The Metamathematics of Putnam's Model-Theoretic Arguments

Tim Button

Abstract. Putnam famously attempted to use model theory to draw metaphysical conclusions. His *Skolemisation* argument sought to show metaphysical realists that their favourite theories have countable models. His *permutation* argument sought to show that they have permuted models. His *constructivisation* argument sought to show that any empirical evidence is compatible with the Axiom of Constructibility. Here, I examine the metamathematics of all three model-theoretic arguments, and I argue against Bays (2001, 2007) that Putnam is largely immune to metamathematical challenges.

Hilary Putnam famously attempted to use model theory to draw metaphysical conclusions. Specifically, he attacked *metaphysical realism*, a position characterised by the following *credo*:

> [T]he world consists of a fixed totality of mind-independent objects. (Putnam 1981, p. 49; cf. 1978, p. 125).
>
> Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things. (1981, p. 49; cf. 1989, p. 214)
>
> [W]hat is epistemically most justifiable to believe may nonetheless be false. (1980, p. 473; cf. 1978, p. 125)

To sum up these claims, Putnam characterised metaphysical realism as an "*externalist* perspective" whose "favorite point of view is a God's Eye point of view" (1981, p. 49). Putnam sought to show that this externalist perspective is deeply untenable. To this end, he treated correspondence in terms of model-theoretic satisfaction. This enabled him to deploy results from model theory against metaphysical realism. In particular, he presented two famous model-theoretic arguments: his Skolemisation argument and his permutation argument.

In this paper, I will investigate the metamathematical underpinnings of both arguments. Since both arguments require only extremely weak model-theoretic resources, it would seem that metaphysical realists cannot reasonably object to Putnam's metaphysical conclusions on purely metamathematical grounds. Timothy Bays, however, has raised a challenge

against Putnam on *exactly* those grounds. Bays' main target is Putnam's less famous constructivisation argument, which seeks to establish that any empirical evidence is compatible with the Axiom of Constructibility. However, Bays thinks that his challenge applies equally well against the Skolemisation argument.

I agree that Bays' challenge poses considerable problems for the constructivisation argument. However, I shall show that it has no impact at all on either the Skolemisation or the permutation arguments. Perhaps Putnam's arguments can be refuted on *other* grounds; but the metamathematics at the heart of Putnams model-theoretic arguments is completely secure.

1 A quick introduction to Bays' dilemma

Putnam's model-theoretic arguments usually have the following structure. The metaphysical realist starts by outlining her favourite theory, T_0, hoping to describe the world as its intended model. Putnam responds by using some model theory, T_1, to produce an apparently unintended model of T_0. The challenge is for the metaphysical realist to explain why this model of T_0 is, indeed, *un*intended.

Putnam's three model-theoretic arguments—constructivisation, Skolemisation and permutation—require different theorems from model theory. Accordingly, they can be presented using different model theories. Nevertheless, the arguments share the following general feature: the metaphysical realist's favourite theory, T_0, and Putnam's model theory, T_1, will typically be *different* theories. Unless T_1 is weaker than T_0, Bays thinks that this causes serious problems for Putnam.

> Putnam faces an *inescapable* dilemma. If he pitches his argument toward philosophers who accept less set theory than he himself does, then these philosophers will reject his argument simply because they reject the set theory used in proving Putnam's key theorem. If he pitches his argument toward philosophers who accept the same set theory that he does, then his argument cannot take care of *these philosophers'* [favourite theories]. (2001, p. 340; cf. 2007, p. 122)

In short, Bays thinks that there are two absolutely general strategies for dealing with Putnam's argument:

(***a***) **Rejection.** A metaphysical realist who accepts T_0 has not yet incurred any commitment to T_1. So she can simply *reject* T_1, and deny that there are any nonstandard models of T_0.

(***b***) **Nonchalance.** A metaphysical realist who accepts T_1 in addition to T_0 can react with *nonchalance* to the (mere) existence of nonstandard models of T_0. She would only care about the existence of nonstandard models of $\mathrm{T}_1 + \mathrm{T}_0$, and Putnam has not (yet) shown that there are any such models.

Bays thinks that any metaphysical realist will follow one strategy or the other. Accordingly, these strategies present a dilemma for Putnam, which I shall call *Bays' dilemma*.[1]

At its best, Bays' dilemma would show that Putnam cannot convince the metaphysical realist that there *definitely is* an unintended model for any theory she cares about. This allows space for Putnam to convince the metaphysical realist that there *might be* such models. Perhaps this mere *possibility* is enough to cause problems for metaphysical realism; if so, we need not trouble ourselves much with Bays' dilemma. Both Bays himself (2001, pp. 340–1 2007, pp. 128–32) and Luca Bellotti (2005, pp. 405–8) suggest that this "fallback response" might be the best reaction to Bays' dilemma.

I would prefer to attack Bays' dilemma head-on. I shall agree that Bays is on to something, in the case of the constructivisation argument. However, in the case of the Skolemisation and permutation arguments, I shall argue that the metaphysical realist *must* accept that there *are* unintended models for the theories she cares about. The prime reason for this is as follows. To apply a Skolemisation or permutation argument against T_0, it is normally sufficient for dialectical purposes to work in theories *weaker* than T_0. Accordingly, neither *rejection* nor *nonchalance* is an option for the metaphysical realist. That is the primary contention of this paper.

To show all this, I shall need to examine a handful of model-theoretic results in close detail. Fortunately, most of the required definitions and results are relatively elementary. Those which I assume as "background knowledge" during the text are discussed at length in §8, which constitutes a technical appendix.

2 The constructivisation argument

In this section, I shall outline Bays' dilemma in more detail, focussing specifically on the constructivisation argument. I agree that Bays' dilemma raises a serious problem for the constructivisation argument, but I shall postpone the question of just *how* serious the problem is until §6.

2.1 Putnam's constructivisation argument

The target of Putnam's constructivisation argument is a metaphysical realist about sets. The metaphysical realist in question endorses the theory ZF, and is now considering whether or not "$\mathbf{V} = \mathbf{L}$" is true. (This is the so-called Axiom of Constructibility, which is independent of ZFC.) Putnam aims to show that the metaphysical realist has no way even to explain *what* she is considering.

Putnam begins his argument by being maximally concessive to the metaphysical realist. In particular, he allows that empirical measurements might seem to tell against the truth of "$\mathbf{V} = \mathbf{L}$". For example, we might

define a subset $s \subseteq \mathbb{N}$ from an ω-sequence of tosses of an idealised utterly random coin, by stipulating that $n \in s$ iff the n^{th} coin-toss lands heads. Since s is generated by an utterly random procedure, we have no reason to suppose that s must be a *definable* subset of $\mathbb{N}$. So it seems that s could end up being "genuinely nonconstructible". In which case, the metaphysical realist can imagine a situation in which empirical evidence seems to establish that "$\mathbf{V} = \mathbf{L}$" is false.

Putnam's aim is to show that, contrary to initial appearances, even this evidence would have no bearing at all on the truth-value of "$\mathbf{V} = \mathbf{L}$". His argument is based upon the following claim (Putnam 1980, p. 468):

Putnam's Claim. For any $s \subseteq \mathbb{N}$, there is an ω-model, $\mathcal{M}$, of $\mathrm{ZF} + \mathbf{V} = \mathbf{L}$, such that s is represented in $\mathcal{M}$.

