Can Modalities Save Naive Set Theory?

Peter Fritz[1], Harvey Lederman[2], Tiankai Liu[3] and Dana Scott[4]

[1]IFIKK, University of Oslo
[2]Department of Philosophy, New York University
[3]Department of Mathematics, University of Utah
[4]Visiting Scholar, University of California, Berkeley

DRAFT of Monday 11th July, 2016 at 05:56

To the memory of Prof. Grigori Mints, Stanford University
Born: June 7, 1939, St. Petersburg, Russia
Died: May 29, 2014, Palo Alto, California

1 Introduction

Considering only pure sets, the naive set comprehension principle says, for any condition, that there is a set containing all and only the sets satisfying the condition. In first-order logic, this can be formulated as the following schematic principle, where φ may be any formula in which y does not occur freely:

$$\exists y \forall x (x \in y \leftrightarrow \varphi) \tag{Comp}$$

Russell's paradox shows that the instance obtained by letting φ be $x \notin x$ is inconsistent in classical logic. One response to the paradox is to restrict naive set comprehension by ruling out this and other problematic instances: only for each of some *special conditions* is it claimed there is a set containing all and only the sets satisfying the condition. Many well-known set theories can be understood as instances of this generic response, differing in how they understand *special*. For example, the axiom schema of separation in Zermelo-Fraenkel set theory (ZF) restricts set comprehension to conditions which contain, as a conjunct, the condition of being a member of some given set:

$$\exists y \forall x (x \in y \leftrightarrow \varphi \wedge x \in z) \tag{Sep}$$

Similarly, in Quine's New Foundations (NF) set comprehension is restricted to conditions which are stratified, where φ is stratified just in case there is a mapping f from individual variables to natural numbers such that for each subformula of φ of the form $x \in y$, $f(y) = f(x) + 1$ and for each subformula of φ of the form $x = y$, $f(x) = f(y)$. Both of these responses block Russell's paradox by ruling out the condition $x \notin x$.

Must every restriction of naive comprehension take the form of simply ruling out certain instances? In this paper, we suggest and explore a different approach. As we have seen, standard set comprehension axioms restrict attention to some *special conditions*: for each of these special conditions, they provide for the existence of a set containing all and only the sets which satisfy it. Our comprehension axiom restricts not the conditions we are allowed to consider, but rather the way in which the sets in question satisfy a given condition: for *every* condition, our axiom asserts the existence of a set containing all and only the sets satisfying that condition *in a special way*. Although we will suggest some more specific ways of understanding *in a special way* below, we will investigate the fruitfulness of this idea abstractly, considering whether there is any way for the qualification *in a special way* to behave according to which the correspondingly restricted version of set comprehension yields an interesting theory of sets. Consequently, we enrich the first-order language of pure set theory by an uninterpreted unary sentential operator $\Box$ which formalizes *in a special way*. Using this operator, the new comprehension principle can be stated as follows:

$$\exists y \forall x (x \in y \leftrightarrow \Box\varphi). \tag{Comp$\Box$}$$

Since the instance for φ being $x \notin x$ is not ruled out, how might Russell's paradox be blocked on this approach? The set claimed to exist by this instance – call it the Russell set – must contain all and only the sets which satisfy, in a special way, the condition of not containing themselves, and so in particular, the Russell set must contain itself if and only if it satisfies, in a special way, the condition of not containing itself. *Prima facie*, this can be verified if the Russell set satisfies the condition of not containing itself without satisfying it in a special way.

Standard set theories such as ZFC are extremely well-developed; why should we set aside these theories and investigate an unfamiliar principle such as (Comp$\Box$)? Our main motivation is exploratory: we wish to see what results from this simple and natural alternative to more standard comprehension principles. The fact that one response to the set-theoretic paradoxes is known to lead to a rich mathematical theory should not prevent us from attempting to understand the consequences of other conceptually simple, mathematically natural replies.

A second motivation for investigating alternative restrictions of naive set comprehension is that separation requires a relatively large number of additional axioms to yield a strong set theory. What if (Comp$\Box$), perhaps together with an axiom of extensionality, gives rise to a useful set theory? Extensionality is the following simple principle:

$$\forall x (x \in y \leftrightarrow x \in z) \rightarrow y = z \tag{Ext}$$

Such an axiomatic system would constitute a considerable gain in simplicity over ZFC. (Simplicity is of course not enough: the axioms of NF, which are just extensionality and set comprehension restricted to stratified conditions, are considerably simpler than the axioms of ZF, but most set theorists still choose to work in the latter.)

A third reason for restricting set comprehension as in (Comp$\Box$) is that this restriction fits certain views in the philosophy of mathematics and logic, on suitable ways of understanding the qualification "in a special way". One example is fictionalism, which will be discussed below. For another example, we can understand *in a special way* as *determinately*. To motivate this idea, consider an analogy to philosophical discussions of truth. Those who wish to preserve classical logic in the face of the liar paradox sometimes introduce an operator to be read "determinately", which distinguishes paradox-free claims from claims which are paradoxical. Equivalences such as $\varphi \leftrightarrow T\ulcorner\varphi\urcorner$ hold (where $\ulcorner\varphi\urcorner$ is the Gödel number of φ) if it is determinate whether φ holds, but not necessarily otherwise.[1] Given the well-known parallels between the liar paradox and Russell's paradox, it is natural to wonder whether a version of this approach can be extended to the set theoretic paradoxes. On this approach, comprehension might be restricted as in (Comp$\Box$), requiring only that for every condition φ, there is a set containing all and only the sets which are *determinately* φ. As described in the previous paragraph, this strategy might block a version of Russell's paradox: if the Russell set does not belong to itself, it may nevertheless not be determinate that it does not belong to itself.

In the modal logics used below, $\Box\top$ is equivalent to $\top$ (the tautological constant), and so the corresponding instance of (Comp$\Box$) asserts the existence of a universal set. A final reason for exploring this comprehension principle is therefore an interest in set theories with a universal set; see Forster (1995) for motivations for admitting a universal set, and for an overview of existing approaches to such set theories.

In this paper, we approach set theories based on (Comp$\Box$) from an abstract perspective, considering different principles governing $\Box$ in the form of different quantified modal logics, asking whether (Comp$\Box$) is consistent in them, and if so, whether any interesting set theory emerges from it. In keeping with the second motivation, of potentially finding axioms systems which are simpler than the standard axioms of ZFC, we will focus on evaluating the strength of modalized comprehension on its own, or in conjunction with (Ext). It would also be interesting to consider the prospects of developing modal set theories based on these as well as additional axioms, but such considerations will mostly be beyond the scope of this paper.

Section 2 formally introduces the quantified modal logics which we will use to investigate (Comp$\Box$). The relevant logics will be standard free extensions of arbitrary normal modal logics. Standard classes of Kripke models will be introduced, and later used in an instrumental capacity to prove consistency and other underivability results.

Section 3 considers (Comp$\Box$) in the strong modal logic **S5**, showing that (Comp$\Box$) is consistent in **S5** and so *a fortiorti* in all weaker modal logics. Unfortunately, while the principle is consistent in these modal logics, the set theory it gives rise to is very weak. Indeed, we show that in **S5**, the non-modal consequences of (Comp$\Box$) are precisely characterized by the extensional

[1] See Bacon (2015, 237); a more standard approach uses a predicate of sentences instead of an operator; for references and problems with this approach, see the first section of Bacon's paper.

theory which states, for any finite number of elements, that there is the set containing them, and the set containing every other element, and that this result extends to any theory obtained by adding any further non-modal principles. Thus there seems to be little hope of restricting naive comprehension using a single modal operator while at the same time preserving the conceptual simplicity of a theory with (modal) comprehension and extensionality alone.

From the results just sketched, it is natural to conclude that (Comp$\Box$) is too weak. How could it be strengthened in a way which preserves the intuitive motivation with which we started? One idea can be motivated by our earlier example of a theory based on interpreting the operator $\Box$ as "determinately". On this interpretation, for every condition, (Comp$\Box$) asserts the existence of a set of all and only the sets which determinately satisfy the condition. But we may alter this principle so that for every condition, it states the existence of a set such that for all sets, *determinately*, the former contains the latter if and only if the latter *determinately* satisfies a given condition. Although we have articulated this idea using a particular interpretation of $\Box$, it can of course be stated fully abstractly. Formally the principle can be stated as:

$$\exists y \forall x \Box (x \in y \leftrightarrow \Box \varphi) \qquad (\Box\text{Comp}\Box)$$

($\Box$Comp$\Box$) turns out to be far stronger than (Comp$\Box$); indeed, as we show in section 4.1, it overshoots its mark, as it is inconsistent in the relatively weak modal logic **T**, and so *a fortiori* in its extensions. In response, one could use two different modal operators in ($\Box$Comp$\Box$), but since this drastically increases the space of available options, we don't consider it here. Instead, we explore the more restricted option of replacing the second occurrence of $\Box$ in ($\Box$Comp$\Box$) by a string of modal operators and negations. In section 4.2, we show exhaustively that such variants are inconsistent in **S4**.

