u)\urcorner) \to \Theta(A, u))$.

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 δ is the least ordinal such that V_δ contains a witness x for v_0 in ψ under the assignment s (namely, a set x for which ψ is satisfied in V_δ when x is assigned to v_0 and s is used for the assignment to the remaining free variables of ψ), and 0 when there is no such witness x at all. Then, for a given α, the closure β of α under F can be defined in L^+ along the lines of the proof for ZFC, and for such β, $\forall v_1 \ldots \forall v_n(\psi \leftrightarrow \psi^{V_\beta})$ is provable in ZFCS. (Here, in the context of ZFCS, "V_β" actually refers to the expansion $\langle V_\beta, \in, \mathrm{Tr} \cap V_\beta \rangle$ of V_β to L^+.) In particular, for any given sentence ϕ of L^+ that is true in V, taking $\psi := (\phi \wedge \forall u(\mathrm{Ax}(u) \to \mathrm{Tr}(u)))$, one gets ZFCS $\vdash \exists \beta(\forall u(\mathrm{Ax}(u) \to \mathrm{Tr}(u)) \wedge \phi)^{V_\beta}$. But[12] ZFCS $\vdash \forall A(\psi^A \leftrightarrow \ulcorner A \vDash \psi \urcorner)$. Hence, ZFCS $\vdash \exists \beta \ulcorner V_\beta \vDash (\forall u(\mathrm{Ax}(u) \to \mathrm{Tr}(u)) \wedge \phi)\urcorner$, and so ZFCS $\vdash \exists \beta(\forall u(\mathrm{Ax}(u) \to \Theta(V_\beta, u)) \wedge \Theta(V_\beta, \ulcorner \phi \urcorner))$. In other words, one has ZFCS $\vdash \mathrm{Tr}(\ulcorner \phi \urcorner) \to \exists \beta \ulcorner V_\beta \vDash \mathrm{ZFC} + \phi \urcorner$. Now, suppose that ϕ is not true. Then ZFCS proves that $\neg\phi$ is true and thus that $V_\beta \vDash \mathrm{ZFC} + \neg\phi$ for some β, and so ϕ 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, which would require second-order machinery, is not needed. In fact, Boolos's paper precisely aims at providing, in terms of plural quantification, the equivalent of a second-order definition of the set of all true sentences of L.[13] The main point is that the proof above, of the existence of the required model V_β, 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 details of the latter theory). This is the main shortcoming of Boolos's modified solution. Indeed, one should argue just from within ZFC, in keeping with the spirit of reflection theorems, which are theorems of ZFC.

Table 2 is a summary of the treatments of the KST problem drawn from Kreisel's and Boolos's original treatments of the VHT problem. The global set-theoretic truth involved in both Kreisel's and Boolos' modified solutions prompts the following predicament: either one defines it as Kreisel does, but then Löb's theorem 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's modified second solution relies on the fact that ZFCS $\vdash \ulcorner \mathrm{ZFC} \vDash \phi \urcorner \to \mathrm{Tr}(\ulcorner \phi \urcorner)$, or equivalently (by completeness) that ZFCS $\vDash \ulcorner \mathrm{ZFC} \vDash \phi \urcorner \to \mathrm{Tr}(\ulcorner \phi \urcorner)$. But then the question naturally arises as to whether such a logical consequence of ZFCS is true itself. Even if that question is

Table 2 KST problem: Is any logical consequence of ZFC true?

	Kreisel's modified view	Boolos's modified view
Truth in the universe	Informal	Formalized through a satisfaction predicate added to the language of ZFC
Answer to the KST problem	Trivialized by Löb's theorem	Requires to shift to ZFCS, a proper extension of ZFC

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's 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, relative 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 formulating 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 consider only 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 be 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 some 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 θ 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's 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 stated, 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"[15] 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.[16] It simply takes literally the idea, expressed by Joel David Hamkins as underpinning the multiverse view, that set theory has become the 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 to 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 with respect to 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 made precise, 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 now has to be set out precisely.

Let V be the background set-theoretic universe, and let $\in$ be 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 $M = \langle |M|, \in_M \rangle$ satisfies the axioms of ZFC, but whose interpretation of membership does not have to coincide with the restriction of $\in$ to $|M|$. When $\in_M \simeq \in \upharpoonright |M|$, M is said to be a *standard* model.[17]

As recalled about Boolos's modified view, the main notions in the metatheory of ZFC can be formalized within L through some usual Gödel numbering.[18] The code of any formula ϕ of L consists then in a sequence $\ulcorner \phi \urcorner$ of numerals and gives rise in any model M of ZFC to an interpretation $\ulcorner \phi \urcorner^M$, 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 $\ulcorner \phi \urcorner^M$ is called an M-formula. It is then possible to define in L the predicate "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 $M \vDash \text{For}(x)$, and an M-proof is an object y in M such that $M \vDash \exists x(\text{For}(x) \wedge \text{Dem}(y, x))$. Moreover, any statement S falling to the basic semantics of L, such as "$N \vDash \phi[s]$," can likewise be coded into a sentence $\ulcorner \text{S} \urcorner$ of L, here $\ulcorner N \vDash \phi[s] \urcorner$.

A given model M is called *ω-standard* if $\in_M$ is transitive and well orders all the finite ordinals of M. Then, for any $a \in |M|$, $M \vDash$ "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 an ω-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-ω-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-ω-standard models is a direct consequence of the compactness theorem for first-order logic.

If M is ω-standard, then the M-formulas (resp., the M-proofs) are in a one-to-one correspondence with the genuine formulas (resp., the genuine proofs) of ZFC. If not, then some M-formulas and M-proofs fail to correspond to any formula or proof of ZFC. Indeed, for any formula ϕ of L, there is a unique element $x = \ulcorner \phi \urcorner$ of $\mathcal{F}$ corresponding to ϕ and thus a unique corresponding M-formula x^M, but the converse is not true in the 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 let N be an element of $|M|$ such that $M \vDash \ulcorner N$ is an L-structure$\urcorner$. This implies that there exist $|N|$, $E^N \in |M|$ such that $M \vDash (N = \langle |N|, E^N \rangle \wedge E^N \subseteq |N| \times |N|)$. One then defines $|N_M| := \{x \in |M| : M \vDash x \in |N|\}$ and $E^N_M := \{(x, y) \in |N_M| \times |N_M| : M \vDash (x, y) \in E^N\}$. The structure $N_M := \langle |N_M|, E^N_M \rangle$ is called the *replica of N in M*. When $M = \langle |M|, \in_M \rangle$ is a *transitive* $\in$*-model* of ZFC (i.e., $\in_M = \in \upharpoonright |M|$ and, for every $x \in |M|$, $x \subseteq M$), the replica N_M of any N in M is isomorphic to N.

Lemma 1.1 ([20]) *For any sentence ϕ of L and any model M of* ZFC*, one has:*

$$M \vDash \ulcorner N \vDash \phi \urcorner \qquad \textit{iff} \qquad N_M \vDash \phi.$$

Proof The proof is by induction on ϕ, which is possible since, by hypothesis, ϕ is a genuine formula of L, so that its interpretation in M is not an M-pseudoformula. Actually, "$N_M \vDash \phi$" and "$M \vDash \ulcorner N \vDash \phi \urcorner$" verify exactly the same recursive clauses. This suffices to conclude. (We remark that, even if N fails to be an L-structure, N_M will be such a structure anyway.) □

Theorem 1.2 ([19], [20]) *Let M be a model of* ZFC. *Then there exists $N \in |M|$ such that $N_M \vDash$* ZFC *(but not necessarily: $M \vDash \ulcorner N \vDash$* ZFC$\urcorner$*).*

Proof We distinguish two cases.