If this Claim is correct, then what are we to make of the metaphysical realist's thought that s is "genuinely nonconstructible"? Only that $\mathcal{M}$ is an *unintended* model. But why is it *un*intended? $\mathcal{M}$ satisfies the central core of our set theory, ZF, "and we have gone to great length to make sure it satisfies all operational constraints as well", since $\mathcal{M}$ represents s (Putnam 1980, p. 469). What more is required to make a model *intended*? That is the initial challenge posed by the constructivisation argument.

2.2 Class-models and Putnam's "proof"

The metaphysical realist might attempt to answer Putnam's challenge directly, and explain why some feature of $\mathcal{M}$ renders it unintended. For example, Bellotti (2005, pp. 401–3) thinks that $\mathcal{M}$ is unintended because it is not wellfounded. However, there is a more direct response. If we can undermine Putnam's Claim, the entire constructivisation argument will unravel. And indeed, Bays (2001, pp. 335–6; 2007, pp. 119–23) and Velleman (1998) have noted a serious problem with Putnam's "proof" of his Claim. This problem depends on noting the crucial difference between set-models and class-models.

Let T_0 be any object theory the metaphysical realist might want to advance. Let T_1 be Putnam's model theory. T_1 is itself, strictly, a set theory; that is, the basic objects that T_1 talks about are sets. (If we like, we can allow urelements into T_1; this will not change anything essential.) When we say, in T_1, that there is a model of some theory T_0, we are saying that T_1 entails the existence of a *set* with certain properties.

Set theories, like T_1, can be harmlessly augmented with talk about definable classes. To keep this harmless, these definable classes must not be thought of as new objects, but "as abbreviations for expressions not involving them" (Kunen 1980, p. 24).[2] For example, for some monadic predicate "F", we might define a class:

$$\mathbf{F} := [x \mid F(x)]$$

We can then treat class membership as an abbreviation for predication, so that "$x \,\varepsilon\, \mathbf{F}$" abbreviates "$F(x)$".

Once we augment T_1 with talk about classes, it becomes possible to say, in T_1, that $\mathcal{A}$ is a *class-model* of T_0. This is to say that, for every sentence ϕ of T_0:

$$\mathrm{T}_1 \vdash \phi^{\mathcal{A}}$$

where $\phi^{\mathcal{A}}$ is the *relativisation* of ϕ to $\mathcal{A}$. The relativisation of a formula is defined recursively (see Kunen 1980, p. 112); in the special case where T_0 is a pure set theory, we simply specify a domain $\mathbf{A}$, and then define:

$$\begin{aligned}
(x = y)^{\mathcal{A}} &\text{ is } (x = y)\\
(x \in y)^{\mathcal{A}} &\text{ is } (x \in y)\\
(\phi \wedge \psi)^{\mathcal{A}} &\text{ is } (\phi^{\mathcal{A}} \wedge \psi^{\mathcal{A}})\\
(\neg\phi)^{\mathcal{A}} &\text{ is } \neg(\phi^{\mathcal{A}})\\
((\exists x)\phi)^{\mathcal{A}} &\text{ is } (\exists x \,\varepsilon\, \mathbf{A})\phi^{\mathcal{A}}
\end{aligned}$$

So, if ϕ is a formula in the austere language of set theory, $\phi^{\mathcal{A}}$ is obtained from ϕ just by restricting all of the quantifiers of ϕ to $\mathbf{A}$. In §4.3, I shall show how to define the relativisation of a formula when T_0 is *not* a pure set theory. However, in discussing how Bays' dilemma applies to the constructivisation argument, it does no harm to assume that the metaphysical realist's favourite theory, T_0, is just the pure set theory ZF.

There are now two types of model on the table: set-models and class-models. In what follows, when I say "model", I invariably mean "set-model", and when I want to talk about class-models, I shall explicitly mention that they are *class*-models. The difference between (set-)models and class-models is absolutely crucial. To see this, consider the Skolem Hull Theorem: for any model $\mathcal{A}$, there is a countable submodel, $\mathcal{B}$, which is elementarily equivalent to $\mathcal{A}$ (see §8.2). Disaster ensues if we forget that "model" here means "set-model", and suppose that the theorem holds for class-models. To see this, note that ZFC trivially entails the existence of a class-model of ZFC, by defining the class-sized domain as $[x \mid x = x]$. If we could apply the Skolem Hull construction to class-models, then we would have a countable class-model, $\mathcal{A}$, of ZFC. Since the domain of $\mathcal{A}$ would be countable, it would be a set (because there is no cofinal map from ω to the class of all ordinals). So we would have proved, in ZFC, the existence of a set-model of ZFC. By Gödel's Second Incompleteness Theorem, we would have proved that ZFC is inconsistent. The clear moral is that the Skolem Hull Theorem does *not* apply to class-models.

With all this in mind, we shall now consider the core of Putnam's "proof" of his Claim:[3]

#Proof: Putnam's Claim amounts to the following Π_2-sentence:

$$\phi := (\forall s)(\exists M)(s \subseteq \mathbb{N} \rightarrow (s \in M \wedge M \vDash \mathrm{ZF} + \mathbf{V} = \mathbf{L}))$$

Shoenfield's Absoluteness Lemma states that $\psi^{\mathbf{L}} \leftrightarrow \psi$, for any Π_2-sentence ψ. So to prove ϕ, it suffices to prove its relativisation to $\mathbf{L}$:

$$\phi^{\mathbf{L}} = (\forall s \,\varepsilon\, \mathbf{L})(\exists M \,\varepsilon\, \mathbf{L})(s \subseteq \mathbb{N} \rightarrow (s \in M \wedge M \vDash \mathrm{ZF} + \mathbf{V} = \mathbf{L}))$$

For every $s \varepsilon \mathbf{L}$, "there is a model—namely $\mathbf{L}$ itself—which satisfies "$\mathbf{V} = \mathbf{L}$" and contains s. By the [Skolem Hull] Theorem, there is a countable submodel[, M,] which is elementary equivalent to $\mathbf{L}$ and contains s.... By [Gödel's Condensation Lemma, M] itself lies in $\mathbf{L}$" (Putnam 1980, p. 468). This establishes $\phi^{\mathbf{L}}$, as required. □

As Bays and Velleman note, the problem here is that $\mathbf{L}$ is a proper class, not a set. We just saw that the Skolem Hull Theorem can only be applied to sets, and not to proper classes. So the "proof" fails.

2.3 Applying Bays' dilemma to the constructivisation argument

To repair the fallacious "proof", we might present the argument in a theory in which "the $\mathbf{L}$ of ZF" (so to speak) can be treated as a set, for then the Skolem Hull Theorem could be applied to it after all. One option is to offer the proof in ZFK, which adds to ZF the claim that there is some inaccessible cardinal κ; and Velleman (1998), Bays (2001, pp. 338, n.6; 2007, pp. 123–4) and Bellotti (2005, p. 396) all consider repairing the proof in this way. Now we can argue thus:

Constructivisation Theorem (ZFK)**.** Let s be a subset of $\mathbb{N}$. There is a model, $\mathcal{M}$, of $\mathrm{ZF} + \mathbf{V} = \mathbf{L}$, such that for each $n \in s$, $\mathcal{M} \vDash (\underline{n} \in \underline{s})$.