The results obtained so far leave open the consistency of ($\Box$Comp$\Box$) in modal logics weaker than **T**. Such logics do not prove the axiom (T): $\Box\varphi \to \varphi$. This might seem non-negotiable, given our earlier schematic reading of the modality as satisfying a condition *in a special way*: whatever satisfies a condition in a special way satisfies it simpliciter. But there are ways of reading $\Box$ on which it is natural to give up (T). One such reading arises naturally from the position of fictionalism in the philosophy of mathematics.[2] One of the fundamental questions in the philosophy of mathematics – some might say, *the* fundamental question – concerns the existence of mathematical objects. In what sense are there sets, for example? An important proposal is that mathematical objects do not in fact exist, although they exist according to a literally false, but nevertheless useful theory. This is a "fictionalist" approach to the philosophy of mathematics not because it absurdly claims that proofs from mathematical axioms are somehow merely fictionally correct, but because it claims that the things such as sets which we take to exist when we are writing and reading mathematics do not really exist. Reading $\Box$ as "in the fiction (of there

[2]See, e.g. Field (1980, 1989), Rosen (2001), Yablo (2001). Balaguer (2008) provides a helpful overview, with many more references.

being sets)”, $\Box\varphi \to \varphi$ is naturally rejected – fictionalists hold that in the fiction, there are sets, but in fact, there are none.

One axis of variation among versions of fictionalism concerns what we take to be the relevant fiction. At least on some versions of fictionalism, it's up to us – it depends on which fiction we choose. We may therefore choose to work in a fiction specified by the axioms of some standard set theory. But fictionalists may also explore more adventurous fictions; in particular, it is an intriguing idea to let the fiction specify what sets there are by making reference to the fiction itself. Therefore, one claim we may choose to be part of the fiction is that for any condition, there is a set containing all and only the sets satisfying the condition in the fiction. Of course, if we read $\Box$ is “in the fiction”, this is just (Comp$\Box$), and an analogous reading exists for ($\Box$Comp$\Box$). We are not aware of any proposal of this kind in the literature, but it seems to us to be sufficiently congenial to fictionalism to be motivated by the general fictionalist project in the philosophy of mathematics.

Section 4.3 considers ($\Box$Comp$\Box$) in the context of modal logics which do not prove (T). We first observe that this principle, as well as (Comp$\Box$), are trivially consistent if the background modal logic does not prove the axiom (D): $\Box\varphi \to \Diamond\varphi$. Reading $\Box$ as “according to the fiction”, we may read $\Diamond$ as “consistent with the fiction”; thus (D) expresses the very natural idea that what holds according to the fiction is consistent with the fiction. We then show that ($\Box$Comp$\Box$) is also inconsistent if **KD** is extended by one of the well-known axioms (4), (B) and (5).

Section ($\Box$Comp$\Box$) covers most of the standard normal modal logics and shows that among them, the principle is only consistent in very weak modal logics, and trivial if consistent. Section 5 therefore returns to the original principle (Comp$\Box$), and considers its prospects in normal modal logics which are not contained in **S5**. Section 5.1 shows that it is inconsistent in any proper extension of **S5**. Section 5.2 considers the logic **KDDc**, axiomatized by (D) and its converse (Dc): $\Diamond\varphi \to \Box\varphi$. On the fictionalist interpretation of the modal operators, this expresses the natural constraint that the fiction be complete: what is consistent with the fiction must hold according to the fiction. We show that (Comp$\Box$) is consistent in **KDDc**; the models used in this proof indicate that in an informal sense, the resultant set theories are significantly stronger than the set theory **S5** + (Comp$\Box$). Indeed, the resulting theory can be consistently extended by principles with which ($\Box$Comp$\Box$) follows from ($\Box$Comp$\Box$), and so the latter principle is also non-trivially consistent in **KDDc**.

We owe our investigation of these questions to the late Grigori “Grisha” Mints. In October of 2009 at Stanford University, Mints asked Scott whether a naive set theory could be consistent in modal logic. At that time Scott could not answer the consistency question, and neither could Mints, though they both agreed that a set theory based on (Comp$\Box$) would probably be very weak. In November 2014, Scott received a notice from Carnegie Mellon that there would be a philosophy seminar on a naive set theory by Lederman (see Field *et al.* (forthcoming)). Scott wrote him for his paper and said, “By the way, there is this question of Grisha Mints, and I wonder if you have an opinion?” Lederman sent back a sketch of a proof of inconsistency for a slightly strengthened version of (□Comp□), which did not quite work out, but the exchange became the basis for sections 4.1-4.2. In the first draft of the paper, Scott and Lederman left open the consistency of (Comp□), although they observed that it was not inconsistent by the analogue of the Russell set alone. Scott and Lederman tried out several model possibilities for the consistency of that principle, without success. In March of 2015 Liu approached them with a related model, which after a small correction gave a consistency proof; a few days later, Fritz approached them with essentially the same model. Fritz then provided the results of section 5, and wrote the present version of the paper.

Before delving into the formal details of this paper, we will situate our comprehension principles (Comp□) and (□Comp□) in the existing literature. Modalized comprehension principles have been studied in a number of different settings.[3] One is intensional higher-order logic (see, e.g., Gallin (1975, p. 77) or Zalta (1988, p. 22)), where a syntactic distinction between types allows for an unrestricted comprehension principle. Such discussions usually work with models with constant (first- and higher-order) domains; for discussions of comprehension principles appropriate for variable domains of all types, see Williamson (2013, chapter 6.3–6.4) and Fritz & Goodman (forthcoming, section 5).

Another form of modal comprehension principles occurs in modal set theories which are obtained by modalizing common set theories. Such a system for metaphysical necessity is presented in Fine (1981); systems for epistemic modalities were developed by several authors in the 1980s – see the contributions by Myhill, Goodman and Ščedrov to Shapiro (1985), or the references in Goodman (1990).

In the lecture at which Mints posed his original question, 2010, Scott presented his Modal ZF, which uses the following modalization of the axiom schema of separation:

$$\exists y \Box \forall x (x \in y \leftrightarrow x \in u \wedge \varphi). \qquad \text{(MZF Comp)}$$

All of the modal comprehension principles mentioned so far differ fundamentally from the naive principles in that they are modalizations of comprehension principles which do not give rise to the Russell paradox, either through employing type distinctions or through restrictions on the formulas with which instances may be obtained. Modalizing naive comprehension in order to make it consistent has been less widespread, but several such strands can be identified in the literature. The first uses modality to make the iterative conception of set explicit by reformulating comprehension to say that at some stage, there is a set defined by a given condition, using a

[3] There are also a number of ways in which modal logic has been used in connection with set theory without considering modalized comprehension principles. E.g., in provability logic, propositional modal logics can be used to study abstract features of a complex predicate expressing "it is provable in ZFC that" in certain first-order theories; see, e.g., Solovay (1976) and Boolos (1995). Propositional modal logics have also been used to characterize certain properties of forcing, roughly interpreting □ is expressing "it is true in all forcing extensions that"; see, e.g., Hamkins (2003), Hamkins & Woodin (2004) and Hamkins & Löwe (2008). For a different connection between modal logic and forcing see Smullyan & Fitting (1996, Part III) and Fitting (2003). For yet another set-theoretic modality, see Blass (1990). Such combinations of modal logic and set theory are less closely related to the topic of this paper than the variant modal set comprehension principles discussed in the following.

possibility operator to formalize "at some stage". Pioneered by Parsons (1983), such principles have been recently investigated in Studd (2013) and Linnebo (2013); see also Hellman (1989) and Linnebo (2010).

The second strand, and the closest to our own work, goes back to Fitch (1966, 1967a).[4] Fitch works in a language extended by a term-forming operator, which creates a term from a variable and a formula, and which we may write as $\{:\}$. He considers the principle:

$$x \in \{x : \varphi\} \leftrightarrow \Box\varphi \tag{Abst$\Box$}$$

in a strengthening of a quantified modal logic based on the propositional modal logic **KD**. In Fitch's extension of **KD**, while one can necessitate instances of theorems of predicate logic, instances of (Abst$\Box$), (K), (D) and the converse Barcan formula, one cannot necessitate formulas proven using instances of the Barcan formula, (T) or (4), although these too are described as "axioms". A concrete example of the failure of necessitation is given in Fitch (1967a, 102-3) (cf. (Fitch, 1967b, p. 107)). Fitch proves the consistency of his system using techniques similar to those used later by Gilmore (1967) and Kripke (1975), but which Fitch had developed as early as Fitch (1942), cf. Fitch (1948, 1963). It is easy to see that both (Comp$\Box$) and ($\Box$Comp$\Box$) are derivable from (Abst$\Box$) by standard quantifier rules and necessitation, so Fitch establishes the consistency of these two principles in his extension of **KD**; a similar result is obtained in section 5.2 by a different construction. Despite these commonalities, Fitch's system differs from the ones investigated below in a number of crucial details. One difference is that we investigate (Comp$\Box$) and ($\Box$Comp$\Box$) separately, rather than (Abst$\Box$), from which both of these principles follow. Another difference is Fitch's assumption of the necessity of membership ($x \in y \rightarrow \Box(x \in y)$), to which we return in section 3. For further discussion of Fitch's work, and the history of his comparative neglect by other authors in the field, see (Cantini, 2009, section 4.2).