First case: M is ω-standard. In this case, all the formulas of L and all the proofs within ZFC may be coded into elements of M in a transparent way. Since by hypothesis M is ω-standard, there cannot be any M-proof which would not code a real proof. This holds in particular for all ZFC-proofs. Now, we exploit the assumption that there exists a model M of ZFC. The very existence of M implies that ZFC is consistent. Hence, $V \vDash \mathrm{Con(ZFC)}$. This means that there is no proof in 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, $M \vDash \mathrm{Con(ZFC)}$.[19] Since the (formalized version of) the completeness theorem for first-order logic is a theorem of ZFC, it can be deduced that $M \vDash \ulcorner$There exists a model of ZFC$\urcorner$. So there is $N \in |M|$ such that $M \vDash \ulcorner N \vDash \mathrm{ZFC}\urcorner$, and consequently (by Lemma 1.1) $N_M \vDash$ ZFC.

Second case: M is not ω-standard. That is, M contains a nonstandard finite ordinal. The idea is to index all the axioms of ZFC by such nonstandard integers of M and to consider the conjunction of all these axioms, which M thinks to be in finite number. The application of the reflection principle to what amounts, from M's point of view, to a mere finite fragment of ZFC actually gives a result involving the whole of ZFC. So let $(A_i)_{i \in \mathbb{N}}$ be a recursive enumeration of the axioms of ZFC. If $M \vDash \exists\alpha \ulcorner V_\alpha \vDash \mathrm{ZFC}\urcorner$, the claim is proved; so suppose $M \vDash \neg\exists\alpha \ulcorner V_\alpha \vDash \mathrm{ZFC}\urcorner$. By compactness (which is a theorem of ZFC and, hence, true in M), $M \vDash \neg\forall n \exists\alpha \ulcorner V_\alpha \vDash A_0 \wedge A_1 \wedge \cdots \wedge A_n \urcorner$. Now, the formula "$\exists\alpha \ulcorner V_\alpha \vDash A_0 \wedge A_1 \wedge \cdots \wedge A_n\urcorner$" is a formula $\chi(\overline{n})$ of L. One has that $M \vDash \exists n \neg\chi(n)$; thus, $M \vDash \{n \in \omega : \neg\chi(n)\} \neq \emptyset$. Besides, $M \vDash$ "Every nonempty subset of ω has a least element" (since it is also a theorem of ZFC). So there exists $n_0 \in \omega^M$ such that $M \vDash (\neg\chi(\overline{n_0}) \wedge \forall n < \overline{n_0}\ \chi(n))$. On the other hand, $M \vDash A_0 \wedge \cdots \wedge A_n$ for each integer n, so, owing to the reflection principle (also true in M), $M \vDash \chi(\overline{n})$ for each standard integer $n \in \omega^M$. Hence, n_0 is necessarily a nonstandard integer of M. But $M \vDash \chi(\overline{n_0} - 1)$; in other words, $M \vDash \exists\alpha \ulcorner V_\alpha \vDash A_0 \wedge \cdots \wedge A_{n_0 - 1}\urcorner$. Since $n_0 - 1$ is also nonstandard, there finally exists an ordinal α of M such that $(V_\alpha^M)_M \vDash$ ZFC. (The notation "$(V_\alpha^M)_M$" 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 $M \vDash \ulcorner V_\alpha \vDash \mathrm{ZFC}\urcorner$. As a matter of fact, from the point of view of M, V_α^M only satisfies a finite number of axioms of 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 ZFC that it contains fails to satisfy some part of ZFC. In comparison to Lemma 1.1, it appears that M thinks of every axiom of ZFC that N satisfies it and yet does not think that N is a model of ZFC.

Gathering the standard case and the nonstandard one gives the following result: any model M of ZFC contains an element whose replica in M is a model of ZFC. □

The upshot of this result is that any model M contains an internal model N in which all the axioms of ZFC are true, even though the statement that all the axioms of ZFC are true in N may be false in M. This is what happens when the axioms of 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 ZFC as being a particular universe. Indeed, we shall define an *internal model* of ZFC as any model of ZFC isomorphic to N_M, where M is a model of ZFC and N is a member of $|M|$. The previous result ensures that any model M of ZFC has internal models. Hence, it becomes possible to define logical consequence from ZFC with respect to any given model M of ZFC and, thus, to tackle the KST problem at the level of models of ZFC. Before pursuing this, it should be explained how seeing a model of 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. (By 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 which are 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-ω-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 Σ_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 stated, 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 aspects: this sentence is true about any model and, as such, is 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 $\exists^{\mathrm{AI}}$, where "$(\exists^{\mathrm{AI}} x)(\phi(x))$" means "the individuals satisfying $\phi(x)$ are too many to form a set." Then "$(\exists^{\mathrm{AI}} x)(x = x)$" 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."[20]

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 reinterpretation 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, we say that M's structure consists of all the first-order sentences without parameters of the language of ZFC satisfied by M, whereas M's point of view consists of all the conditions of the form "$N_M \models \Gamma$" that are realized, where Γ is any set of sentences.[21] It could be argued that no sentence φ can be said to hold from the point of view of a model M^* of ZFC $+ \neg\,\mathrm{Con}(\mathrm{ZFC})$, even if φ is true *in* M^*. According to M^*, there is no model of ZFC in which ϕ 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 φ 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 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 $M_1' := (M_1)_M \vDash$ ZFC. So the next internal model will be internal to M_1', not to M_1. 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 $\cdots \in |M_2| \in |M_1| \in |M|$ 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, which does not imply that M necessarily says of M_1 that it is a model of ZFC (i.e., $M \vDash \ulcorner M_1 \vDash \text{ZFC}\urcorner$). Otherwise stated, 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 ϕ is a 2-*logical consequence of* ZFC if and only if, for any $M \vDash$ ZFC and any $N \in |M|$, $N_M \vDash$ ZFC implies $N_M \vDash \phi$.

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

Theorem 2.2 ([17]) *Let N be a model of* ZFC. *Then there is a model $M \equiv N$ such that $\exists \Omega \in |M|\ \forall a \in \Omega\ \ M \simeq_a \Omega_M$, where "$M \simeq_a \Omega_M$" means that there is an isomorphism $f : M \to \Omega_M$ such that $\forall x \in |M| \cap a\ \ f(x) = x$.*

Corollary 2.3 *Let ϕ be a sentence of L. Then ϕ is a 2-logical consequence of* ZFC *if and only if it is a logical consequence of* ZFC.

Proof Any logical consequence of ZFC is by definition a 2-logical consequence of ZFC. Conversely, suppose ϕ is a 2-logical consequence of ZFC, and let N be any model of ZFC. By Theorem 2.2, there exists $M \equiv N$, $\Omega \in |M|$, and $a \in \Omega$ such that $M \simeq_a \Omega_M$. So, in particular, one has $M \equiv \Omega_M$. Since by hypothesis $\Omega_M \vDash \phi$, one finally gets $N \vDash \phi$, and this holds for any $N \vDash$ ZFC. □

An alternative definition would consist in identifying a 2-*logical* consequence* of ZFC with an L-sentence that 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 ϕ is a 2-logical consequence* of ZFC, and let M be a model of ZFC. The first-order theory Th(M) of M in L is consistent, and so[22] admits a countable recursively saturated model M'. By a theorem of Schlipf,[23] M' belongs to some model M of ZFC, and so by hypothesis $M' \vDash \phi$, so $\phi \in \text{Th}(M)$, that is, $M \vDash \phi$. Hence, ϕ 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. In 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 analogues, 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, 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 ϕ be an L-sentence and M be a model of ZFC. Then ϕ is called an *M-logical consequence of* ZFC, written ZFC $\vDash_M \phi$, if and only if, for every $N \in |M|$, $N_M \vDash$ ZFC implies $N_M \vDash \phi$.[24]

Definition 2.5 (Internal logical consequence of ZFC) Let ϕ be an L-sentence. Then ϕ is called an *internal logical consequence of* ZFC, written ZFC $\vDash^{\mathrm{i}} \phi$, if and only if ZFC $\vDash_M \phi$ for any model M of ZFC.