Proof Sketch. The core of the argument is that, by Shoenfield's Absoluteness Lemma, it suffices to prove:

$$\phi^{L_\kappa} := (\forall s \in L_\kappa)(\exists M \in L_\kappa)(s \subseteq \mathbb{N} \rightarrow (s \in M \wedge M \vDash \mathrm{ZF} + \mathbf{V} = \mathbf{L}))$$

Given any $s \in L_\kappa$, let M be a countable Skolem Hull of L_κ containing s. Then $M \equiv L_\kappa \vDash \mathrm{ZF} + \mathbf{V} = \mathbf{L}$ (see Kunen 1980, pp. 132, 169–70). Moreover $M \subsetneq L_\kappa$, so $M \in L_\kappa$, by Gödel's Condensation Lemma. □

The parenthetical remark after the theorem's name indicates that this is a theorem of ZFK. However—and this is crucial—this proof cannot be given in ZF, since ZF cannot prove the existence of L_κ. This gives the metaphysical realist two options. On the one hand, since ZFK is strictly stronger than ZF, the metaphysical realist can avoid Putnam's argument by simply *rejecting* ZFK. On the other hand, she might accept ZFK and so accept that there are nonstandard models of ZF, but she could remain *nonchalant* about the existence of nonstandard models of ZF, for only a nonstandard

model of ZFK would worry her. Bays' dilemma thus successfully applies to the argument just given.

This is already a fair criticism of Putnam's 1980-argument. However, it is worth asking whether the applicability of Bays' dilemma is an *intrinsic* feature of the constructivisation argument, or whether it is merely a contingent artefact of the *particular* proof we just considered. Bays argues that it is intrinsic, for the following reason: Putnam's Claim entails the existence of a model of ZF, but Gödel's Second Incompleteness Theorem shows that ZF cannot itself prove the existence of any such model; so Putnam's Claim cannot be proved in ZF or any weaker theory. Consequently, Bays' dilemma can be applied *whenever* we try to argue for something like Putnam's Claim. For this reason, Bays calls his dilemma "*inescapable*" (2001, pp. 338–40).

I entirely agree with Bays that the constructivisation argument faces difficulties. However, I am less certain about Bays' diagnosis that these difficulties are "*inescapable*". To explain why, it will help to set aside the constructivisation argument for now (until §6), and turn my attention to the Skolemisation and permutation arguments. These arguments are, in any case, of wider philosophical interest than the constructivisation argument, because they are much more central to Putnam's model-theoretic assault on metaphysical realism.

3 The Skolemisation Argument

Bays (2001, pp. 338–40, n.7) claims that his dilemma hamstrings the Skolemisation argument in just the same way that it hamstrings the constructivisation argument. In this section, I shall show that Bays is mistaken. The Skolemisation argument uses a theorem which can be proved, with full generality, in very weak model theories. Accordingly, Bays' dilemma does not threaten the argument.

3.1 The Skolemisation argument

Putnam's target is a metaphysical realist who has advanced a theory, T_0, whose intended interpretation is uncountable.[4] Putnam's Skolemisation argument proceeds by invoking the Completeness Theorem of first-order logic.

Completeness Theorem. Let T be any consistent countable set of sentences of a first-order language. There is a model $\mathcal{N} \vDash \mathrm{T}$ whose domain is $\mathbb{N}$. □

Accordingly, if the metaphysical realist's theory T_0 has any models, it has an unintended countable model $\mathcal{N}_0$. From a sufficiently broad perspective, this Skolemisation argument and the constructivisation argument raise the same challenge for the metaphysical realist. The metaphysical realist must

explain why $\mathcal{N}_0$ is an *unintended* model, even though $\mathcal{N}_0$ models the metaphysical realist's theory, T_0. (Putnam 1980, pp. 465–6; also see Skolem 1922. Putnam 1980 actually uses the Skolem Hull Theorem, rather than the Completeness Theorem; I discuss this version of the argument in §5.)

The metaphysical realist has a number of lines of response to this argument. She might attempt to explain that T_0 is a theory expressed in "full" second-order logic, in which case, the Completeness Theorem for first-order logic does not apply to it. More generally, she might attempt to explain why some feature of $\mathcal{N}_0$ renders it unintended. These kinds of reactions are well-explored elsewhere—perhaps most famously in Lewis 1984—and I shall not discuss them here. I wish to investigate whether Bays' dilemma offers a *new* line of response to Putnam's model-theoretic arguments.

3.2 Raising and finessing Bays' Dilemma

The Completeness Theorem is a theorem of WKL_0 (Simpson 1999, pp. 139–41).[5] If the metaphysical realist rejects WKL_0, she may be able to reject the Completeness Theorem too. In that case, we will not be able to run the Skolemisation argument against her. However, WKL_0 is an extremely weak mathematical theory; it is *much* weaker than Z, for example. So, in what follows, I shall assume that the metaphysical realist's theory, T_0, is stronger than WKL_0, and so that she accepts the Completeness Theorem. That is, I shall assume that Putnam's opponents are metaphysical realists about a reasonable chunk of mathematics. I shall also assume, for the whole of this subsection, that T_0 is an effectively axiomatisable theory. (I shall consider negation-complete theories in the next subsection.)

Being stronger than WKL_0, T_0 contains more than enough arithmetic to formalise talk about any effectively axiomatisable theory. So T_0 can formalise talk about itself: where $\mathscr{L}$ is the language of T_0, there is an $\mathscr{L}$-sentence which formalises such English expressions as "T_0 is a consistent countable set of sentences".[6] Now consider this proposition:

T_0-conditional (T_0)**.** If T_0 is a consistent countable set of sentences, then there is a model $\mathcal{N}_0 \vDash T_0$ whose domain is $\mathbb{N}$. □

We ought to think of the T_0-conditional as a long formal sentence in the language $\mathscr{L}$. Crucially, this $\mathscr{L}$-sentence is a *theorem of* T_0 (as indicated by the parenthetical remark after the theorem's name). Recall that the Completeness Theorem is a theorem of WKL_0, and so is a theorem of T_0; the T_0-conditional is then obtained simply by instantiating (the formal code for the effective axiomatisation of) T_0 into the Completeness Theorem. So the metaphysical realist must think that the T_0-conditional is true.

Of course, the T_0-conditional does not tell us that T_0 *has* some non-standard model $\mathcal{N}_0$. To show that, we would need to discharge the conditional's antecedent. On pain of Gödel's Second Incompleteness Theorems, the metaphysical realist cannot prove, in T_0, that T_0 is a consistent countable set of sentences. To prove that T_0 is consistent, she would have to

move to some other theory, T_1, such as $T_0 + \mathrm{Con}(T_0)$. This is where Bays' dilemma is supposed to arise (recall §2.3). On the one hand, the metaphysical realist may *reject* T_1. She will then accept the conditional, but can happily reject the consequent. On the other hand, she may accept T_1, and so accept that there is an unintended model of T_0, but treat this result with *nonchalance*, since she would only care about an unintended model of T_1.