A third strand starts with Aczel & Feferman (1980), who save the naive comprehension principle from inconsistency by replacing its material biconditional with a binary intensional operator. Feferman (1984, Section 12) shows that if we abbreviate $\varphi \equiv \top$ (where $\equiv$ is their binary intensional operator) by $\Box\varphi$, we obtain from his models a consistency proof for a modal logic which validates (Comp$\Box$), along with the Barcan formula, the modal schemes (K), (T) and (4), and a number of other strong principles. Like Fitch, Aczel and Feferman consider a principle which features set-abstracts, and not merely the quantifiers as in more standard comprehension principles. Unlike Fitch's system, however, the logic in Feferman's system is not a normal modal logic; as Feferman observes (1984, p. 100) there are false instances of $\Box(\varphi \vee \neg\varphi)$ in his model construction. The principle $\neg\Box\varphi \rightarrow \Box\neg\Box\varphi$ also has false instances; the logic is thus substantially weaker than the one we show to be consistent with (Comp$\Box$). In Feferman's model, as in Fitch's (but not in ours) $x \in y \rightarrow \Box(x \in y)$.

[4]The original technical paper Fitch (1967a) contained an error, pointed out in a review by Rundle (1969); a correction appeared as Fitch (1970).

A fourth strand is the following comprehension principle, proposed by Krajíček (1987):

$$\exists y \forall x((\Box x \in y \leftrightarrow \Box \varphi) \wedge (\Box \neg x \in y \leftrightarrow \Box \neg \varphi)) \tag{MCA}$$

Krajíček proves that this principle is inconsistent in **S5**, and it seems still to be an open problem whether it is consistent in the relatively weak modal logic **T** (see Krajíček (1988) and Kaye (1993)).

2 Logics and Models

In this section, we set out the logics and models with which we will investigate our modalized set comprehension principles. Let $\mathcal{L}_0^\Box$ be a language of propositional modal logic, based on a countably infinite set of proposition letters, from which formulas are constructed using Boolean operators $\neg$ and $\wedge$ and a unary modal operator $\Box$. A set of formulas in this language is a *normal modal logic* if it contains all propositional tautologies and the distributivity axiom $(K) = \Box(p \to q) \to (\Box p \to \Box q)$, and is closed under the rules of modus ponens, uniform substitution and necessitation. The normal modal logic *axiomatized* by formulas $A_1, \ldots, A_n$, written $\mathbf{KA_1 \ldots A_n}$, is the smallest normal modal logic containing $A_1, \ldots, A_n$.

Let $\mathcal{L}_1(\in)$ be a first-order language based on a countably infinite set of individual variables, using two binary relation symbols $=$ and $\in$ to form atomic predications, from which formulas are constructed using Boolean operators $\neg$ and $\wedge$, and universal quantifiers $\forall x$. Let $\mathcal{L}_1^\Box(\in)$ be the extension of this language by $\Box$. Other common symbols, such as $\vee$, $\to$, $\exists$, $\Diamond$, $\neq$ and $\notin$ will be used as metalinguistic abbreviations in the usual way, and Ex will be used to abbreviate $\exists y(y = x)$, where y is the first variable distinct from x in some given order of the variables. Let $\mathcal{S}_1(\in)$ and $\mathcal{S}_1^\Box(\in)$ be the sets of sentences, i.e., closed formulas, of these langauges.

For any normal modal logic Λ, let $Q\Lambda$ be the set of $\mathcal{L}_1^\Box$-formulas derivable in the following axiomatic calculus, adapted from Hughes & Cresswell (1996, chapters 16 & 17):

(Λ)	any substitution instance of a theorem of Λ
$(\forall 1E)$	$(\forall x\varphi \wedge Ey) \to \varphi[y/x]$
$(\forall^{\to})$	$\forall x(\varphi \to \psi) \to (\forall x\varphi \to \forall x\psi)$
(VQ)	$\varphi \leftrightarrow \forall x\varphi$, provided x is not free in φ
(UE)	$\forall xEx$
$(I1)$	$x = x$
$(I2)$	$x = y \to (\varphi \to \psi)$, where φ and ψ differ only in that φ has free x in places where ψ has free y
(LNI)	$x \neq y \to \Box(x \neq y)$
(MP)	From φ and $\varphi \to \psi$, derive ψ.
(N)	From φ, derive $\Box\varphi$.
(UG)	From φ, derive $\forall x\varphi$.
$(UGL\forall^n)$	From $\varphi_1 \to \Box(\varphi_2 \to \cdots \to \Box(\varphi_n \to \Box\psi)\ldots)$, derive $\varphi_1 \to \Box(\varphi_2 \to \cdots \to \Box(\varphi_n \to \Box\forall x\psi)\ldots)$, where x is not free in $\varphi_1, \ldots, \varphi_n$.

For any further axioms or axiom schemas $A_1, \ldots, A_n$, let $Q\Lambda + A_1 + \cdots + A_n$ be the set of formulas derivable in the axiomatic calculus obtained by adding $A_1, \ldots, A_n$ to the above.

The axioms and rules of $Q\Lambda$ may look unnatural and overly complicated. The reason for choosing this system is that (i) it includes classical first-order logic, in the sense that all *sentences* of standard first-order logic are derivable, (ii) it smoothly combines with arbitrary normal modal logics and (iii) it does not prove the following two schematic principles, known respectively as the Barcan formula and its converse:

(BF)	$\forall x\Box\varphi \to \Box\forall x\varphi$
(CBF)	$\Box\forall x\varphi \to \forall x\Box\varphi$

These two principles have been at the heart of debates in the metaphysics of necessity and possibility. If the modal operators $\Box$ and $\Diamond$ are read as "necessarily" and "possibly", respectively, the Barcan formula entails that if it is possible that something exists, then there is something which is possibly identical to it. Given that necessarily, everything is self-identical, an instance of the converse Barcan formula allows us to show that everything exists necessarily. Needless to say, both of these results are highly controversial. Similarly, in the present setting, there is no obvious reason why the claim that every set satisfies a condition in a special way should be equivalent to the claim that in a special way that every set satisfies that condition. Of course, it is interesting whether such a principle can be consistently added to a given theory in $\mathcal{L}_1^{\Box}(\in)$, and we will return to this question at various points below.

Another aspect of $Q\Lambda$ worth mentioning is the axiom (LNI) and the fact that an analogous principle for $=$ is derivable (see the proof of Lemma 3.2). These assumptions are natural on the metaphysical interpretation of modal operators just mentioned; in the present setting, they are merely included for simplicity, as they lead to a natural logic for identity which corresponds to a standard treatment of identity in Kripke models. Although this won't be explicitly discussed in the following, it is clear that many of the following results do not depend on the identity axioms. In particular, none of the derivations showing that particular modalized comprehension principles are inconsistent in a given modal logic features the identity symbol; *a fortiori* these results do not depend on laws governing identity.

One useful feature of $Q\Lambda$ is that it admits the rule of substitution of equivalents: if $\varphi \leftrightarrow \psi$ is provable, then so is $\chi \leftrightarrow \chi'$, where χ' differs from χ only in having ψ in some places where χ has φ; see Hughes & Cresswell (1996) for further discussion. Appeals to this fact, as well as other elementary features of $Q\Lambda$ such as its including classical propositional logic, will not be explicitly recorded in derivations and other arguments in the following.

Turning to model theory, a Kripke frame (for $\mathcal{L}_0^\Box$) is a pair $\langle W, R\rangle$ consisting of a set W, the "possible worlds", and a binary "accessibility" relation $R \subseteq W^2$; validity of $\mathcal{L}_0^\Box$-formulas relative to a class of Kripke frames is defined as usual. In the following, we will often exploit well-known results to the effect that an $\mathcal{L}_0^\Box$-formula is a theorem of some normal modal logic just in case it is valid on some class of frames.

A Kripke model (for $\mathcal{L}_1^\Box(\in)$) is a structure $\langle W, R, D, V\rangle$ such that $\langle W, R\rangle$ is a Kripke frame, D is a function mapping each world to a set, and V maps each world to a binary relation on $\bigcup_{w\in W} D(w)$. Truth of a formula is defined relative to a world w and an assignment function a, with the crucial clauses as follows:

$M, w, a \vDash x = y$ iff $a(x) = a(y)$

$M, w, a \vDash x \in y$ iff $\langle a(x), a(y)\rangle \in V(w)$

$M, w, a \vDash \Box\varphi$ iff for all $v \in W$, if wRv then $M, v, a \vDash \varphi$

$M, w, a \vDash \forall x\varphi$ iff for all $o \in D(w)$, $M, w, a[o/x] \vDash \varphi$

where $a[o/x]$ maps x to o and every other variable y to $a(y)$. Validity on a model is defined as truth in all worlds on all assignment functions, and validity on a class of frames as validity on all models based on a frame in the class. A routine induction shows that if Λ is valid on a class of Kripke frames C, then $Q\Lambda$ is valid on the class of Kripke models based on a frame in C, and that this extends to adding $(BF)/(CBF)$ if the class of models is decreasing/increasing (if wRv then $D(w) \supseteq D(v)/D(w) \subseteq D(v)$). To show the consistency of, say, $Q\Lambda + (BF) + (\text{Comp}\Box)$ it therefore suffices to construct a non-empty class of decreasing Kripke models based frames validating Λ on which $(\text{Comp}\Box)$ is valid.