The intuitive meaning of ZFC $\vDash_M \phi$ is that ϕ would be a logical consequence of ZFC *were M the background universe*. The intuitive meaning of ZFC $\vDash^{\mathrm{i}} \phi$, then, is that ϕ 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 ZFC $\vDash_M \phi$, that is, with respect to the local prism of the universe that constitutes some model M of ZFC. Now, a few things naturally deserve to be studied.

The first one bears on the relation between ZFC $\vDash_M \phi$ and $M \vDash \phi$. One may take as an example the natural model of ZFC that is V_θ, where θ is the first strongly inaccessible ordinal. In this case, one knows, by a result of Montague and Vaught,[25] that there is an ordinal $\theta^* < \theta$ such that $\langle V_{\theta^*}, \in\rangle \equiv \langle V_\theta, \in\rangle$, with $\langle V_{\theta^*}, \in\rangle \in V_\theta$. Suppose ZFC $\vDash_{V_\theta} \phi$. Then, by definition, $(V_{\theta^*})_{V_\theta} \vDash \phi$. But V_θ is a transitive $\in$-structure, and so $(V_{\theta^*})_{V_\theta} \simeq V_{\theta^*}$. Consequently, $V_{\theta^*} \vDash \phi$, and finally $V_\theta \vDash \phi$. Hence, ZFC $\vDash_{V_\theta} \phi$ implies $V_\theta \vDash \phi$ for any L-sentence ϕ.

Now we call a cardinal γ a *universe cardinal* if and only if $V_\gamma \vDash$ ZFC, and let γ_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 κ one gets $V_\kappa \vDash$ ZFC + WAU. But of course, by minimality, $V_{\gamma_0} \nvDash$ WAU, so that in fact (because $\gamma_0 < \kappa$) $V_{\gamma_0} \in V_\kappa$ and $V_\kappa \vDash \ulcorner V_{\gamma_0} \nvDash \text{WAU}\urcorner$, resulting in $(V_{\gamma_0})_{V_\kappa} \vDash$ ZFC + ¬WAU. Consequently, $M \vDash \phi$ does not entail ZFC $\vDash_M \phi$.

The converse general question, as to whether ZFC $\vDash_M \phi$ entails $M \vDash \phi$, receives a positive answer.

Theorem 2.6 *Let ϕ be an L-sentence, and let M be a model of* ZFC *such that* ZFC $\vDash_M \phi$. *Then $M \vDash \phi$.*

Proof Suppose that $M \nvDash \phi$. This proves that ZFC + $\neg\phi$ is consistent. The proof of Theorem 1.2 can then be rewritten, with ZFC + $\neg\phi$ replacing ZFC. One thus concludes that there exists $N \in |M|$ such that $N_M \vDash$ ZFC + $\neg\phi$; hence, ZFC $\nvDash_M \phi$. Accordingly, ZFC $\vDash_M \phi$ implies $M \vDash \phi$. □

As noticed, the two conditions ZFC $\vDash_M \phi$ and $M \vDash \phi$ are not equivalent in general. This fact is still compatible with ZFC $\vDash_M \phi$ for *each* M if and only if $M \vDash \phi$ for *each* M, as what follows proves.

Corollary 2.7 *Let ϕ be an L-sentence. Then* ZFC $\vDash \phi$ *if and only if* ZFC $\vDash^{\mathrm{i}} \phi$.

Proof By generalization over M, the previous theorem guarantees that ZFC $\vDash^{\mathrm{i}} \phi$ implies ZFC $\vDash \phi$. Conversely, suppose that ϕ is a logical consequence of ZFC. Then, in particular, $N_M \vDash \phi$ for any internal model N_M of ZFC, so, by definition, ZFC $\vDash_M \phi$, and this holds for any model M of ZFC.[26] □

Table 3 Answers to the KST problem.

	Kreisel's modified view	Boolos's modified view	Model-scaled view	Generalization to every M
ϕ is a logical consequence of ZFC	ZFC $\models^{+} \phi$	ZFC $\models \phi$	ZFC $\models_M \phi$	ZFC $\models^{\mathrm{i}} \phi$
ϕ is true	ϕ is informally true	$\ulcorner\phi\urcorner$ is in the extension of the truth predicate added to L	ϕ is true in M	ϕ is true in every M
Answer to the KST question	Yes	Yes	Yes	Yes (equivalence)

We are now in a position to get back to the KST problem. Set at the level of models of ZFC, that problem consists in the following two questions: Is any L-sentence true in M if it is an M-logical consequence of ZFC? Is any L-sentence true in all models of ZFC if it is an internal logical consequence of ZFC? Both answers are positive, as noted in Table 3, along with Kreisel's answer and Boolos's answer.

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 ZFC (as expressed by the previous column). Truth in every model of ZFC, and thus derivability in ZFC, does not purport to capture what it means for a sentence of the language L of ZFC to be true, but only generalizes truth in M in the same way as ZFC $\models^{\mathrm{i}} \phi$ generalizes M-logical consequence to any M. Thus, the bottom rightmost 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 The framework of internal models allows one to address and to 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 just 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 it is 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, Theorem 2.7 supports the conclusion that nothing new will ever come out of the consideration of all models of ZFC whatsoever and, thus, 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 $\models_M$, simply defined by the following: for each recursively enumerable set of sentences T in L and each sentence ϕ of L, $T \models_M \phi$ if and only if, for any internal L-structure N_M in M, if $N_M \models T$, then $N_M \models \phi$. But it results from the contrapositive of Theorem 2.6 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 one of 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 $M \vDash \ulcorner N \vDash$ the first n axioms of ZFC$\urcorner$, 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 ϕ such that $M \vDash \phi$, there exists some internal model N_M of M in which ϕ is true as well (since ZFC $\nvDash_M \neg\phi$). In which cases is it possible to assume the internal model N_M to be the same for all sentences ϕ 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 ϕ of L are now coded by numerals $n(\phi)$ 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:*

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

where $\mathrm{st}(A) = \{n \in \mathbb{N} : M \vDash \overline{n} \in A\}$. *Then there is $N \in |M|$ such that $N_M \equiv M$ if and only if* $\mathrm{Th}(M) \in \mathrm{St}(M)$.

Proof Suppose that there exists $N \in |M|$ such that $N_M \equiv M$. Then $\mathrm{Th}(M) = \mathrm{Th}(N_M)$. Now, $N_M \vDash \sigma$ if and only if $M \vDash \ulcorner N \vDash \sigma \urcorner$ for any sentence σ of L. Besides, one can write a formula $\mathrm{Sat}(N, x)$ in two variables N, x formalizing the statement that N is an L-structure and that x codes an L-sentence true in N. Therefore, $A_N = \{\alpha \in \omega^M : M \vDash \mathrm{Sat}(N, x)[x = \alpha]\}$ 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 $\mathrm{Th}(N_M) = \mathrm{st}(A_N)$ belongs to $\mathrm{St}(M)$, and so does $\mathrm{Th}(M)$.