We can immediately rule out the strategy of rejecting T_1. A metaphysical realist about T_0 thinks that T_0 is true. Since truth entails consistency, she must accept both T_0 and $\mathrm{Con}(T_0)$. That is just to accept T_1, as Bellotti (2005, p. 405) has noted.

The metaphysical realist must therefore attempt to remain nonchalant; Bays must emphasise that Putnam has shown nothing yet about T_1. However, Putnam can run exactly the preceding argument against T_1. T_1 is strictly stronger than WKL_0, and is still effectively axiomatisable. So we have:

T_1-conditional (T_0)**.** If T_1 is a consistent countable set of sentences, then there is a model $\mathcal{N}_1 \vDash T_1$ whose domain is $\mathbb{N}$. □

Again, this formal $\mathscr{L}$-sentence is a theorem of T_0. And again, the metaphysical realist thinks that T_1 is true, so she can hardly deny that T_1 is consistent. So she must accept a strictly stronger theory, $T_2 = T_1 + \mathrm{Con}(T_1)$, and attempt to remain nonchalant. But by exactly the same argument, involving the T_2-conditional (another theorem of T_0), she must ascend to a stronger theory, $T_3 = T_2 + \mathrm{Con}(T_2)\ldots$. In short, the metaphysical realist is forced to commit to every theory in an iterated *consistency-sequence*. This is an ordinal sequence of theories such that:[7]

$$T_{\alpha+1} := T_\alpha + \mathrm{Con}(T_\alpha) \quad \text{for some appropriate definition of "Con(X)"}$$
$$T_\alpha := \bigcup_{\beta<\alpha} T_\beta \qquad \text{when } \alpha \text{ is a limit ordinal}$$

No matter how far the metaphysical realist runs along this consistency-sequence, Putnam's model-theoretic arguments are waiting for her. At every stage α, T_α is an effectively axiomatised theory and the T_α-conditional is a theorem of T_0. And at every stage α, she is committed to the antecedent of the T_α-conditional. So, every time she moves from one theory to the next in the consistency-sequence, she merely steps out of the frying pan and into another, very slightly larger, frying pan.

The crucial point is this. The metaphysical realist *herself* always supplies the assumption that her favourite theory has a model. So Putnam's argument only requires the conditional: If a theory has any models, then it has an unintended model. *This is the fundamental reason why the Skolemisation argument resists Bays' dilemma.*

In desperation, the metaphysical realist might make the following argument. The Skolemisation argument starts with the metaphysical realist

presenting some theory, T_0, and Putnam then uses the Completeness Theorem to generate a countable model $\mathcal{N}_0 \vDash \mathrm{T}_0$. Now, if the metaphysical realist is allowed to add new sentences to her theory, moving from T_0 to T_1, we have no guarantee that $\mathcal{N}_0 \vDash \mathrm{T}_1$. For this reason, it seems important to keep the metaphysical realist's theory *fixed*. However, the metaphysical realist has just agreed that she is committed to every theory in an infinite sequence of theories. She may then claim that *no* single fixed theory captures all of her commitments. In which case the metaphysical realist might object that "Putnam's argument goes wrong at a very early point", because the "whole apparatus of (fixed) theories and models seems decidedly inappropriate."[8]

I cannot see how this response would help the metaphysical realist. The metaphysical realist has simply pointed out that she cannot articulate all of her commitments. This is just *another* problem for her position, and it does not help her to deal with the Skolemisation argument. *Whatever* claims the metaphysical realist manages to make, Putnam can show her that those claims have a countable model. This is all that the Skolemisation argument requires.

3.3 A detour through negation-complete theories

We have seen that the Skolemisation argument is untouched by Bays' dilemma. However, it is worth noting that Putnam's "apparatus of (fixed) theories and models" is entirely *unobjectionable* in the present context. To show this, I shall embark on a brief but interesting detour, concerning negation-complete theories.

Recall the *credo* of metaphysical realism, as explained at the start of this paper. The *credo* states: there is some fixed totality of objects, which we hope that our best theory truly describes, where truth consists of some sort of correspondence relation. So let $\mathscr{L}$ be the language of T_0. Since the world is some structure, $\mathcal{W}$, we can introduce $\mathrm{T}_\star$ as the set of all $\mathscr{L}$-sentences satisfied by $\mathcal{W}$. More briefly: $\mathrm{T}_\star$ contains every truth expressible in $\mathscr{L}$.

Putnam (1983, p. ix, n) maintained that we can run a Skolemisation argument *even* against a metaphysical realist who advances $\mathrm{T}_\star$. It would not matter much if Putnam were wrong about this: since $\mathrm{T}_\star$ is negation-complete, and hence not effectively axiomatisable, no metaphysical realist could ever *actually* advance $\mathrm{T}_\star$. However, the purpose of this detour is to show that Putnam is, indeed, correct here. The detour is interesting, in the present context, because it vindicates the "apparatus of (fixed) theories and models".

In §3.2, I showed how to run an argument against a metaphysical realist who advocates an effectively axiomatisable theory, T_α: simply note that T_0 entails the T_α-conditional, and then sit back and allow the metaphysical realist's own assumptions to do the rest. So, to run a Skolemisation argument against $\mathrm{T}_\star$, we might want to start by asking whether T_0 entails

the following sentence:

#$\mathrm{T}_\star$-conditional. If $\mathrm{T}_\star$ is a consistent countable set of sentences, then there is a model $\mathcal{N}_\star \vDash \mathrm{T}_\star$ whose domain is $\mathbb{N}$.

This question is a trick-question. $\mathrm{T}_\star$ is negation-complete, sound and stronger than WKL_0. Accordingly, $\mathrm{T}_\star$ is not effectively axiomatisable, and so no formula of $\mathscr{L}$ defines (the axiom base of) $\mathrm{T}_\star$. So "$\mathrm{T}_\star$ is a consistent set of sentences" is not even a *sentence* of $\mathscr{L}$. *A fortiori*, "the $\mathrm{T}_\star$-conditional" (scare-quotes needed) is not even an $\mathscr{L}$-sentence. If we want to convince the metaphysical realist that $\mathrm{T}_\star$ has a countable model, we clearly cannot follow the route pursued in §3.2.

Here is an alternative route. No $\mathscr{L}$-sentence expresses that $\mathrm{T}_\star$ is consistent, but nothing stops the metaphysical realist from expanding her language. Let $\mathscr{K}$ be a language in which one can express that $\mathrm{T}_\star$ is consistent (perhaps $\mathscr{K}$ contains a truth predicate for $\mathscr{L}$). Once she can express the thought that $\mathrm{T}_\star$ is consistent, the metaphysical realist must immediately accept that $\mathrm{T}_\star$ is consistent, since she thinks that $\mathrm{T}_\star$ is true. An appropriate completeness theorem will now force her to accept that $\mathrm{T}_\star$ has a countable model, as required.

However, the argument just given took place neither *within* T_0 nor $\mathrm{T}_\star$, but within some *wider* formal semantic theory. So our metaphysical realist may respond to this argument as follows. She first claims that she cannot talk about models for $\mathrm{T}_\star$ *at all* while using (the language of) $\mathrm{T}_\star$. She explains that, if she wants to talk about models for $\mathrm{T}_\star$, then she must move to some wider *formal* semantic theory, S (given in the richer language, $\mathscr{K}$). However, she continues, once she has accepted S, she no longer cares that $\mathrm{T}_\star$ has unintended models; she would only care if the theory $\mathrm{T}_\star + \mathrm{S}$ had unintended models, and she has not yet been shown that it does.