3 (Comp□) in S5

Recall the basic modalized comprehension principle:

$$\exists y\forall x(x \in y \leftrightarrow \Box\varphi). \qquad (\text{Comp}\Box)$$

We first consider $(\text{Comp}\Box)$ in the strong modal logic **S5**, the normal modal logic axiomatized by the following two axioms:

$$(T) \qquad \Box p \to p$$
$$(5) \qquad \Diamond p \to \Box\Diamond p$$

We show that $Q\mathbf{S5} + (\text{Comp}\Box)$ is consistent. In fact, this will fall out of a theorem which is stronger in several ways. First, the theorem will precisely characterize the non-modal fragment of $Q\mathbf{S5} + (\text{Comp}\Box)$ as the first-order theory axiomatized by the following two axiom schemas:

$$(F) \qquad \exists y \forall x \left(x \in y \leftrightarrow \bigvee_{i \leq n} x = z_i\right)$$
$$(CF) \qquad \exists y \forall x \left(x \in y \leftrightarrow \bigwedge_{i \leq n} x \neq z_i\right)$$

Informally, this theory can be summed up by saying that for any finite number of sets, there is the set containing them and the set not containing them. Second, the characterization of the non-modal fragment is preserved under adding any further non-modal assumptions: for any set of non-modal sentences Γ, the non-modal fragment of $Q\mathbf{S5} + (\text{Comp}\Box) + \Gamma$ is $(F) + (CF) + \Gamma$. Third, the theorem will prove that this is holds for all normal modal logics contained in **S5** which contain the weak modal logic **D**, axiomatized by the single axiom:

$$(D) \qquad \Box p \to \Diamond p$$

Theorem 3.1. *Let Λ be a normal modal logic such that $\mathbf{D} \subseteq \Lambda \subseteq \mathbf{S5}$, $\Gamma \subseteq \mathcal{S}_1(\in)$, and $\varphi \in \mathcal{S}_1(\in)$.*

$$\varphi \in Q\Lambda + \text{Comp}\Box + \Gamma \text{ iff } \varphi \in (F) + (CF) + \Gamma.$$

We split up the proof into several lemmas. First, we establish the right-to-left direction by showing that the modal theory proves (F) and (CF):

Lemma 3.2. *If Λ is a normal modal logic containing (D), then $(F), (CF) \in Q\Lambda + \text{Comp}\Box$.*

Proof. The necessity of identity is derivable using the axioms and rules of $Q\Lambda$; see Hughes & Cresswell (1996, p. 313):

$$(LI) \qquad x = y \to \Box x = y$$

Using (D), (LI) and (LNI), the following two are easily derived:

$$(1) \qquad x = y \leftrightarrow \Box x = y$$
$$(2) \qquad \Diamond x = y \leftrightarrow \Box x = y$$

Note that the following is a theorem of **D**:

$$(3) \qquad \bigwedge_{i \leq n} (\Diamond p_i \leftrightarrow \Box p_i) \to \left(\Box \bigvee_{i \leq n} p_i \leftrightarrow \bigvee_{i \leq n} \Box p_i\right)$$

Instantiating with $x = z_i$ for p_i and appealing to (1) and (2), we obtain:

$$(4) \qquad \Box \bigvee_{i \leq n} x = z_i i \leftrightarrow \bigvee_{i \leq n} x = z_i$$

The following is an instance of Comp□:

$$(5) \qquad \exists y \forall x \left(x \in y \leftrightarrow \Box \bigvee_{i \leq n} x = z_i \right)$$

(F) follows from (4) and (5). The proof of (CF) is analogous. QED

For the converse direction, consider any $\varphi \in \mathcal{S}_1(\in)$ such that $\varphi \notin (F) + (CF) + \Gamma$. By the completeness of first-order logic, there is a model of $(F) + (CF) + \Gamma$ in which φ is false. Let A be its domain and $E \subseteq A^2$ be the relation used to interpret $\in$. We show that $\varphi \notin Q\Lambda + (\text{Comp}\Box) + \Gamma$, for any given normal modal logic Λ contained in **S5**, by constructing a Kripke model $M = \langle W, R, D, V \rangle$ based on a frame validating **S5** which validates $(\text{Comp}\Box)$ and Γ such that $D(w) = A$ and $V(w) = E$ for some $w \in W$.

The idea behind the model construction is to let the interpretation of $\in$ vary among worlds to such an extent that the only witnesses required for the validity of (Comp□) correspond to the finite and cofinite sets – which can be specified using $=$ alone. This can be done by using permutations of A as worlds, interpreting $\in$ accordingly. As we show, it is not even necessary to include all permutations. More precisely, for any permutation π of A, let the set of elements of A not mapped to themselves by π be called the *support* of π; let W be the set of permutations of A with finite support. Let $R = W^2$ – thereby ensuring that the underlying frame validates **S5** – and D be the constant function to A. For all $\pi \in W$, let $V(\pi) = \pi(E) = \{\langle \pi(o_1), \pi(o_2) \rangle : \langle o_1, o_2 \rangle \in E\}$.

To prove that M validates (Comp□), define the extension of a formula with a distinguished variable relative to a world and assignment:

$$[\![\varphi(x)]\!]_{M,\pi,a} = \{o \in A | M, \pi, a[o/x] \vDash \varphi\}.$$

We show how to apply a permutation of A to worlds and assignment functions, and prove that the extension of a formula φ is invariant if we apply to the world and assignment function any permutation of A which maps all parameters of φ to themselves. It follows that the extension of a formula is determined entirely by its parameters, in the sense that it either contains all other elements of A or none of them; consequently, it expresses a set which is finite or cofinite in A. Along the way, we establish that all worlds agree on the sentences they validate, from which the validity of Γ follows. The following makes this strategy precise.

For any permutations π and ρ, let $\pi\rho$ be the composition of π and ρ; for $X \subseteq W$, let $\pi X = \{\pi\rho : \rho \in X\}$; for $O \subseteq A$, let $\pi(O) = \{\pi(o) : o \in O\}$. For any assignment function a, let $\pi(a)$ be the assignment function such that $\pi(a)(z) = \pi(a(z))$. In the following lemmas, unless noted otherwise, π and ρ are arbitrary member of W, a is an arbitrary assignment function, and φ an arbitary formula.

Lemma 3.3. *$M, \pi, a \vDash \varphi$ iff $M, \rho\pi, \rho(a) \vDash \varphi$.*

Proof. By induction on the complexity of φ; only the case for $\in$ is interesting:

$M, \pi, a \vDash x \in y$ iff
$\langle a(x), a(y)\rangle \in \pi(E)$ iff
$\langle \rho(a)(x), \rho(a)(y)\rangle \in \rho\pi(E)$ iff
$M, \rho\pi, \rho(a) \vDash \varphi$. QED

Since by construction all elements of Γ are true at the identity permutation (on any assignment), this lemma already establishes that Γ is valid on M.

Lemma 3.4. $\rho([\![\varphi(x)]\!]_{M,\pi,a}) = [\![\varphi(x)]\!]_{M,\rho\pi,\rho(a)}$.

Proof. $\rho([\![\varphi(x)]\!]_{M,\pi,a})$
$= \{\rho(o) | o \in A \text{ and } M, \pi, a[o/x] \vDash \varphi\}$
$= \{o \in A | M, \pi, a[\rho^{-1}(o)/x] \vDash \varphi\}$
$= \{o \in A | M, \rho\pi, \rho(a[\rho^{-1}(o)/x]) \vDash \varphi\}$ (by the previous lemma)
$= \{o \in A | M, \rho\pi, \rho(a)[o/x] \vDash \varphi\}$
$= [\![\varphi(x)]\!]_{M,\rho\pi,\rho(a)}$ QED

Lemma 3.5. *If $\rho(a(z)) = a(z)$ for all variables z free in φ, $\rho([\![\Box\varphi(x)]\!]_{M,\pi,a}) = [\![\Box\varphi(x)]\!]_{M,\pi,a}$.*

Proof. $\rho([\![\Box\varphi(x)]\!]_{M,\pi,a})$
$= [\![\Box\varphi(x)]\!]_{M,\rho\pi,\rho(a)}$ (by the previous lemma)
$= \bigcap_{\sigma \in W} [\![\varphi(x)]\!]_{M,\sigma,\rho(a)}$
$= \bigcap_{\sigma \in W} [\![\varphi(x)]\!]_{M,\sigma,a}$ (since $\rho(a(z)) = a(z)$ for all variables z free in φ)
$= [\![\Box\varphi(x)]\!]_{M,\pi,a}$ QED

Lemma 3.6. *If $O \subseteq A$ is finite and $O' \subseteq A$ is such that $\pi(O') = O'$ for all $\pi \in W$ such that $\pi(o) = o$ for all $o \in O$, then O' is finite or cofinite in A.*

Proof. Assume for the sake of contradiction that O' is infinite and coinfinite in A. Then there are $o_1, o_2 \in A \backslash O$ such that $o_1 \in O'$ and $o_2 \notin O'$. Now consider the transposition $(o_1 o_2)$ which switches o_1 and o_2. $(o_1 o_2) \in W$, but $(o_1 o_2)(O') \neq O'$. But this contradicts the assumption. QED

Lemma 3.7. (Comp$\Box$) *is valid in the model M.*

Proof. Consider any φ in which y is not free, any $\pi \in W$ and any assignment a. It suffices to show that $M, \pi, a \vDash \exists y \forall x (x \in y \leftrightarrow \Box\varphi)$. Note that by the preceding two lemmas, $X = [\![\Box\varphi(x)]\!]_{M,\pi,a}$ is finite or cofinite in A. By assumption, there is an $o \in A$ such that for all $o' \in A$, $\langle o', o\rangle \in E$ iff $o' \in \pi^{-1}(X)$. Hence for all $o' \in A$, $\langle o', \pi(o)\rangle \in \pi(E)$ iff $o' \in X$, so $\pi(o)$ witnesses the existential claim. QED

This completes the proof of Theorem 3.1. The construction presented here restricts the worlds to permutations *with finite support* only to demonstrate concretely that the relevant models need only contain countably many worlds. One could arrive at the same conclusion more abstractly by including all permutations, and then appealing to an analog of the Löwenheim-Skolem theorem such as the one established in (Kripke, 1959, p. 7).