Conversely, suppose that $\mathrm{Th}(M)$ is in the standard system of M: $\mathrm{Th}(M) = \{n \in \mathbb{N} : M \vDash s(n)\} = \mathrm{st}(B)$ for some $B \in |M|$ with $M \vDash B \subseteq \omega$. Hence, $n \in \mathrm{Th}(M)$ if and only if $n \in \mathbb{N}$ and $M \vDash \overline{n} \in B$. One can assume that, for any $x \in B$, $M \vDash \ulcorner x$ is the code of some sentence of $L\urcorner$. Besides, there exists a function g, definable in L, such that, for any finite family F of formulas of L, $n(\bigwedge_{\phi \in F} \phi) = g(\langle n(\phi) : \phi \in F\rangle)$, where the integers $n(\phi)$ are ordered increasingly. Then one states the following definition in M: $M \vDash \forall x \in B\ j(x) = g(\langle y : y \in B, y \leq x\rangle)$. Now suppose $M \vDash \forall x \in B\ \exists N\ \mathrm{Sat}(N, j(x))$. Then, by compactness (which holds in M), $M \vDash \exists N\ \forall x \in B\ \mathrm{Sat}(N, x)$. In particular, there exists $N \in |M|$ such that, for any $n \in \mathrm{Th}(M)$, $M \vDash \mathrm{Sat}(N, \overline{n})$, and so $N_M \equiv M$. On the other hand, if $M \nvDash \forall x \in B\ \exists N\ \mathrm{Sat}(N, j(x))$, then one has that $M \vDash \exists x_0 \in B\ (\forall x \in B\ (x < x_0 \to \exists N\ \mathrm{Sat}(N, j(x))) \wedge \neg\exists N\ \mathrm{Sat}(N, j(x_0)))$. For any $n \in \mathrm{Th}(M)$, $M \vDash \bigwedge_{i \in \mathrm{Th}(M),\, i \leq n} s(i)$, so by reflection and because $\{\overline{n}^M : n \in \mathrm{Th}(M)\}$ is an initial segment of B, $M \vDash \exists \alpha_n\ \mathrm{Sat}(V_{\alpha_n}, s(n))$, which means that $M \vDash \overline{n} < x_0$. Thus, x_0 is nonstandard. But since $M \vDash \exists N\ \mathrm{Sat}(N, j(x_0 - 1))$, once again there exists $N \in |M|$ such that, for any $n \in \mathrm{Th}(M)$, $M \vDash \mathrm{Sat}(N, \overline{n})$, hence such that $N_M \equiv M$. □

The upshot of this result is that for any transitive $\in$-model $\langle |M|, \in \rangle$ of ZFC, there is $x \in |M|$ such that $\langle x, \in \rangle \equiv \langle |M|, \in \rangle$ if and only if, for any M-definable subtheory S of $\mathrm{Th}(M)$,[27] $M \vDash SM(\mathrm{S})$, where "$SM(\mathrm{S})$" is a shorthand for the sentence (in the language of 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 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 ZFC cannot be in M's standard system. As a matter of fact, suppose that $\mathrm{Th}(M)$ is in $\mathrm{St}(M)$, which means that $\mathrm{Th}(M)$ 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, $M \vDash \phi(x)$ if and only if $M \vDash \neg s(x)$. Since s is definable, $\phi(x)$ is a formula without parameters. By Kleene's fixed point theorem, there is a sentence ψ such that $\mathrm{ZFC} \vdash \psi \leftrightarrow \phi(\overline{n(\psi)})$. But then $M \vDash \psi$ if and only if $M \vDash \neg\psi$. Since it is a fact (established by Ali Enayat)[29] that every countable model of 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 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 $M \vDash$ "α 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-ω-standard. The idea would be to establish, for some nonstandard integer n_0 of M, that $M \vDash \forall \varphi(\vec{x}) \in \Sigma_{n_0}\ \forall \vec{a} \in |N|^k\ (\phi(\vec{a}) \leftrightarrow \ulcorner N \vDash \phi(\vec{a})\urcorner)$. But this cannot be obtained by any application of the reflection schema. In fact, the set of sentences true in (M, V_α^M) (that is, in the expansion of M into an interpretation of $L \cup \{\bar{a} : a \in V_\alpha^M\}$) 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* ZFC *and α an ordinal of M. Then there exists $N \in |M|$ such that (i) $V_\alpha^M \subseteq |N|$ and (ii) $(N_M, V_\alpha^M) \equiv (M, V_\alpha^M)$, if and only if $\exists s : (V_\alpha^M)^{<\omega^M} \to \wp(\omega^M)$, $s \in |M|$, such that $\forall \vec{a} \in V_\alpha^M$* $\mathrm{st}(s(\vec{a})) = \mathrm{Th}(M, \vec{a})$.

Again, the condition of this theorem really divides the models of ZFC into two camps. Indeed, the minimal model M_0 of 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 satisfies any of the first n formulas of L in 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 ϕ_n's is also realizable by some ordinal β of M. For such ordinal β, 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 analogue of second-order set theory. In fact, its objects are only classes, sets (denoted with lowercase 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 κ. 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: $\forall W_1 \ldots \forall W_n \exists Y(x \in Y \leftrightarrow \exists X(x \in X \wedge \phi(x, W_1, \ldots, W_n)))$ 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 $\forall x(\exists X\ x \in X \to x \in V)$. The fact that the theory MK is strictly stronger than ZFC comes from the impredicativity of the comprehension schema. It is a proper (nonconservative) extension of ZFC. But it does not preclude Theorems 1.2 and 2.8 and Corollaries 2.3 and 2.7 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 if and only if 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 urelement (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 $\mathbb{A}$ (i.e., to some model of KPU) and the statement of a completeness theorem (Barwise completeness theorem)[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 $L_{\mathbb{A}}$ corresponding to an admissible set $\mathbb{A}$ is defined as the set of all formulas φ of $L_{\infty,\omega}$ whose codes belong to $\mathbb{A}$. One can then focus on results establishing (1) the conditions on which an $L_{\mathbb{A}}$-sentence is satisfied in a structure internal to $\mathbb{A}$, or in the reverse direction, (2) the conditions on which an $L_{\mathbb{A}}$-sentence is valid with respect to an admissible set $\mathbb{B}$ to which $\mathbb{A}$ is internal. Because admissible sets are supposed to be transitive $\in$-models of KPU, any admissible set $\mathbb{A} \in \mathbb{B}$ coincides with the corresponding internal model $\mathbb{A}_{\mathbb{B}}$.