The response we are imagining resurrects Bays' strategy of *nonchalance*, in the special case where the metaphysical realist's favourite theory is negation-complete. Here, though, the metaphysical realist's nonchalance rests upon the following:

(***c***) **Model-formalism.** The metaphysical realist maintains that talking about models only makes sense within the context of some formal model theory, given in some formal language.

Unfortunately for this metaphysical realist, such *model-formalism* is absolutely incompatible with metaphysical realism. I shall spend the remainder of this subsection explaining why.

We must again recall the metaphysical realist's *credo* that truth involves some sort of correspondence. Putnam suggested that we discuss the correspondence relation in terms of model-theoretic satisfaction. If the metaphysical realist had some *other* way to explain the correspondence relation, she should have spoken up much earlier. It would have saved us all a lot of time, since if we can discuss correspondence without invoking

model theory, model-theoretic arguments will just be *irrelevant* to metaphysical realism. So, we must suppose that the metaphysical realist agrees that correspondence is to be explained via models.

Suppose, now, that our metaphysical realist has adopted the strategy of *model-formalism*, in an attempt to resurrect the strategy of *nonchalance*. As a model-formalist, she thinks that *all* talk about models must take place within a formal model theory. If she is dealing with a negation-complete theory, such as $T_\star$, she must think that she cannot talk about models for $T_\star$ within $T_\star$ (for if she *could* do that, then she would have to accept that $T_\star$ has a countable model, about which she could not remain nonchalant.) So if she wants to talk about *correspondence* for $T_\star$, she must move to a wider *formal* semantic theory, S. But it is not indicative of metaphysical realism to discuss, in a formal theory S, the models of some formal theory $T_\star$; it simply indicates a willingness to engage in a certain branch of mathematics, namely, model theory. Nor is it indicative of metaphysical realism to talk in S about *correspondence* for $T_\star$; even the most ardent anti-realist can do that, so long as they are antirealists *about S itself* (or about some yet-wider theory). So, how *is* the metaphysical realist to indicate her metaphysical realism?

This question reveals an essential tension between metaphysical realism and model-formalism. Metaphysical realism demands that there is an *externalist* perspective on formal theories. Model-formalism demands that any perspective concerning correspondence must be *internal* to some formal theory. Model-formalism, then, yields an even stronger internalist conclusion than Putnam himself drew from his model-theoretic arguments: *Model-formalism denies that metaphysical realists can even give any content to the picture of an externalist perspective* (cf. Putnam 1992, pp. 353–4).

Evidently, the metaphysical realist cannot hope to resurrect Bays' dilemma by appealing to model-formalism. More generally, the metaphysical realist must avoid model-formalism entirely. She must maintain that her central *credo* "can only be stated with 'typical ambiguity'—i.e. it transcends complete formalization in any one theory" (Putnam 1978, p. 125). She must treat metaphysical realism as an *intuitive picture*, rather than as something totally formalisable. She might attempt to flesh out this picture via a theological parable (cf. Putnam 1983, pp. ix–x). The metaphysical realist imagines God establishing a correspondence relation between the theories of Earthbound humans and the objects of the world. As she sits by God's side and watches God at work, she sees that God is engaged in a practice much like that of building a model for a formal theory. She sees that certain features of God's correspondence relation behave like model-theoretic satisfaction. Of course, this *is* a metaphor: she is not really watching God build a model of her theory; she is simply *using* that theory and hoping that she is thereby saying true things (she is hoping that God is smiling on her). Nonetheless, outside the metaphor, she can retain the insight that certain features of correspondence behave like model-theoretic satisfaction.

In short, the metaphysical realist should hold that formal model theory helps *explicate* her claim that truth consists in some sort of correspondence relation. (This is to be contrasted with the model-formalist's mantra that talk about models is only meaningful *at all* within the context of some formal model theory.) However, this explication comes with a price: when the metaphysical realist sat with God, she saw that God had many possible correspondence relations to choose from. In particular, she saw that $T_\star$ has a model which is countable from God's perspective, and many other "unintended" models besides. She must tackle Putnam's model-theoretic arguments head-on.

3.4 Two metamathematical properties

The detour of the previous subsection was illuminating, but it is liable to distract attention from the central point of this paper. Allow me to repeat that central point. The Skolemisation argument avoids Bays' dilemma thanks to two metamathematical properties of the Completeness Theorem:

(**1**) It tells us that *any* consistent set of sentences has an "unintended" model.

(**2**) It is provable in an extremely *weak* model theory.

Of course, these properties do not allow Putnam to offer a formal *proof* that the metaphysical realist's favourite theory has a countable model. Fortunately, Putnam is not arguing with a theorem-prover; he is arguing with a philosopher with substantial prior metaphysical commitments. All Putnam needs to do is convince the metaphysical realist that her theories have countable models. He can do this by virtue of the two metamathematical properties just mentioned. By property (**1**), any theory which is any good has a countable model. By property (**2**), the metaphysical realist can see this with full generality—and so for any theory which she might *ever* be right to accept—on the basis of an extremely weak model theory, which she already *does* accept.

4 The permutation argument

I now turn to Putnam's permutation argument. Bays does not discuss the permutation argument in the context of his "dilemma"; my talk of "Bays' dilemma", in the context of the permutation argument, is by analogy with Bays' treatment of the Skolemisation and constructivisation arguments. However, the permutation argument can finesse Bays' dilemma in just the same way as the Skolemisation argument. The reason for this is that the permutation theorem exhibits the two crucial metamathematical features just mentioned.

4.1 Putnam's permutation argument

Intuitively, the permutation argument allows us to generate an unintended model by "shuffling" the reference and correspondence relations of the intended model. Formally, we employ the following theorem (see §8.1):

Permutation Theorem. Let T be a theory with a non-trivial model. T has multiple distinct isomorphic models. □

So, if $\mathcal{W}$ models the metaphysical realist's favourite theory T_0, then there is some permuted model, $\mathcal{P}$ which, by construction, is isomorphic to $\mathcal{W}$. *A fortiori*, both models make exactly the same sentences of T_0 true. Furthermore, since they are isomorphic, no possible sentence of the object language can be added to T_0 to tell them apart. So we have absolutely free choice as to whether to treat correspondence as given by model-theoretic satisfaction in $\mathcal{W}$, or by model-theoretic satisfaction in $\mathcal{P}$. This is Putnam's permutation argument (1981, pp. 32–8, 217–8).

4.2 Finessing Bays' dilemma

The question we need to address is familiar. When we permute the metaphysical realist's favourite theory, T_0, must we work in a theory that is strictly stronger theory than T_0, or can we work in some weaker theory?

It should come as no surprise to learn that the Permutation Theorem is provable in an *extremely* weak set theory. In particular, one can present the Permutation Theorem in theories *much* weaker than Z–I, that is, Zermelo set theory without an Axiom of Infinity (see §8.1). If the metaphysical realist does not accept some such theory, she cannot do much mathematics at all. So, as in my discussion of the Skolemisation argument, I shall assume that the metaphysical realist accepts enough mathematics to prove the Permutation Theorem.