$(F) + (CF)$ is a very weak set theory; Theorem 3.1 therefore shows that (Comp$\Box$) is a weak comprehension principle on its own. But it does more: it also shows that considering only the non-modal fragment, the deductive strength of (Comp$\Box$) is precisely captured by $(F) + (CF)$. This suggests that if we were to try to enrich (Comp$\Box$) by supplementary axioms to obtain an interesting set theory, we might need to do so using modal axioms, as it seems *prima facie* unlikely that $(F) + (CF)$ could play a useful role in a stronger extensional set theory with a universal set. Any further modal axioms which are to strengthen the resulting set theory may not already be validated by the models constructed above. It is therefore of interest to note some natural modal principles which they validate. Since the models constructed above have constant domains, they validate (BF) and (CBF). The next proposition records three further validities.

Proposition 3.8. *The following are valid on any model M as above:*

$$\begin{aligned} &(\text{Comp}\Diamond) \quad \exists y \forall x (x \in y \leftrightarrow \Diamond \varphi(x)) \\ &(\text{Mem}) \quad \Diamond x \in y \\ &(\text{Non}) \quad \Diamond x \notin y \end{aligned}$$

Proof. The proof of (Comp$\Diamond$) is analogous to that of (Comp$\Box$). For (Mem), consider any $\pi \in W$ and assignment function a. Let $o \in A$ such that $\langle o', o \rangle \in E$ for all $o' \in A$, and $\rho = (a(y), o)$. Then $\langle a(x), a(y) \rangle \in \rho(E)$, and so $M, \rho, a \vDash x \in y$, as required. The proof of (Non) is similar. QED

Another modal principle valid on the models discussed here is a homogeneity schema discussed in Fine (1978). Calling a formula *de dicto* if it contains no subformula of the form $\Box\varphi$ where φ has a free variable, Fine shows that if this schema added to a modal theory, every formula is equivalent, modulo the expanded theory, to a *de dicto* one, and that the expansion is conservative over the fragment of *de dicto* formulas.

Since all of these are already valid on the models constructed above, they cannot be used to strengthen $Q\mathbf{S5}$ + (Comp$\Box$). Two natural principles which are not valid on these models are the rigidity of membership, and the rigidity of non-membership:

$$x \in y \rightarrow \Box x \in y. \qquad (\text{Rig}\in)$$

$$x \notin y \rightarrow \Box x \notin y. \qquad (\text{Rig}\notin)$$

However, it is easy to see that adding these leads to inconsistency. In fact, in the presence of the following axiom, which is provable in **S5**), adding the former is sufficient for inconsistency:

$$(B) \qquad p \rightarrow \Box\Diamond p$$

However, (Rig$\in$) by itself does not lead to inconsistency in $Q\mathbf{D}$ + (Comp$\Box$); this follows from the consistency results in Fitch (1967a).

Proposition 3.9.

(i) $Q\mathbf{D} + (\text{Comp}\Box) + (\text{Rig}{\in}) + (\text{Rig}{\notin})$ *is inconsistent.*

(ii) $Q\mathbf{KDB} + (\text{Comp}\Box) + (\text{Rig}{\in})$ *is inconsistent.*

Proof.

(i)

(1)	$y \in y \to \Box y \in y$	$(\text{Rig}{\in})$
(2)	$\Box y \in y \to \neg\Box y \notin y$	(D)
(3)	$\neg(y \in y \land \Box y \notin y)$	(1), (2)
(4)	$y \notin y \to \Box y \notin y$	$(\text{Rig}{\notin})$
(5)	$\neg(y \notin y \land \neg\Box y \notin y)$	(4)
(6)	$\neg(y \in y \leftrightarrow \Box y \notin y)$	(3), (5)
(7)	$(\forall x(x \in y \leftrightarrow \Box x \notin x) \land Ey) \to (y \in y \leftrightarrow \Box y \notin y)$	$(\forall 1E)$
(8)	$Ey \to \neg\forall x(x \in y \leftrightarrow \Box x \notin x)$	(6), (7)
(9)	$\forall y(Ey \to \neg\forall x(x \in y \leftrightarrow \Box x \notin x))$	(8), (UG)
(10)	$\forall y Ey$	(UE)
(11)	$\forall y \neg\forall x(x \in y \leftrightarrow \Box x \notin x)$	(9), (10), $(\forall^{\to})$
(12)	$\neg\exists y \forall x(x \in y \leftrightarrow \Box x \notin x)$	(11)

(ii) In place of step (4) we insert the following derivation:

$(4a)$	$\Diamond y \in y \to \Diamond\Box y \in y$	$(\text{Rig}{\in})$, (N)
$(4b)$	$\Diamond\Box y \in y \leftrightarrow \neg\Box\Diamond y \notin y$	Definition
$(4c)$	$y \notin y \to \Box\Diamond y \notin y$	(B)
$(4d)$	$y \notin y \to \neg\Diamond y \in y$	(4a-c)

The rest of the proof is exactly as above.

QED

4 $(\Box\text{Comp}\Box)$

S5 is a very strong modal logic; most common modal logics are contained in it. As the results of the previous section indicate that even in such a strong background logic, $(\text{Comp}\Box)$ only leads to a relatively weak set theory, it is natural to strengthen it by replacing its material biconditional by a strict biconditional:

$$\exists y \forall x \Box(x \in y \leftrightarrow \Box\varphi) \qquad (\Box\text{Comp}\Box)$$

This principle was the subject of Mints's original question; it turns out to be inconsistent in the relatively weak modal logic **T**, axiomatized by the single axiom (T). Had he seen the weakness of (Comp$\Box$) and our first contradiction using ($\Box$Comp$\Box$), we suspect Mints himself would have considered variants of ($\Box$Comp$\Box$) in which the second modal operator is replaced by some other modality M, i.e., any string of modal operators and negations:

$$\exists y \forall x \Box (x \in y \leftrightarrow M\varphi). \qquad (\Box\text{CompM})$$

Accordingly, we will also consider these principles systematically below.

4.1 Inconsistency in T

The inconsistency of ($\Box$Comp$\Box$) in **T** follows from the fact that the negation of the following principle, for $M = \Box$, is derivable in $Q\mathbf{T}$:

$$\exists y \forall x \Box (x \in y \leftrightarrow Mx \notin x) \qquad (\Box\text{RussellM})$$

The argument will be factored into a propositional and a quantified part; the latter part is most conveniently formulated more generally for arbitrary modalities, so as to be of further use below.

Lemma 4.1. $\neg\Box(p \leftrightarrow \Box\neg p) \in \mathbf{T}$.

Proof. By a standard completeness result, it suffices to show that this formula is valid on reflexive frames. Assume for contradiction that $\Box(p \leftrightarrow \Box\neg p)$ is true in some world w of a model based on a reflexive frame. Then $p \leftrightarrow \Box\neg p$ is true in every accessible world. If p is true in some such world, then $\Box\neg p$ is true there as well, leading to contradiction by reflexivity. So p is false in all such worlds, which means that $\Box\neg p$ is true in w. By reflexivity, $\neg p$ is true in w. But also by reflexivity, $p \leftrightarrow \Box\neg p$ is true in w, and so p is true in w. $\lightning$. QED

Lemma 4.2. *For any normal modal logic* Λ, *if* $\neg\Box(p \leftrightarrow M\neg p) \in \Lambda$, *then* $\neg(\Box\text{RussellM}) \in Q\Lambda$.

Proof.

(1)	$\neg\Box(y \in y \leftrightarrow My \notin y)$	(Λ)
(2)	$(\forall x \Box(x \in y \leftrightarrow Mx \notin x) \wedge Ey) \to \Box(y \in y \leftrightarrow My \notin y)$	$(\forall 1E)$
(3)	$Ey \to \neg\forall x \Box(x \in y \leftrightarrow Mx \notin x)$	(1), (2)
(4)	$\forall y(Ey \to \neg\forall x \Box(x \in y \leftrightarrow Mx \notin x))$	(3), (UG)
(5)	$\forall y Ey$	(UE)
(6)	$\forall y \neg\forall x \Box(x \in y \leftrightarrow Mx \notin x)$	(4), (5), $(\forall^{\to})$
(7)	$\neg\exists y \forall x \Box(x \in y \leftrightarrow Mx \notin x)$	(6)

QED

Corollary 4.3. $\neg(\Box\text{Russell}\Box) \in Q\mathbf{T}$.