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

As for question (2), we have the following fact.[32]

Proposition 2.10 *Let $\mathbb{A}$, $\mathbb{B}$ be two admissible sets such that $\mathbb{A} \in |\mathbb{B}|$ and $\mathbb{B} \models \ulcorner \mathbb{A}$ is countable$\urcorner$. Then, for any $L_{\mathbb{A}}$-sentence ϕ, one has that $\mathbb{B} \models \ulcorner \phi$ is valid$\urcorner$ if and only if $\mathbb{A} \models \exists P \ulcorner P$ is a proof of $\phi\urcorner$.*

Proof Suppose $\mathbb{B} \models \ulcorner\phi$ is valid$\urcorner$. Then (by the Barwise completeness theorem, which can be derived within KPU), $\mathbb{B} \models \ulcorner\exists P\ P$ is an infinitary proof of $\phi\urcorner$; hence, there really exists such an infinitary proof P of ϕ; hence (again owing to the Barwise completeness theorem) $\mathbb{A} \models \exists P \ulcorner P$ is a proof of $\phi\urcorner$. Conversely, suppose $\mathbb{A} \models \exists P \ulcorner P$ is a proof of $\phi\urcorner$. Then ϕ is valid. Now, the set $\mathrm{Sat}(\mathbb{A})$ of all the valid $L_{\mathbb{A}}$-sentences is Σ_1 on $\mathbb{A}$ and, hence, an object of $\mathbb{B}$ (because $\mathbb{HYP}_{\mathbb{A}} \subseteq \mathbb{B}$). By absoluteness, $\phi \in \mathrm{Sat}(\mathbb{A})$ implies that $\mathbb{B} \models \phi \in \mathrm{Sat}(\mathbb{A})$ and, thus, that $\mathbb{B} \models \ulcorner\phi$ is valid$\urcorner$. □

Corollary 2.11 *Let $\mathbb{A}$ and $\mathbb{B}$ be admissible sets, with the same hypotheses as above. Then, for any $L_{\mathbb{A}}$-sentence ϕ, ZFC $\models_{\mathbb{B}} \phi$ if and only if ϕ 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 $\mathbb{A}$ be an admissible set, and let $\mathbb{A}' \in |\mathbb{A}|$. According to the Barwise completeness theorem, the set $E_{\mathbb{A}}$ of valid $\mathbb{A}$-sentences of $L_{\infty,\omega}$ is $\Sigma_{\mathbb{A}}$ on $\mathbb{A}$, which means that there exists a Σ-formula $F(x, \vec{a})$ with parameters in $\mathbb{A}$ such that $\sigma \in E_{\mathbb{A}}$ implies $\mathbb{A} \models F(\ulcorner\sigma\urcorner, \vec{a})$. Besides, $L_{\mathbb{A}'}$ is a $\Delta_{\mathbb{A}}$-subset of $\mathbb{A}$. Likewise, the main syntactic and semantical notions relative to $L_{\mathbb{A}'}$ are Δ on $\mathbb{A}$. In particular, there is a $\Delta_{\mathbb{A}}$-formula $S(x, y)$ such that, for any $L_{\mathbb{A}}$-sentence σ, $\mathbb{A} \models S(\mathbb{A}', \ulcorner\sigma\urcorner)$ if and only if $\mathbb{A}' \models \sigma$. Moreover, the members of $\mathbb{A}$ which are themselves $L_{\mathbb{A}}$-structures form also a $\Delta_{\mathbb{A}}$-subset of $\mathbb{A}$. Let $G_{\mathbb{A}}(x)$ denote the corresponding $\Delta_{\mathbb{A}}$-formula. Now, we say that an admissible set $\mathbb{A}$ is *reflective* if, for any $L_{\mathbb{A}}$-sentence σ, σ is a valid sentence if and only if $\mathbb{A} \models S(\mathbb{A}', \ulcorner\sigma\urcorner)$ for any $L_{\mathbb{A}}$-structure $\mathbb{A}' \in \mathbb{A}$, that is, if and only if $\mathbb{A} \models \forall \mathbb{A}' (G_{\mathbb{A}}(\mathbb{A}') \to S(\mathbb{A}', \ulcorner\sigma\urcorner)$. As this last formula is $\Pi_{\mathbb{A}}$, the set $E_{\mathbb{A}}$ of all valid $L_{\mathbb{A}}$-sentences is $\Delta_{\mathbb{A}}$. One may partially characterize the admissible sets $\mathbb{A}$ such that $E_{\mathbb{A}}$ is $\Sigma_{\mathbb{A}}$ and not $\Delta_{\mathbb{A}}$. As a matter of fact, we put, for any admissible $\mathbb{A}$, $\mathbb{A}^+ = \cap\{\mathbb{B} : \mathbb{A} \in \mathbb{B},\ \mathbb{B} \text{ is admissible}\}$; $\mathbb{A}^+$ is itself an admissible set. An admissible set of this form is called a *next admissible set*. For any next admissible set $\mathbb{A}$, $E_{\mathbb{A}}$ is provably not $\Delta_{\mathbb{A}}$.[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 chosen to settle the KST problem. That setting has important conceptual advantages, 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 is 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 if and only if 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 differences with the latter are 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 as follows.

Definition 3.1 Let M be a model of ZFC, and let ϕ be a sentence of the language L of ZFC. Then "$M \vDash \Diamond\phi$" is a shorthand for the existence of some $N \in |M|$ such that $N_M \vDash \mathrm{ZFC} + \phi$.

Defining $\Box\phi$ as $\neg\Diamond\neg\phi$ as usual, one has that $M \vDash \Box\phi$ is equivalent to $\mathrm{ZFC} \vDash_M \phi$.

An alternative presentation could be useful in making it appear that in fact all models of ZFC are involved in making up a huge modal frame. We say that a model N of ZFC is *accessible** from a given model M if and only if, for any L-sentence ϕ, $\mathrm{ZFC} \vDash_M \phi$ implies $N \vDash \phi$. Then, we say that $\Box^*\phi$ is true in M if and only if ϕ is true in every model of ZFC accessible* from M.

Lemma 3.2 *For any L-sentence ϕ and any model M of* ZFC*, $M \vDash \Box^*\phi$ if and only if $M \vDash \Box\phi$.*

Proof The forward implication is immediate. For the converse implication, it is sufficient to remark that $\mathrm{ZFC} \vDash_M \phi$ implies that $N \vDash \phi$ for any model N accessible* from M. □

Proposition 3.3 *There is no formula $P(x)$ of L such that, for any L-sentence ϕ and any model M of* ZFC*, $M \vDash \Diamond\phi$ if and only if $M \vDash P(\overline{n(\phi)})$.*

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

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, $\Diamond A$ being defined as $\neg\Box\neg A$. A formal system of modal logic in L' is said to be *normal* if the following hold: (i) all tautologies of propositional calculus; (ii) the axiom $\mathbf{K} = \Box(p \to q) \to (\Box p \to \Box q)$; (iii) the rule of uniform substitution (if $A(p_1, \dots, p_n)$ is a theorem, so is $A[B_k/p_k]$) ($k = 1, \dots, n$); (iv) modus ponens; (v) the rule of necessitation (if A is a theorem, so is $\Box A$). Important modal axioms are: $p \to \Diamond p$, or, equivalently, $\Box p \to p$ (**T**), $\Diamond\Diamond p \to \Diamond p$ (**4**), $\Diamond p \to \Box\Diamond p$ (**5**), and $\Box(\Box p \to p) \to \Box p$ (**GL**). S4 is the normal modal system that contains the axioms **T** and **4**; S5 is S4 + **5**.

Definition 3.4 An *interpretation* i of L' into L is a map that assigns to each propositional variable p an arbitrary L-sentence.

For any such interpretation i and any L-structure M, it is possible to define inductively "$M \vDash i(A)$" for every modal formula A as follows:

- $M \vDash i(\neg A)$ if and only if $M \nvDash i(A)$;
- $M \vDash i((A \wedge B))$ if and only if $M \vDash i(A)$ and $M \vDash i(B)$;
- $M \vDash i(\Box A)$ if and only if ZFC $\vDash_M i(A)$.

Definition 3.5 Given a formula A of L' and a model M of ZFC, A is *modal-internally valid in* M, written $M \vDash_{\text{IML}} A$, if and only if, for any interpretation i of L' into L, $M \vDash i(A)$.

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 IML.