The same key factors are now in place for the permutation argument as they were for the Skolemisation argument. In particular:

(**1**) The Permutation Theorem tells us that *any* set of sentences with a model has a permuted model.

(**2**) The Permutation Theorem can be proved, with complete generality, in extremely *weak* model theories.

Accordingly, if Bays attempted to deploy his dilemma against the permutation argument, we could reply with exactly the same argument as we considered in the case of the Skolemisation argument. By (**1**), any theory which is any good has a permuted model. By (**2**), the metaphysical realist can see this with full generality—and so for any theory which she might *ever* be right to accept—on the basis of an extremely weak model theory, which she already *does* accept. In short, the permutation argument absolutely resists Bays' dilemma.

The Permutation Theorem comes with an added bonus. The Completeness Theorem produces an "unintended" model from a given *theory*. Hypothetically at least, this allows the metaphysical realist space to respond by rejecting the "apparatus of (fixed) theories and models", and such thoughts occupied much of §§3.2–3.3. When considering the Permutation Theorem, we can streamline this discussion considerably, since

(**3**) The Permutation Theorem generates an "unintended" model from a given *model*.

The metaphysical realist thinks of the world as the intended model of some theory (perhaps "God's theory"). Accordingly, the Permutation Theorem immediately shows the metaphysical realist the following: *If there is a world at all, then there is a permuted world.*

4.3 A detour through permuted class-models

We have seen that the permutation argument resists Bays' dilemma admirably. Before moving on, I would like to embark on a second brief detour. This time, I would like to explore how the permutation argument applies to *class*-models.[9]

In §2.2, I explained how to relativise a sentence, ϕ, to a class-model, $\mathcal{A}$. There, I assumed that $\mathcal{A}$ was a class-model of pure set theory. However, we can liberalise the definition of the relativisation of a sentence, to allow class-models of any theory. As before, we first define the domain of $\mathcal{A}$ as a class, $\mathbf{A}$. We next define an *interpretation* class-function, $\mathbf{I}^{\mathcal{A}} : [x \mid x = x] \longrightarrow \mathbf{A}$, whose primary role is to assign "interpretations" to the individual constants of the language of T_0. For each atomic predicate "R" in the language of T_0, we also define a suitable class $\mathbf{R}^{\mathcal{A}}$. Now we define $\phi^{\mathcal{A}}$ recursively:

$$\begin{aligned}
(x = y)^{\mathcal{A}} &\text{ is } (\mathbf{I}^{\mathcal{A}}(x) = \mathbf{I}^{\mathcal{A}}(y)) \\
(R(x_1, \ldots, x_n))^{\mathcal{A}} &\text{ is } (\langle \mathbf{I}^{\mathcal{A}}(x_1), \ldots, \mathbf{I}^{\mathcal{A}}(x_n) \rangle \, \varepsilon \, \mathbf{R}^{\mathcal{A}}) \\
(\phi \wedge \psi)^{\mathcal{A}} &\text{ is } (\phi^{\mathcal{A}} \wedge \psi^{\mathcal{A}}) \\
(\neg\phi)^{\mathcal{A}} &\text{ is } \neg(\phi^{\mathcal{A}}) \\
((\exists x)\phi)^{\mathcal{A}} &\text{ is } (\exists x \, \varepsilon \, \mathbf{A})\phi^{\mathcal{A}}
\end{aligned}$$

Armed with this apparatus, we could run a class-level version of the permutation argument. We would start by fixing some non-trivial class-sized bijection $\boldsymbol{\pi}$ on the class-domain $[x \mid x = x]$. We would then define two class-models, $\mathcal{W}$ and $\mathcal{P}$, as follows:

$$\begin{aligned}
\mathbf{W} &:= [x \mid x = x] & \mathbf{P} &:= [x \mid x = x] \\
\mathbf{I}^{\mathcal{W}} &:= [\langle x, x \rangle \mid x = x] & \mathbf{I}^{\mathcal{P}} &:= \boldsymbol{\pi} \\
\mathbf{R}^{\mathcal{W}} &:= [\langle x_1, \ldots, x_n \rangle \mid R(x_1 \ldots x_n)] & \mathbf{R}^{\mathcal{P}} &:= [\langle \boldsymbol{\pi}(x_1), \ldots, \boldsymbol{\pi}(x_n) \rangle \mid R(x_1 \ldots x_n)]
\end{aligned}$$

for each predicate "R". Now, as long as T_0 is powerful enough to talk about class-models and to prove the Permutation Theorem (which is not

to ask very much), we can prove the existence of these class-models for T_0 *within* T_0 itself. In which case, we can simply ask the metaphysical realist, *within* her favourite theory, why she thinks that "R" refers to $\mathbf{R}^{\mathcal{W}}$, rather than $\mathbf{R}^{\mathcal{P}}$. When the question is put this way, Bays' dilemma never even gets a foot in the door.

Sadly, the question itself is illegitimate, since it treats "$\mathbf{R}^{\mathcal{W}}$" and "$\mathbf{R}^{\mathcal{P}}$" as names for genuine objects. Of course, there are theories, such as MK (Morse–Kelley class theory), which have *bona fide* objects called "classes". But, if MK is consistent, it obviously cannot prove the existence of a class-model for MK, where "class" is used in the sense employed by MK itself. So MK, and other theories like it, are irrevelant to this attempt to finesse Bays' dilemma. In the present context, as in §2.2, we must not think of classes as *bona fide* objects. Given this, there is literally *no* question of whether "R" refers to $\mathbf{R}^{\mathcal{W}}$ or $\mathbf{R}^{\mathcal{P}}$. The question simply dissolves.

In slightly more detail: to consider the truth of a sentence ϕ in the class-model $\mathbf{P}$ is to consider the relativised sentence $\phi^{\mathbf{P}}$. Logical connectives are unaffected by their relativisation and, in the case of $\mathbf{P}$, the relativisation of quantifiers to the domain $\mathbf{P}$ has no effect, since $\mathbf{P}$ is $[x \mid x = x]$. So it suffices to consider the relativisation of each atomic sentence $(R(a_1 \dots a_n))^{\mathcal{P}}$, namely:

$$\begin{aligned}(R(a_1 \dots a_n))^{\mathcal{P}} \text{ is } & \langle \mathbf{I}^{\mathcal{P}}(a_1), \dots, \mathbf{I}^{\mathcal{P}}(a_n)\rangle \,\varepsilon\, \mathbf{R}^{\mathcal{P}} \\ \text{is } & \langle \boldsymbol{\pi}(a_1), \dots, \boldsymbol{\pi}(a_n)\rangle \,\varepsilon\, \mathbf{R}^{\mathcal{P}} \\ \text{is } & R(a_1 \dots a_n)\end{aligned}$$

That is, the relativised sentence simply abbreviates the unrelativised sentence. In short, once we remove all talk about classes—as we are honour-bound to do, since classes are not *bona fide* objects—it becomes clear that class-models pose no problems for metaphysical realists.