4.2 Inconsistency of Variants in S4

We consider $\Box$CompM in the more restrictive setting of **S4**, the normal modal logic axiomatized by T and the following axiom:

$$(4) \qquad \Box p \to \Box\Box p$$

Say that two modalities M and N are equivalent in a normal modal logic Λ if $Mp \leftrightarrow Np \in \Lambda$. Up to equivalence in **S4**, there are fourteen modalities; see, e.g., Hughes & Cresswell (1996, p. 55). A useful way of generating them is as follows: let the *dual* of a modality M be $\neg M \neg$, and the *inner negation* of M be $M\neg$. Up to equivalence in **S4**, every modality can be generated from $-$ (the empty sequence), $\Box$, $\Box\Diamond$ and $\Box\Diamond\Box$ by taking duals and inner negations.

We show that in Q**S4**, $\Box$CompM is inconsistent for every modality M. We start with the four basic modalities:

Lemma 4.4. *$\neg\Box(p \leftrightarrow M\neg p) \in$ **S4** for every modality $M \in \{-, \Box, \Box\Diamond, \Box\Diamond\Box\}$.*

Proof. The case of $-$ is immediate, and the case of $\Box$ follows from Lemma 4.1. The remaining two cases can be established by similar arguments using the completeness of **S4** with respect to reflexive and transitive frames. QED

The next lemma extends this result to the duals of the basic modalities:

Lemma 4.5. *For any normal modal logic Λ and modality M, if $\neg\Box(p \leftrightarrow M\neg p) \in \Lambda$, then $\neg\Box(p \leftrightarrow \neg M\neg\neg p) \in \Lambda$.*

Proof. Only propositional logic is required:

$$\begin{array}{ll} (1) & \neg\Box(p \leftrightarrow M\neg p) \\ (2) & \neg\Box(\neg p \leftrightarrow \neg M\neg p) \\ (3) & \neg\Box(p \leftrightarrow \neg M\neg\neg p) \end{array}$$

QED

Corollary 4.6. *For every modality M, Q**S4** proves the negation of an instance of $\Box$CompM.*

Proof. If M is a member or a dual of a member of $\{-, \Box, \Box\Diamond, \Box\Diamond\Box\}$, it follows from Lemmas 4.2, 4.4 and 4.5 that $\neg(\Box\text{RussellM}) \in Q$**S4**. If not, there is such a modality N such that M is equivalent, in **S4**, to $N\neg$. As just noted, $\neg(\Box\text{RussellN}) \in Q$**S4**, so $\neg(\Box\text{RussellN}\neg\neg) \in Q$**S4**. The negation of the corresponding instance of ($\Box$CompM) is therefore derivable from Q**S4**. QED

The present results leave open the non-trivial consistency of ($\Box$CompM) in normal modal logics Λ weaker than **S4**. However, in such a setting, we may well be faced with infinitely many distinct modalities, up to equivalence in Λ, which makes a general assessment of the situation difficult. We therefore return to ($\Box$Comp$\Box$) for the remainder of this section.

4.3 Extensions of D

Since ($\Box$Comp$\Box$) is inconsistent in **T**, let us consider normal modal logics weaker than, or incomparable to, **T**. We first show that the principle is trivially consistent in logics which fail to prove the axiom (D). We then consider three natural extensions of **D**, namely **KD4**, **KD5** and **KDB**. We show that ($\Box$Comp$\Box$) is inconsistent in all of them. The question of the consistency of the principle in **D** itself will be considered later, and answered in the affirmative.

The consistency of ($\Box$Comp$\Box$) in normal modal logics not containing (D) extends to (Comp$\Box$), and these results are conveniently established together:

Proposition 4.7. *For any normal modal logic Λ such that $(D) \notin \Lambda$, $Q\Lambda +$ (Comp$\Box$) $+$ ($\Box$Comp$\Box$) is consistent.*

Proof. If $(D) \notin \Lambda$, then $\Lambda \subseteq$ **Ver**, the normal modal logic axiomatized by the axiom $\Box p$ (see Hughes & Cresswell (1996, p. 67)). Let $M = \langle W, R, D, V \rangle$ be a model such that W is a singleton $\{w\}$, R is the empty relation, $D(w)$ is non-empty, and $V(w) = D(w)^2$. Since $\Box p$ is valid on the underlying frame, so is Λ. ($\Box$Comp$\Box$) is trivially validated, and the validity of (Comp$\Box$) follows from the existence of a universal set – indeed, every element is a universal set. QED

From the model construction in the proof, it is clear that the result could be strengthened by adding various further principles, such as (BF) and (CBF), and that the resulting theories are uninteresting.

The inconsistency of ($\Box$Comp$\Box$) in extensions of **KD** will be based on the following lemma concerning the propositional logics:

Lemma 4.8.

$$\neg\Box(p \leftrightarrow \Box\neg p) \in \mathbf{KD4}.$$

$$\neg\Box(p \leftrightarrow \Box\Diamond\neg p) \in \mathbf{KDB}.$$

$$\Box\neg\Box(p \leftrightarrow \Box\neg p) \in \mathbf{KD5}.$$

Proof. As with similar arguments above, these can be established model-theoretically using the corresponding classes of frames, satsfying seriality and, respectively, transitivity, symmetry and euclideanness. QED

Using Lemma 4.2, the derivability of the negation of the relevant instance of ($\Box$Comp$\Box$) follows immediately in the first two cases. The third case requires a slight modification of the argument:

Lemma 4.9. *For any normal modal logic Λ containing (D), if $\Box\neg\Box(p \leftrightarrow \Box\neg p) \in \Lambda$, then $\neg\Box$($\Box$Russell$\Box$) $\in Q\Lambda$.*

Proof.

(1)	$\Box\neg\Box(y \in y \leftrightarrow My \notin y)$	(Λ)
(2)	$\Box((\forall x\Box(x \in y \leftrightarrow Mx \notin x) \wedge Ey) \to \Box(y \in y \leftrightarrow My \notin y))$	$(\forall 1E), (N)$
(3)	$(\Box\neg\Box p \wedge \Box(q \to \Box p)) \to \Box\neg q$	(Λ)
(4)	$\Box\neg(\forall x\Box(x \in y \leftrightarrow Mx \notin x) \wedge Ey)$	$(1) - (3)$
(5)	$\Box\forall y(Ey \to \neg\forall x\Box(x \in y \leftrightarrow Mx \notin x))$	$(UGL\forall^0)$
(6)	$\Box\forall y Ey$	$(UE), (N)$
(7)	$\Box\forall y\neg\forall x\Box(x \in y \leftrightarrow Mx \notin x)$	$(5), (6), (\forall^{\to}), (N), (\Lambda)$
(8)	$\neg\Box\exists y\forall x\Box(x \in y \leftrightarrow Mx \notin x)$	$(7), (D)$

QED

5 (Comp□) beyond S5

The previous section suggests that in strengthening (Comp□) to (□Comp□), we may have overshot our mark: the principle is so strong that it is inconsistent even in weak normal modal logics like **T**. So maybe (Comp□) was a better principle to work with after all. Of course, the relatively uninteresting models of it in **S5** mean that we need to consider it in the context of logics which are not contained in **S5**, but there are a number of interesting such logics. In this section, we explore such options. We first show that we can't simply strengthen **S5**, as (Comp□) is inconsistent in any proper extension of **S5**. We then consider a natural normal modal logic incomparable to **S5**. We show that (Comp□) is consistent in it, and exhibits at least *prima facie* interesting behaviour.

5.1 Inconsistency in Proper Extensions of S5

The result that (Comp□) is inconsistent in any proper extensions of **S5** relies on the fact that any such extension contains an instance of the following schema, for some $n < \omega$:

(Alt$_n$) $\bigvee_{i \leq n} \Box \left(\bigwedge_{j<i} p_j \to p_i\right)$

Indeed, the inconsistency can be derived in any normal modal logic containing such an instance and (T).

Proposition 5.1. *For any normal modal logic Λ which contains (T) and* (Alt$_n$) *for some $n < \omega$, $Q\Lambda +$* (Comp□) *is inconsistent.*

Proof. We work in $Q\Lambda +$ (Comp□). Define the following formulas, for every $n < \omega$ and variable z:

$$\varphi_n(z) := \bigwedge_{i<n} y_i \notin y_i \to z \notin z$$

$$\psi_n := \forall x(x \in y_n \leftrightarrow \Box\varphi_n(x))$$

In order to keep the derivations readable, they will be presented in a somewhat abbreviated form. Note that the following version of universal instantiation is derivable in $Q\Lambda$; see Hughes & Cresswell (1996, p. 294):

$(\forall 1')$ $\forall y(\forall x\varphi \to \varphi[y/x])$

We derive, for any $n < \omega$, $\forall y_0 \ldots \forall y_n \left(\bigwedge_{i\leq n} \psi_i \to \bigwedge_{i\leq n} y_i \notin y_i\right)$ by induction on n:

(1)	$\forall y_0 \ldots \forall y_{n-1} \left(\bigwedge_{i<n} \psi_i \to \bigwedge_{i<n} y_i \notin y_i\right)$	IH
(2)	$\forall y_n(\psi_n \to (y_n \in y_n \leftrightarrow \Box\varphi_n(y_n)))$	$(\forall 1')$
(3)	$\forall y_0 \ldots \forall y_n \left(\bigwedge_{i\leq n} \psi_i \to \left(\bigwedge_{i<n} y_i \notin y_i \wedge (y_n \in y_n \leftrightarrow \Box\varphi_n(y_n))\right)\right)$	(1), (2)
(4)	$\bigwedge_{i<n} y_i \notin y_i \to (\Box\varphi_n(y_n) \to y_n \notin y_n)$	(T)
(5)	$\forall y_0 \ldots \forall y_n \left(\bigwedge_{i\leq n} \psi_i \to \bigwedge_{i\leq n} y_i \notin y_i\right)$	(3), (4)

Starting from this theorem for $n = m$, consider the following derivation:

(1)	$\forall y_0 \ldots \forall y_m \left(\bigwedge_{i\leq m} \psi_i \to \bigwedge_{i\leq m} y_i \notin y_i\right)$	
(2)	$\forall y_0 \ldots \forall y_m \left(\left(\bigwedge_{i\leq m} \psi_i \wedge \bigwedge_{i\leq m} y_i \notin y_i\right) \to \bigwedge_{i\leq m} \neg\Box\varphi_i(y_i)\right)$	$(\forall 1')$
(3)	$\forall y_0 \ldots \forall y_m \left(\bigwedge_{i\leq m} \psi_i \to \bigwedge_{i\leq m} \neg\Box\varphi_i(y_i)\right)$	(1), (2)
(4)	$\forall y_0 \ldots \forall y_m \left(\bigwedge_{i\leq m} \psi_i \to \neg \bigvee_{i\leq m} \Box\varphi_i(y_i)\right)$	(3)
(5)	$\bigvee_{i\leq m} \Box\varphi_i(y_i)$	(Alt_m)
(6)	$\forall y_0 \ldots \forall y_m \neg \bigwedge_{i\leq m} \psi_i$	(4), (5)
(7)	$\exists y_0 \ldots \exists y_m \bigwedge_{i\leq m} \psi_i$	$(\mathrm{Comp}\Box)$

The inconsistency follows from 6 and 7. QED

Corollary 5.2. *For any normal modal logic Λ which properly extends* **S5**, *$Q\Lambda + (\mathrm{Comp}\Box)$ is inconsistent.*

Proof. By a theorem due to Scroggs (1951), there is an $n < \omega$ such that $(\mathrm{Alt}_n) \in \Lambda$. QED

5.2 KDDc

We now turn to a normal modal logic incomparable in strength with **S5**: **KDDc**, the normal modal logic axiomatized by the axioms (D) and the converse of (D):

(D_c) $\Diamond p \to \Box p$

The system **KDDc** is atypical in allowing the distribution of $\Box$ across all Boolean operators: for example in this system $\neg\Box\varphi$ is equivalent to $\Box\neg\varphi$ and $\Box(\varphi \vee \psi)$ is equivalent to $\Box\varphi \vee \Box\psi$. This makes the theory particularly interesting as a fictionalist theory: although not everything which is true in the fiction is true, the fiction is complete in the sense that $\Box\varphi \vee \Box\neg\varphi$ is a theorem schema. If we add (BF) and (CBF), this property of movement extends to the quantifiers as well: $\exists x\Box\varphi$ is now equivalent to $\Box\exists x\varphi$. Using this further property, together with one application of necessitation, every instance of $(\Box\mathrm{Comp}\Box)$ is derivable from a corresponding instance of (Comp$\square$). As mentioned below, the following result can be strengthened by adding (BF) and (CBF), so it also establishes the consistency of ($\square$Comp$\square$) in **KDDc**, and therefore in **KD**, as claimed above.

Proposition 5.3. Q**KDDc** + (Comp$\square$) *is consistent.*

Proof. To show that **KDDc** + (Comp$\square$) is consistent, we construct a Kripke model; although not every theorem of **KDDc** + (Comp$\square$) is true in every world of this model, every theorem is true in some world, and this suffices for consistency.

Let $M = \langle W, R, D, V \rangle$ such that $W = \mathbb{N}$, $R = \{\langle w+1, w \rangle : w \in W\}$, D is the constant function to $\mathbb{N}$, and V is defined as follows:

$$V(0) = D^2$$

$$V(w+1) = \{\langle r, t \rangle : M, w \vDash \beta(t)[r]\}$$

where β is some fixed bijection from D to the set of tuples $\langle \varphi, \bar{s} \rangle$ such that φ is a formula with $n+1$ free variables and $\bar{s}$ is a tuple of n elements of D. For $\beta(t) = \langle \varphi, \bar{s} \rangle$, read $M, w \vDash \beta(t)[r]$ as saying that φ is true in w with the first n free variables in it (according to some fixed enumeration) interpreted using $\bar{s}$ and the remaining variable interpreted as r.

Say that a formula φ is true in $w \in W$ iff φ is true in w on each interpretation of the free variables of φ. Consider a variant axiomatization of Q**KDDc** in which the schema **KDDc** is replaced by substitution instances of the axioms of **KDDc**. For each $w \in W$, it is routine to observe that each axiom of Q**KDDc** except for (D) is true in w, and that truth in w is closed under the rules of Q**KDDc** except for (N). Furthermore, each instance of (Comp$\square$) is true in w; as can be seen by distinguishing two cases: For $w = 0$, the claim is true since $\square\varphi$ is always true in 0, so there only needs to be a universal set in 0, which is guaranteed by the construction of V. For all other worlds, the claim follows by construction of V as well. Finally, note that any substitution instance of (D) in any $w > 0$.

We now show that if φ can be derived in **KDDc** + (Comp$\square$) by a proof using n applications of (N), then φ is true in $n+1$; this can be done by an induction on n. (It is $n+1$ rather than n because (D) need not be true in 0.) It follows that Λ is consistent: If $\bot$ were derivable, there would be a proof of $\bot$; letting n be the number of applications of (N) in this proof, it would follow that $\bot$ us true in $n+1$, which is false. QED

As the model used in this proof employs constant domains, the result can be strengthened by adding (BF) and (CBF).

The theory Q**KDDc** + (Comp$\square$) appears informally to be stronger than the theory Q**S5** + (Comp$\square$). For simplicity, consider the extensions of these systems by (BF) and (CBF). Whereas the proof of Theorem 3.1 shows that (Comp$\square$) has constant-domain models with a universal accessibility relation in which (necessarily) all sets have finitely or cofinitely many members, no constant domain model on a frame validating **KDDc** can have this property: It is easy to see that all worlds of such a model must have infinite domains, and validate $(F) + (CF)$. A suitable instance of (Comp$\Box$) asserts the existence of a set whose elements are exactly those who satisfy the necessitation of the claim that they have exactly n members, for some positive natural number n; since all worlds have infinite domains and validate $(F) + (CF)$, there must be infinitely members of this set as well as infinitely many non-members. Indeed, it can be shown that Q**KDDc** $+ (BF) + (CBF) +$ (Comp$\Box$) *proves* schematically, for each natural number n, that such a set is distinct from every set containing exactly n elements, and distinct from every set excluding exactly n elements, since this system is complete with respect to the class of constant-domain models which validate (Comp$\Box$) and are based on a frame validating **KDDc** (a proof of the completeness of the system is omitted, but the claim is a relatively straightforward consequence of the completeness of Q**KDDc** $+ (BF) + (CBF)$ with respect to constant-domain models based on frames validating **KDDc**, which itself can be established by a routine argument).

Without considering further set-theoretic axioms, Q**KDDc** $+$ (Comp$\Box$) is therefore the most promising system considered in this paper. Yet, it also suffers from serious limitations. In particular, the theory does not prove natural principles about the closure of sets under Boolean operations. Consider the following three principles:

(Neg)	$\exists y \forall z (z \in y \leftrightarrow z \notin x)$
(Con)	$\exists z \forall w (w \in z \leftrightarrow (z \in x \land z \in y))$
(Union)	$\exists z \forall w (w \in z \leftrightarrow (z \in x \lor z \in y))$.

The last of these is most familiar, as the axiom of union; the second is usually obtained as an instance of separation; and the first is a common principle in set theories with a universal set. Each of these principles is consistent with Q**KDDc** $+$ (Comp$\Box$), but so are their negations. To see this, we can simply modify the above model construction so that at some world, the mapping is still an injection from instances of comprehension to individuals in the domain but is not a surjection. Which individuals are members of the sets not in the image of the mapping can be chosen arbitrarily without affecting the validity of (Comp$\Box$). We can thus use them to create counterexamples to each of the three principles. On its own, Q**KDDc** $+$ (Comp$\Box$) therefore also does not seem to be a promising set theory.

6 Conclusion

We began our investigation with the modal principle (Comp$\Box$). The most commonly studied modal logics are sublogics of **S5**. In all such logics the principle (Comp$\Box$) is consistent, but on its own, it leads to a weak set theory. This suggested that we strengthen the comprehension principle, to ($\Box$Comp$\Box$). This alternative principle turned out to be inconsistent in all extensions of the modal system **T**, as well as in a number of standard extensions of **KD**. These results led us to examine (Comp$\Box$) in proper extensions of **S5**, where we showed it to be inconsistent. We also considered it and ($\Box$Comp$\Box$) in the system **KDDc**, where we showed them to be consistent, although in certain ways also limited in strength.