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 $M \vDash i(A(p_1, \ldots, p_n))$, for any interpretation i and any model M. There is no such interpretation as $i^* : p_k \mapsto i(B_k)$, but one can proceed by induction on ϕ. The only case worth considering is $A = \Box B$. So suppose $M \vDash i(\Box B(p_1, \ldots, p_n))$, for any interpretation i and any model M. In other words, $N_M \vDash i(B(p_1, \ldots, p_n))$ for any internal model N_M of any model M. Let us consider a fixed model M; as a consequence of Theorem 2.6, $M \vDash i(B(p_1, \ldots, p_n))$ and, by the induction hypothesis, $M \vDash i(B[B_k/p_k])$. This is true for any model M and so, in particular, for any internal model N_M of M. Thus, ZFC $\vDash_M i(B[B_k/p_k])$; hence, $M \vDash i(\Box(B[B_k/p_k]))$, that is, $M \vDash i(A[B_k/p_k])$, which is true again for any interpretation i and any model M. Finally, as to the necessitation rule, suppose $M \vDash i(A)$ for any model M and, thus, in particular for any internal model N_M of any model M. Then $M \vDash i(\Box A)$, and this holds for any interpretation i and any model M. □

One also has the following result.

Proposition 3.7 *One has that* **T** $\in$ IML.

Proof Let M and i be any model of ZFC and any interpretation of L' into L, respectively, and suppose that $M \nvDash i(p)$. It then follows from Theorem 2.6 that ZFC $\nvDash_M i(p)$. In other words, $M \vDash i(\Box p)$ only if $M \vDash i(p)$, for any interpretation i. So $M \vDash_{\text{IML}}$ **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 *One has that* **GL** $\notin$ IML.

Proof Let M be a model of ZFC, and let i be some interpretation of L' into L. By definition, $M \vDash i(\Box(\Box p \rightarrow p))$ if and only if $N_M \vDash i(\Box p \rightarrow p)$ for any model N_M of ZFC internal to M. But Theorem 1.2 and thus the axiom **T** (looked at as a schematic set-theoretic fact) are themselves theorems of ZFC and, hence, are IML-valid in any such N_M. So $M \vDash_{\text{IML}} \Box(\Box p \rightarrow p)$, and so $M \vDash i(\Box(\Box p \rightarrow p))$, whatever i may be. Still, $M \vDash i(\Box p)$ does not hold as soon as $i(p) \notin \mathrm{Th}(M)$, since in that case there exists an internal model N_M of M for which $N_M \vDash \mathrm{ZFC} + \neg i(p)$, so that $\mathrm{ZFC} \nvDash_M i(p)$. □

Definition 3.9 Given a class $\mathcal{K}$ of models of ZFC, "$M \vDash^{\mathcal{K}} i(A)$" is defined as in Definition 3.4, except for the last clause, which is replaced with $M \vDash^{\mathcal{K}} i(\Box A)$ if and only if, for every internal model N_M of M *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 $\mathcal{K} \vDash^*_{\mathrm{IML}} A$.

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 that* $\mathcal{T} \vDash^*_{\mathrm{IML}}$ S4.

Proof Firstly (by Propositions 3.6 and 3.7), $\mathcal{T} \vDash^*_{\mathrm{IML}}$ KT. Furthermore, owing to the Jensen–Karp theorem,[37] 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 $M \vDash i(\Diamond p)$, even if $N \notin |M|$. 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 $M \vDash^{\mathcal{T}} i(\Box\Box p)$. This holds for any interpretation i and any M in $\mathcal{T}$, so $\mathcal{T} \vDash^*_{\mathrm{IML}}$ **4**. □

Proposition 3.11 *Let $\mathcal{S}$ be the class of all standard models of* ZFC*. One has that* $\mathcal{S} \nvDash^*_{\mathrm{IML}}$ **5**.

Proof By minimality, the minimal model M_0 satisfies ¬SM(ZFC), where "SM(ZFC)" is the L-sentence asserting the existence of a standard model of ZFC. Now, M_0 is isomorphic to L^M_γ for some standard model M of ZFC, where γ is the ordinal of all M-ordinals. By the Mostowski collapsing lemma, M can be taken to be a transitive $\in$-model, so that M_0 is the real L_γ. Since any internal model of M_0 is nonstandard, $M_0 \vDash^{\mathcal{S}} \neg i(\Diamond p)$, 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 $(M_0)_{M_2} = M_0 \nvDash^{\mathcal{S}} i(\Diamond p)$, so that $M_2 \nvDash^{\mathcal{S}} i(\Box\Diamond p)$. Still, considering the internal (standard) model $M_1 = L_{\gamma+1} = (M_1)_{M_2}$ of M_2, one has that $M_1 \vDash$ SM(ZFC), since $M_1 \vDash \ulcorner M_0 \vDash \mathrm{ZFC}\urcorner$, so that $M_2 \vDash^{\mathcal{S}} i(\Diamond p)$ for any interpretation i assigning SM(ZFC) to p. Hence, $M_2 \nvDash^{\mathcal{S}} i(\mathbf{5})$ 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 Definitions 3.4 and 3.5).

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 $\mathcal{K} \models_{\text{IML}} A$.

Remark 3.14 For any strongly downward stable class $\mathcal{K}$ of models of ZFC, $\mathcal{K} \models^*_{\text{IML}} A$ if and only if $\mathcal{K} \models_{\text{IML}} A$.

The class $\mathcal{M}$ of all models of ZFC is strongly stable for trivial reasons. A better understanding of IML requires getting hold of other stable classes.

Lemma 3.15 *One has that $\mathcal{S}$ and $\mathcal{T}$ 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-ω-standard models of* ZFC *are both strongly downward stable.*

Proof Let $M \in \mathcal{R}$ and $N \in |M|$ be such that $N_M \models$ ZFC. First, N_M is obviously countable. Second, M, as any recursively saturated model of ZFC, is non-ω-standard. Indeed, the type $p(x) = \{\text{Ord}(x)\} \cup \{x > \overline{n} : n \in \mathbb{N}\} \cup \{x < \omega\}$ 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-ω-standard model of ZFC is recursively saturated. Indeed, consider a non-ω-standard model M of ZFC, $N \in |M|$ such that $N_M \models$ ZFC, and $p(x) = (\phi_n(x))_{n\in\mathbb{N}}$ a recursive type in L. Suppose that $p(x)$ is finitely realizable in N_M: $N_M \models \exists x \bigwedge_{k<n} \phi_k(x)$ for any $n \in \mathbb{N}$. In other words, $M \models \ulcorner N \models \exists x \bigwedge_{k<n} \phi_k(x)\urcorner$, which can be written $M \models \Phi(\overline{N}, \overline{n})$ with $\Phi \in L$. Now, by compactness, $M \models \Phi(\overline{N}, \overline{c})$ 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 $\overline{k}^M < c$, and thus for any $k \in \mathbb{N}$ whatsoever, $N_M \models \phi_k(x)[b]$. 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-ω-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 $N \models \exists R\phi(R, \vec{m})$ with $M \prec N$ and $\vec{m} \in |M|^k$, $M \models \exists R\phi(R, \vec{m})$, where "$R$" stands for a second-order variable.