For exactly the same reason, though, class-models provide no comfort for metaphysical realists. The metaphysical realist was happy to talk about set-models in order to explicate talk about correspondence (see §3.3). It is easy to explicate correspondence in terms of a set-model: there is a set which is the *world*, and a set of *words*, and an interpretation function (also a set) between the two, which is to be thought of as *reference*. In the case of a class-model, properly speaking, there *is* no domain, nor *is* there any function from the words to the world, since there are (speaking strictly) no classes. Talk of "correspondence" between a theory and a class-model must be treated in an utterly *deflationary* sense.

In short, class-models—permuted or otherwise—have no interesting bearing on metaphysical realism. This concludes our second detour.

5 The Skolemisation argument again

We have seen that Bays' dilemma affects neither the most straightforward version of the Skolemisation argument, nor the permutation argument.

These form the mainstay of Putnam's model-theoretic arguments against metaphysical realism. Before I pass final judgment on the constructivisation argument, it is worth briefly commenting on three variations of Putnam's Skolemisation argument.

The Skolemisation argument raises problems for metaphysical realists who want their favourite theory to have an uncountable model. In §3, we based this argument on the Completeness Theorem. But there are many other "Skolemising" results which we could use instead, such as (see §8.2):

Skolem Hull Theorem (ZD_ω)**.** Given any structure $\mathcal{A}$ and any countable set $S \subseteq A$, there is a countable structure $\mathcal{H} \prec \mathcal{A}$ with $S \subseteq H$. □

The Skolem Hull Theorem is a theorem of ZD_ω (that is, Zermelo set theory with the Axiom of Countable Dependent Choice). And, as with the Completeness Theorem, if the metaphysical realist's favourite theory is at least as strong as ZD_ω, then the metaphysical realist must accept the Skolem Hull Theorem. Moreover, unlike the Completeness Theorem, the Skolem Hull Theorem has property (**3**): it generates a countable submodel from a given model. Accordingly, it immediately shows the metaphysical realist the following: *if there is a world at all, then there is a countable world.*

Now, when considering a Skolemisation argument against a metaphysical realist about the pure sets, we only need to worry about the interpretation of "$\in$". However, it is reasonably common to insist that any *intended* interpretation of "$\in$" must be transitive.[10] If it is legitimate for the metaphysical realist to do this, then we face a difficulty: neither the Completeness Theorem nor the Skolem Hull Theorem generate transitive models, so the Skolemisation arguments depending upon them collapse.[11]

However, this difficulty can be dealt with by moving from the Skolem Hull Theorem to the following result (see §8.3):

Transitive Skolem Theorem (ZD_ω)**.** Given any transitive model of pure set theory, $\mathcal{B}$, there is a countable transitive model $\mathcal{A} \equiv \mathcal{B}$. □

If we generate $\mathcal{A}$ using the Transitive Skolem Theorem, then $\in^{\mathcal{A}}$ is transitive. Accordingly, if we rely upon the Transitive Skolem Theorem, we can dodge altogether the question of whether it is legitimate to insist that the intended interpretation is transitive. As a further bonus, the Transitive Skolem Theorem retains property (**3**).

The Skolem Hull and Transitive Skolem Theorems are obviously very powerful weapons against a metaphysical realist. However, more powerful weapons come with higher price-tags. The proofs of these theorems depend (essentially) on the Axiom of Countable Dependent Choice (see §8.4). By contrast, the Completeness Theorem can be proved without *any* choice principle. Accordingly, for all their virtues, we might doubt whether these two flashier theorems have property (**2**).

More specifically, we might worry that a metaphysical realist may be able to avoid Putnam's Skolemisation argument as follows. She first argues

that the intended interpretation of "$\in$" is transitive. She then *rejects* the model-theoretic resources needed to prove Transitive Skolem's Theorem; for example, she may accept ZF, but accept no choice principles. She finally maintains that she has been given no reason to believe that her favourite set theory, ZF, has any countable transitive interpretations.

In practice, I cannot imagine many contemporary metaphysical realists making this argument. Rightly or wrongly, most self-defined "realists" now seem to accept ZFC. Since ZFC is much stronger than ZD_ω, such metaphysical realists must accept Transitive Skolem's Theorem. That said, I must leave the ultimate decision on this point to the conscience of individual metaphysical realists.

However, even if the metaphysical realist accepts no choice principles, there is some mileage in the Transitive Skolem Theorem. In ZF, we can prove that ZFC is consistent if ZF is consistent (see Kunen 1980, p. 175). So if a metaphysical realist accepts that ZF is true, we can convince her that it is *consistent* to believe that there is some transitive countable model of her favourite theory. This raises the *possibility* that there is some countable transitive model, and this mere possibility may be sufficient to cause the metaphysical realist some worries. (In some sense, this recalls the "fallback response" to Bays' dilemma, mentioned in §1.)

Michael Potter has suggested that we may be able to do better still against metaphysical realists who are chary of choice. Potter draws attention to a close relative of the Transitive Skolem Theorem:

Submodel Skolem Theorem (ZF)**.** For any transitive model $\mathcal{B} \vDash \mathrm{ZF}$, there is a countable transitive model $\mathcal{A} \vDash \mathrm{ZF}$ such that $\mathcal{A} \subseteq \mathcal{B}$. □

Unlike the Transitive Skolem Theorem, the Submodel Skolem Theorem can be proved *without* any choice principles (see §8.4). But, Potter suggests, we can use the Submodel Skolem Theorem in any philosophical argument where we might have wanted to use the Transitive Skolem Theorem. For this reason, Potter claims that "the issue about the use of choice here is a red herring" (2004, p. 241).

I think that this may be mistaken. The Submodel Skolem Theorem is certainly often mentioned in the literature on Skolem's paradox (though see my cautionary comments at the end of §8.3 and §8.4). However, the Submodel Skolem Theorem has some obvious drawbacks. Recall that we want to base our model-theoretic arguments on theorems with nice metamathematical properties. In particular, we want to show that *any* consistent set of sentences has an unintended model. The Submodel Skolem Theorem does not show this. Given any theory T which extends ZF and has a transitive model $\mathcal{B}$, the Submodel Skolem Theorem guarantees that there is some countable model $\mathcal{A} \vDash \mathrm{ZF}$ such that $\mathcal{A} \subseteq \mathcal{B}$. It does *not*, though, prove that $\mathcal{A} \vDash \mathrm{T}$. Accordingly, the metaphysical realist may be able to make statements in the *object language* that rule out $\mathcal{A}$ as an intended model. To illustrate this: for the proof of the Submodel Skolem Theorem given in §8.4, $\mathcal{A} \vDash \mathbf{V} = \mathbf{L}$; so if T contains "$\mathbf{V} \neq \mathbf{L}$", then $\mathcal{A} \nvDash \mathrm{T}$.

The Submodel Skolem Theorem therefore lacks one of the crucial metamathematical properties that enabled us to defend the model-theoretic arguments against Bays' dilemma; unlike the Skolem Hull Theorem and the Transitive Skolem Theorem, it lacks property (**1**). Accordingly, the status of the Axiom of Countable Dependent Choice is not a "red-herring".

6 The constructivisation argument again

With these lessons in mind, I shall end by returning to the constructivisation argument. In §2, I noted that the Constructivisation Theorem could be proved, for models of ZF, in the strong system ZFK. I explained why this was problematic, and asked whether the problem was *intrinsic* to the constructivisation argument, or an artefact of the *particular* proof of the Constructivisation Theorem. I can now answer that question.