Overall, we have seen that if modalities can save naive set theory along the lines of (Comp$\Box$), they probably can't do so on their own – as far as the modal systems we have investigated are concerned, (Comp$\Box$) seems to require supplementation by further set-theoretic axioms, most likely themselves modal, to yield any interesting set theory. Such additional axioms were not considered here; the prospects for constructing set interesting theories based on $Q\Lambda +$ ($\Box$Comp$\Box$) and additional set-theoretic axioms, for some normal modal logic Λ, is a wide open field for future research.

There are a number of interesting further questions left open by the results established here. One concerns a variant of ($\Box$Comp$\Box$) in which the universal quantifier and the outer modal operator are interchanged:

$$\exists y \Box \forall x (x \in y \leftrightarrow \Box \varphi) \qquad (\Box\text{Comp}\Box')$$

In some settings, such as $Q\mathbf{T}$, it is not hard to see how to adapt the proof of the inconsistency of ($\Box$Comp$\Box$) to show the inconsistency of ($\Box$Comp$\Box'$), but in other settings, such as **KD4**, the question whether this variant principle is consistent seems harder to settle.

Further open questions arise from normal modal logics incomparable in strength to **S5** apart from **KDDc**. A particularly interesting example is **S4M**, the normal extension of **S4** by the axiom

$$(M) \qquad \Box\Diamond p \to \Diamond\Box p.$$

In frames validating **S4M**, each world can access a world which can only access itself; it follows that there can be no model of (Comp$\Box$) based on a frame validating **S4M**. Yet this does not mean that $Q\mathbf{S4M} +$ (Comp$\Box$) is inconsistent, since $Q\mathbf{S4M}$ is incomplete with respect to the class of such models; see Hughes & Cresswell (1996, pp. 265–270). In case $Q\mathbf{S4M} +$ (Comp$\Box$) turns out to be consistent, it would also be interesting to consider certain normal extensions of **S4M**, in particular **K2**, which adds the converse of M.

A final area of open questions concerns the strength of $Q\mathbf{KDDc} +$ (Comp$\Box$). Even though it was shown above not to prove certain basic set-theoretic principles, much about its strength is left open. One way of measuring its strength is in terms of interpretability: which standard theories, such as Peano Arithmetic, can be interpreted in it? As usual, different notions of interpretability are available, and in the present setting, we can additionally choose between interpreting the relevant theory in the full modal system or its non-modal fragment. The non-modal fragment may also be characterized directly, analogously to the characterization of the non-modal fragment of $Q\mathbf{S5} +$ (Comp$\Box$) as $(F) + (CF)$ in Theorem 3.1. In fact, none of the results established above immediately rules out the hypothesis that the non-modal fragment of $Q\mathbf{KDDc} +$ (Comp$\Box$) is $(F) + (CF)$ as well.

References

Aczel, Peter, & Feferman, Solomon. 1980. Consistency of the unrestricted abstraction principle using an intensional equivalence operator. *In:* Seldin, J.P., & Hindley, J.R. (eds), *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*. New York: Academic Press.

Bacon, Andrew. 2015. Can the Classical Logician Avoid the Revenge Paradoxes? *Philosophical Review*, **124**(3), 299–352.

Balaguer, Mark. 2008. Fictionalism in the philosophy of mathematics. *Stanford Encyclopedia of Philosophy*.

Blass, Andreas. 1990. Infinitary combinatorics and modal logic. *The Journal of Symbolic Logic*, **55**(02), 761–778.

Boolos, George. 1995. *The logic of provability*. Cambridge University Press.

Cantini, Andrea. 2009. Paradoxes, Self-Reference and Truth in the 20th Century. *In:* Woods, John, & Gabbay, Dov M. (eds), *Handbook of the History of Logic, volume 5: Logic from Russell to Church*. Amsterdam: Elsevier.

Feferman, Solomon. 1984. Toward Useful Type-Free Theories, I. *The Journal of Symbolic Logic*, **49**, 75–111.

Field, Hartry. 1980. *Science Without Numbers*. Princeton University Press.

Field, Hartry. 1989. *Realism, mathematics & modality*. Basil Blackwell.

Field, Hartry, Lederman, Harvey, & Øgaard, Tore Fjetland. forthcoming. Prospects for a Naive Theory of Classes. *Notre Dame Journal of Formal Logic*.

Fine, Kit. 1978. Model Theory for Modal Logic – Part II. The Elimination of *de re* Modality. *Journal of Philosophical Logic*, **7**, 277–306.

Fine, Kit. 1981. First-Order Modal Theories I—Sets. *Noûs*, **15**, 177–205.

Fitch, Frederic B. 1942. A basic logic. *The journal of symbolic logic*, **7**(03), 105–114.

Fitch, Frederic B. 1948. An extension of basic logic. *The Journal of Symbolic Logic*, **13**(02), 95–106.

Fitch, Frederic B. 1963. The system CΔ of combinatory logic. *The Journal of Symbolic Logic*, **28**(01), 87–97.

Fitch, Frederic B. 1966. A consistent modal set theory (Abstract). *The Journal of Symbolic Logic*, **31**, 701.

Fitch, Frederic B. 1967a. A complete and consistent modal set theory. *The Journal of Symbolic Logic*, **32**, 93–103.

Fitch, Frederic B. 1967b. A Theory of Logical Essences. *The Monist*, **51**, 104–109.

Fitch, Frederic B. 1970. Correction to a Paper on Modal Set Theory. *The Journal of Symbolic Logic*, **35**, 242.

Fitting, Melvin. 2003. Intensional logic—beyond first order. *Pages 87–108 of: Trends in Logic*. Springer.

Forster, T. E. 1995. *Set Theory with a Universal Set*. Second edn. Oxford: Clarendon Press.

Fritz, Peter, & Goodman, Jeremy. forthcoming. *Higher-Order Contingentism, Part 1: Closure and Generation*. *Journal of Philosophical Logic*.

Gallin, Daniel. 1975. *Intensional and Higher-Order Modal Logic*. Amsterdam: North-Holland.

Gilmore, P. C. 1967 (Jan 24). *The consistency of a positive set theory*. Tech. rept. RC-1754. IBM Research Report.

Goodman, Nicolas. 1990. Topological models of epistemic set theory. *Annals of Pure and Applied Logic*, **46**, 147–167.

Hamkins, Joel, & Löwe, Benedikt. 2008. The modal logic of forcing. *Transactions of the American Mathematical Society*, **360**(4), 1793–1817.

Hamkins, Joel David. 2003. A simple maximality principle. *The Journal of Symbolic Logic*, **68**(02), 527–550.

Hamkins, Joel David, & Woodin, W Hugh. 2004. The necessary maximality principle for ccc forcing is equiconsistent with a weakly compact cardinal. *arXiv preprint math/0403165*.

Hellman, Geoffrey. 1989. *Mathematics without numbers: towards a modal-structural interpretation: towards a modal-structural interpretation*. Oxford University Press.

Hughes, George Edward, & Cresswell, Maxwell John. 1996. *A new introduction to modal logic*. London: Routledge.

Kaye, Richard. 1993. Review. *Notre Dame Journal of Formal Logic*, **34**, 302–309.

Krajíček, Jan. 1987. A Possible Modal Formulation of Comprehension Scheme. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **33**, 461–480.

Krajíček, Jan. 1988. Some Results and Problems in the Modal Set Theory MST. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **34**, 123–134.

Kripke, Saul A. 1959. A Completeness Theorem in Modal Logic. *The Journal of Symbolic Logic*, **24**, 1–14.

Kripke, Saul A. 1975. Outline of a theory of truth. *Journal of Philosophy*, **72**(19), 690–716.

Linnebo, Øystein. 2010. Pluralities and sets. *The Journal of Philosophy*, **107**(3), 144–164.

Linnebo, Øystein. 2013. The potential hierarchy of sets. *The Review of Symbolic Logic*, **6**(02), 205–228.

Parsons, Charles. 1983. Sets and Modality. *In: Mathematics in Philosophy*. Ithaca: Cornell University Press.

Rosen, Gideon. 2001. Nominalism, naturalism, epistemic relativism. *Noûs*, **35**(s15), 69–91.

Rundle, Bede. 1969. Review. *The Journal of Symbolic Logic*, **34**, 125.

Scott, Dana. 2010 (June). *A Boolean-Valued Modal Set Theory*. Unpublished Oxford Lectures, available online at: www.homepages.inf.ed.ac.uk/als/ScottInScotland/ScottLect3.pdf.

Scroggs, Schiller Joe. 1951. Extensions of the Lewis System S5. *The Journal of Symbolic Logic*, **16**(2), 112–120.

Shapiro, Stewart (ed). 1985. *Intensional Mathematics*. Amsterdam: North-Holland.

Smullyan, Raymond, & Fitting, Melvin. 1996. *Set theory and the continuum problem*. Oxford University Press.

Solovay, Robert M. 1976. Provability interpretations of modal logic. *Israel journal of mathematics*, **25**(3-4), 287–304.

Studd, James. 2013. The iterative conception of set. *Journal of Philosophical Logic*, **42**(5), 697–725.

Williamson, Timothy. 2013. *Modal Logic as Metaphysics*. Oxford University Press.

Yablo, Stephen. 2001. Go figure: A path through fictionalism. *Midwest studies in philosophy*, **25**(1), 72–102.

Zalta, Edward N. 1988. *Intensional Logic and the Metaphysics of Intentionality*. MIT Press.