In other words, M is resplendent if and only if M satisfies a Σ^1_1-formula $\exists R\phi(R, \vec{x})$ as soon as it satisfies all its first-order consequences $\{\psi(\vec{x}) \in L_0 : \phi(R, \vec{x}) \models \psi(\vec{x})\}$. Any resplendent structure is recursively saturated. In addition, Jon Barwise and Jean-Pierre Ressayre proved independently that any countable recursively saturated structure is resplendent.[38] So in fact $\mathcal{R}$ coincides with the class of all countable resplendent (non-ω-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."[39] Gitman and Hamkins's 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 *One has that* $\mathcal{R} \vDash_{\text{IML}}$ 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 $N = \langle |N|, E\rangle \in |M|$ such that $N_M \vDash \text{ZFC} + i(\Diamond p)$. So there exists $\alpha = \langle \nu, \eta\rangle \in |N_M|$ such that $\alpha_{N_M} \vDash \text{ZFC} + \phi$, where $\phi = i(p)$. Now, for any sentence θ of L, consider the following sentence θ^* of $L(P^{(1)}, R^{(2)})$: $(\ulcorner\langle P, R\rangle$ is an L-structure$\urcorner \wedge \exists x \forall y (Py \leftrightarrow y \in x) \wedge \exists r \forall y \forall z (Ryz \leftrightarrow (y,z) \in r) \wedge \ulcorner\langle P, R\rangle \vDash \theta\urcorner)$. Let T_0 be any finite fragment of $\text{ZFC} + \phi$, and let σ_0 be the conjunction of all the members of T_0. One has that $N_M \vDash \ulcorner\alpha \vDash \sigma_0\urcorner$, so $\langle N_M, \nu_{N_M}, \eta_{N_M}\rangle \vDash \sigma_0^*(P, R)$, and so $M \vDash \ulcorner N \vDash (\ulcorner\nu$ and η are sets $\urcorner \wedge \ulcorner\langle \nu, \eta\rangle \vDash \sigma_0\urcorner)\urcorner$. Hence (for the same reasons as in the proof of Lemma 1.1), $M \vDash (\ulcorner\nu_N$ and η_N are sets $\urcorner \wedge \ulcorner\langle \nu_N, \eta_N\rangle \vDash \sigma_0\urcorner)$, with $\nu_N = \{x \in |N| : N \vDash x \in \nu\}$ and $\eta_N = \{(x,y) \in \upsilon_N \times \upsilon_N : N \vDash (x,y) \in \eta\}$. (By comprehension, ν_N and η_N, as interpreted in M, are indeed members of $|M|$.) Thus, $\langle M, (\nu_N)_M, (\eta_N)_M\rangle = \langle M, \nu_{N_M}, \eta_{N_M}\rangle \vDash \sigma_0^*(P, R)$; ν_{N_M} and η_{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 $(\text{ZFC} + \phi)^*$. Now, by a result due to Jon Barwise,[40] 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 $(\text{ZFC} + \phi)^*$ in $L(P, R)$. Thus, there are $A \subseteq |M|$ and $B \subseteq |M| \times |M|$ such that $\langle M, A, B\rangle \vDash (\text{ZFC} + \phi)^*$. This implies that there are a and b in $|M|$ such that $a_M = A$, $b_M = B$, and $M \vDash \ulcorner\langle a, b\rangle \vDash \theta\urcorner$ for any $\theta \in \text{ZFC} + \phi$. So $\langle a_M, b_M\rangle \vDash \text{ZFC} + \phi$, with $\phi = i(p)$, and so $M \vDash i(\Diamond p)$. □

Definition 3.19 A modal theory Λ is IML-*complete with respect to a class* $\mathcal{K}$ *of models of* ZFC if, for any formula A of L', $\Lambda \vdash A$ if and only if $\mathcal{K} \vDash_{\text{IML}} A$.

Theorem 3.20 *One has that* S4 *is* IML-*complete with respect to* $\mathcal{R}$.

Proof Owing to the preceding theorem, it only remains to show that if S4 $\nvdash A$, then there is an interpretation i and $M \in \mathcal{R}$ such that $M \nvDash i(A)$. One proceeds by induction on A. Given a propositional variable p, S4 $\nvdash p$ but, for an interpretation i assigning to p an L-sentence which is not a theorem of ZFC, there is a model N of ZFC such that $N \nvDash i(p)$. 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 $\text{Th}(N)$ of N. For $A = \neg B$, suppose that, for any interpretation i of L' and any model M in $\mathcal{R}$, $M \vDash i(\neg B)$, that is, $M \vDash \neg i(B)$. As a consequence, for any fixed interpretation i, the theory ZFC $+ i(B)$ is inconsistent. Otherwise, there are a model $\overline{M}$ of that theory and a countable recursively saturated model of $\text{Th}(\overline{M})$, and thus a model $M \in \mathcal{R}$ of $i(B)$. So, under the assumption that ZFC is consistent, ZFC $\vDash \neg i(B)$ for any interpretation i of L' into L. This means that B is an antilogy and thus that A is a theorem of S4. For $A = (B \wedge C)$, S4 $\nvdash (B \wedge C)$ implies S4 $\nvdash B$ or S4 $\nvdash C$. For the sake of argument, say S4 $\nvdash B$. But then (by the induction hypothesis) there is an interpretation i and $M \in \mathcal{R}$ such that $M \nvDash i(B)$, thus such that $M \nvDash (i(B) \wedge i(C))$, and so such that $M \nvDash i(A)$. Finally, consider $A = \Box B$, where B is any formula, the induction hypothesis being that if S4 $\nvdash B$, then there is an interpretation i and $M \in \mathcal{R}$ such that $M \nvDash i(B)$. Suppose that S4 $\nvdash A$. Then (owing to the necessitation rule) S4 $\nvdash B$, so (by the induction hypothesis) there is an interpretation i and $M \in \mathcal{R}$ such that $M \vDash \neg i(B)$. Now, by [18, Theorem III.2.6], there is a non-ω-standard model U of ZFC and $M' \in |U|$ such that $M \simeq M'_U$. So $U \vDash i(\Diamond\neg B)$, hence $U \vDash i(\neg\Box B)$, thus $U \nvDash i(A)$, and by [6, Corollary 8], U can be taken to be in $\mathcal{R}$, which allows one to conclude that there is an interpretation i and $M \in \mathcal{R}$ such that $M \nvDash i(A)$. □

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 to embrace internal models, so that not only truth in any given model of ZFC, but also logical consequence of ZFC with respect to 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 ill-founded 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. This 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. See [11, pp. 89–91].

3. See [11, pp. 90–91].

4. See [21, pp. 279–80].

5. See [15].

6. See [3, p. 83].

7. See [3, pp. 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 $Y\langle u, v\rangle$. Relativize all quantifiers $\forall v$ and $\exists v$ in the result to the formula Xv; that is, replace contexts $\forall v(\ldots)$ by $\forall v(Xv \to \cdots)$ and contexts $\exists v(\ldots)$ by $\exists v(Xv \wedge \ldots)$. 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 [16] should be mentioned as an attempt, in the wave of Boolos's plural interpretation, to provide an account of logical consequence for the language of second-order set theory.

9. See [16, p. 322].

10. Any class V_β, considered as the predicate "$x \in V_\beta$," is definable in L.

11. See [4, pp. 89–98].

12. See for instance [12, Lemma 10.1, p. 144].

13. Indeed, Boolos inductively defines a new predicate $\mathrm{Sat}(R, s, F)$ to the effect that the ordered pairs of second-order variables and sets plurally referred to by R and the assignment s (which assigns a set to each first-order variable) satisfy the formula F. See [3, pp. 80–82].

14. See [9, pp. 142–43].

15. See [7] for a thorough defense and illustration of that view.

16. See [7]: "The multiverse view is one of higher-order realism—Platonism about universes—and I defend it as a realist position asserting the actual existence of the alternative set-theoretic universes into which our mathematical tools have allowed us to glimpse."

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 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 $x \in |M|$. When $M \subseteq N$ and $x_M = x_N$ for any $x \in |M|$, N is said to be an *end extension* of M.

18. See [4, pp. 89–98]. See also [12, pp. 38–42, 143–46].

19. Another way of setting out this argument is as follows. For any integer n, one has (see for example [12, pp. 133–41]) ZFC $\vDash$ Con(n first axioms of ZFC). Now, let M be an ω-standard model of ZFC. For any a, if $M \vDash$ "a is a finite integer", then a is (really) a finite integer. So $M \vDash \forall x$ ("x is a finite integer" $\rightarrow$ Con(x first axioms of ZFC)), therefore, by compacity (which is indeed a theorem of ZFC), $M \vDash$ Con(ZFC).

20. See [9, p. 136].

21. Since 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 Γ, it is necessary to distinguish, when M is not ω-standard, between $M \vDash \ulcorner N \vDash \Gamma \urcorner$ and $N_M \vDash \Gamma$.

22. See [10, p. 70].

23. See [18, Theorem III.2.6]: Let N be a countable recursively saturated model of ZFC. Then there is a non-ω-standard model M of ZFC such that $N \in |M|$.

24. The definiens cannot be "$M \models \forall N(\ulcorner N \models \mathrm{ZFC}\urcorner \to \ulcorner N \models \phi\urcorner)$," inasmuch as a model of ZFC may not recognize any of its members as a model of ZFC.

25. See [14, Theorem 6.8].

26. Applying the completeness theorem, one can deduce that there is a proof of ϕ from the axioms of ZFC, which is encoded in M by some M-proof. Hence, $M \models \ulcorner \mathrm{ZFC} \vdash \phi\urcorner$, so that (by soundness) one has, for any $N \in |M|$, $M \models \ulcorner N \models \mathrm{ZFC}\urcorner$ only if $N \models \phi$. But, as already noted, this is not sufficient to conclude that $\mathrm{ZFC} \models_M \phi$.

27. This means that there is a formula $\varphi_S(x)$ of L such that $\phi \in S$ if and only if $M \models \varphi_S(\overline{n(\phi)})$.

28. See [22].

29. See [5, Remark 2.8.1].

30. In fact, it is not necessary to require pointwise definability: any ω_1-well-founded DO model satisfies $\mathrm{Th}(M) \notin \mathrm{St}(M)$, 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 Ord^M contains as a transitive element a DO model N of $\mathrm{Th}(M)$ (see [5, Theorem 2.12]), one has that, unlike M, N itself has no elementary equivalent internal model.

31. See [1, pp. 98–99].

32. See [1, Exercise 5.11].

33. See [13, p. 263].

34. 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.

35. See [8].

36. As we have seen, $M \models \Diamond\phi$ is equivalent to $M \models \exists N \ulcorner N \models \mathrm{ZFC} + \phi\urcorner$ only in the case where M is ω-standard.

37. See [2, pp. 170–73].

38. See [18, Corollary III.1.8 and Theorem III.1.9], respectively.

39. See [6].

40. See [18, Theorem III.1.7].

References

[1] Barwise, J., *Admissible Sets and Structures*, Springer, Berlin, 1975. MR 0424560. 76

[2] Boolos, G., *The Logic of Provability*, Cambridge University Press, Cambridge, 1993. MR 1260008. 76

[3] Boolos, G., *Logic, Logic, and Logic*, Harvard University Press, Cambridge, Mass., 1998. MR 1675856. 74, 75

[4] Drake, F. R., *Set Theory: An Introduction to Large Cardinals*, vol. 76 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam, 1974. MR 3408725. 75

[5] Enayat, A., "Models of set theory with definable ordinals," *Archive for Mathematical Logic*, vol. 44 (2005), pp. 363–85. MR 2140616. DOI 10.1007/s00153-004-0256-9. 76

[6] Gitman, V., and J. D. Hamkins, "A natural model of the multiverse axioms," *Notre Dame Journal of Formal Logic*, vol. 51 (2010), pp. 475–84. MR 2741838. DOI 10.1215/00294527-2010-030. 73, 76

[7] Hamkins, J. D., "The set-theoretical multiverse," *Review of Symbolic Logic*, vol. 5 (2012), pp. 416–49. MR 2970696. DOI 10.1017/S1755020311000359. 75

[8] Hamkins, J. D., and B. Löwe, "The modal logic of forcing," *Transactions of the American Mathematical Society*, vol. 360 (2008), pp. 1793–817. MR 2366963. DOI 10.1090/S0002-9947-07-04297-3. 76

[9] Jané, I., "Reflections on Skolem's relativity of set-theoretical concepts," *Philosophia Mathematica (3)*, vol. 9 (2001), pp. 129–53. MR 1830693. DOI 10.1093/philmat/9.2.129. 75

[10] Keisler, H. J., "Fundamentals of Model Theory," pp. 47–103 in *Handbook of Mathematical Logic, Part A*, edited by J. Barwise, North-Holland, Amsterdam, 1977. MR 0491125. 75

[11] Kreisel, G., "Informal rigour and completeness proofs," pp. 78–94 in *The Philosophy of Mathematics*, edited by J. Hintikka, Oxford University Press, London, 1969. 74

[12] Kunen, K., *Set Theory: An Introduction to Independence Proofs*, Elsevier, Amsterdam, 1980. MR 0597342. 75

[13] Makkai, M., "Admissible sets and infinitary logic," pp. 233–81 in *Handbook of Mathematical Logic, Part A*, edited by J. Barwise, North-Holland, Amsterdam, 1977. MR 0491125. 76

[14] Montague, R., and R. L. Vaught, "Natural models of set theories," *Fundamenta Mathematicae*, vol. 47 (1959), pp. 219–42. MR 0124203. 76

[15] Mostowski, A., "Some impredicative definitions in the axiomatic set-theory," *Fundamenta Mathematicae*, vol. 37 (1951), pp. 111–24. MR 0041083. 74

[16] Rayo, A., and G. Uzquiano, "Toward a theory of second-order consequence," *Notre Dame Journal of Formal Logic*, vol. 40 (1999), pp. 315–25. MR 1845629. DOI 10.1305/ndjfl/1022615612. 75

[17] Ressayre, J.-P., "Modèles non standard et sous-systèmes remarquables de ZF," pp. 47–147 in *Modèles non standard en arithmétique et théorie des ensembles*, edited by J.-P. Reyssaire and A. J. Wilkie, Publications mathématiques de l'Université Paris VII, Paris, 1983. Zbl 0631.03048. 61

[18] Schlipf, J. S., "A guide to the identification of admissible sets above structures," *Annals of Mathematical Logic*, vol. 12 (1977), pp. 151–92. MR 0485330. 73, 76, 77

[19] Schlipf, J. S., "Toward model theory through recursive saturation," *Journal of Symbolic Logic*, vol. 43 (1978), pp. 183–206. MR 0491130. 58

[20] Suzuki, Y., and G. Wilmers, "Non-standard models for set theory," pp. 278–314 in *The Proceedings of the Bertrand Russell Memorial Logic Conference (Uldum, 1971)*, edited by J. L. Bell, J. C. Cole, G. Priest, and A. B. Slomson, Bertrand Russell Memorial Logic Conference, University of Leeds, Leeds, 1973. MR 0351814. 57, 58

[21] Tait, W. W., "Foundations of set theory," pp. 273–90 in *Truth in Mathematics*, edited by H. G. Dales and G. Oliveri, Oxford University Press, New York, 1998. MR 1688347. 74

[22] Wilmers, G., "Internally standard set theories," *Fundamenta Mathematicae*, vol. 71 (1971), pp. 93–102. MR 0309730. 76

Acknowledgments

I wish to thank Joel David Hamkins, to whom Section 2.4 owes a lot, for very helpful advice. I also wish to thank all anonymous referees for their very constructive comments.

Department of Philosophy
Université Paris Nanterre
Nanterre
France
bhalimi@u-paris10.fr