The proof of the Constructivisation Theorem that we considered was not given, with complete generality, in a weak model theory. Accordingly, the Theorem may lack property (**2**). However, it is still not clear whether or not this is an *intrinsic* defect of the constructivisation argument. For all I know, a hitherto-undiscovered but cunning proof of a suitable theorem might exhibit property (**2**), whilst allowing us to run some constructivisation argument.

We cannot, though, show that *any* consistent set of sentences has a constructible model for, quite obviously, ZF + $\mathbf{V} \neq \mathbf{L}$ does not have a constructible model. Accordingly, any theorem upon which Putnam could base a constructivisation argument *must* lack property (**1**). As with the Submodel Skolem Theorem, then, we could always in principle show that Putnam's models are unintended simply by making statements in the object language. For example, we could simply insist that "$\mathbf{V} \neq \mathbf{L}$" must be satisfied. This is Bellotti's reaction to the constructivisation argument: he objects that "among our best axioms there is presumably some axiom (e.g., the existence of a measurable cardinal) ruling out $\mathbf{V} = \mathbf{L}$" (2005, p. 406).

Naturally, this raises the question of whether it is *legitimate* simply to state in the object language something which entails or contradicts "$\mathbf{V} = \mathbf{L}$". I do not know how to settle this question, without embarking on the question of what legitimates set-theoretic statements *in general*. Is the Axiom Scheme of Replacement legitimate? What about the Axiom of Powersets? What, if anything, justifies Choice rather than Determinacy? What about other candidate axioms? These are extremely difficult questions, but until we have some general method for answering them, I do not see how we can address whether or not it is legitimate to add some axiom which either entails or contradicts "$\mathbf{V} = \mathbf{L}$".

Strikingly, though, the Skolemisation and permutation arguments do *not* require this prior discussion. This is why property (**1**) is so important. It guarantees that, *whatever* theory the metaphysical realist advances, she will have to deal with unintended models of that theory.

7 Concluding Remarks

Bays presented a dilemma against Putnam's model-theoretic arguments. I have shown that the dilemma has no purchase, so long as we can base the argument in question on a theorem with two key metamathematical properties:

(**1**) The theorem must tell us that *any* worthwhile set of sentences has an unintended model of the required kind.

(**2**) The theorem must be provable, with full generality, in a *weak* model theory.

Other properties may be desirable; for example, the ability to generate unintended models from a given *model*, or the ability to generate *transitive* models. But whenever our theorem has properties (**1**) and (**2**), the dialectic kicks in against the metaphysical realist. She must accept that her favourite theory has an unintended model.

Ultimately, then, the metaphysical realist must attempt to specify in general what makes a model (un)intended. Of course, there are many things that the metaphysical realist might say at this point. Indeed, this paper leaves open the possibility that the metaphysical realist can decisively *refute* Putnam's arguments. All I have shown is that metaphysical realists cannot question the metamathematics of the mainstay of Putnam's model-theoretic arguments. The Skolemisation and the permutation arguments are immune to metamathematical challenge.

8 Technical appendix

In this appendix, I state (and prove) most of the results discussed in the main paper. The main exception to this is the Completeness Theorem. The specific proof of the Theorem in WKL_0 (discussed in §3) appears in Simpson 1999, and I have nothing to add to Simpson's presentation.

Before hitting the proofs, we need some notational conventions. Z is standard *Zermelo set theory*. That is to say that Z has the axioms: Extensionality, Powersets, Pairs, Union, Infinity, Regularity, Purity and all Separation instances. Two of these axioms require comment.

First: Regularity guarantees that the sets are wellfounded. This is easily achieved by coupling Replacement with some axiom of Foundation. However, Z *lacks* Replacement. So, I shall sometimes (silently) assume that the set theory is presented as a theory of levels, as in Potter 2004. (Potter also shows how to develop a Replacement-free theory of ordinals.)

Second: most metaphysical realists about sets do not think that *everything* is a set. So it is obviously unreasonable to lumber metaphysical realists with the Axiom of Purity. If we were to drop the assumption of Purity, we could obtain (essentially) the same results; however, the Appendix

would become longer and some of the proofs would become less clear. This is why I include the Axiom of Purity among my assumptions.

I shall work *in* Z and related theories, and I shall also discuss models *of* Z and related theories. To discuss the related set theories, we require some naming-conventions. The conventions I employ are best illustrated by example. The theory ZF–IP is Zermelo set theory ("Z"), with the Axiom scheme of Replacement ("F"), but without Infinity ("–I") and without Powersets ("–P"). To further illustrate these conventions, here is a trivial result, provable in both ZF–IP and Z–I (Zermelo set theory without Infinity):

Proposition 1 (ZF–IP and Z–I)**.** For any x and y, both of the following exist: $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$ and $x \times y = \{\langle s, t\rangle \mid s \in x \wedge t \in y\}$.

Proof. By Extensionality, note the following identities:

$$\langle x, y\rangle = \{\{x, x\}, \{x, y\}\}$$
$$x \times y = \bigcup\{\{\langle s, t\rangle \mid s \in x\} \mid t \in y\}$$
$$x \times y = \{\langle s, t\rangle \in \mathcal{P}(\mathcal{P}(x \cup y)) \mid s \in x \wedge t \in y\}$$

The first construction involves Pairs. The second involves Replacement and Union. The third involves Powersets, Union and Separation. □

The theory (or theories) in which the proof is given are stated in parentheses after the theorem's name. The following (standard) notational conventions are also employed:

- Calligraphic fonts are used for models: $\mathcal{A}, \mathcal{B}, \ldots$
- Italicised fonts are used for that model's set-sized domain: $A, B, \ldots$
- Arbitrary finite sequences of objects, such as $a_1, \ldots, a_n$, or $b_1, \ldots, b_m$, are indicated by overlining: $\overline{a}, \overline{b}, \ldots$
- Suppose $a \in A$, and that $\mathcal{A}$ satisfies the atomic sentence "$F(a)$", where "a" is the name of a. In this case, we write $\mathcal{A} \vDash F(\underline{a})$, or $a \in F^{\mathcal{A}}$.

8.1 The Permutation Theorem

Lemma 2 (ZF–IP and Z–I)**.** Let $\mathcal{A}$ be any non-trivial structure. That is, suppose $\mathcal{A}$ contains at least two objects and:

- an object picked out by an individual constant of the signature; or
- a relation that is non-empty and non-universal; or
- a function that is non-empty and non-universal.

Then there is a structure $\mathcal{B} \cong \mathcal{A}$, such that $B = A$ but $\mathcal{B} \neq \mathcal{A}$.

Proof. First, we define a bijection $\pi : A \longrightarrow A$ other than identity. Let $a_0, \ldots, a_i \in A$ be distinct objects, and define a predicate of the model theory:

$$Fy := (\exists s)(\exists t)\Big(\langle s,t\rangle = y \wedge \big(\textstyle\bigwedge_{k=0}^{i}(s \neq a_k) \rightarrow s = t\big) \wedge \textstyle\bigwedge_{k=0}^{i}(s = a_k \rightarrow t = a_{i-k})\Big)$$

By Separation in the model theory, $(\forall x)(\{y \in x \mid Fy\}$ exists). Instantiate x with $A \times A$, using Proposition 1. This set is the required bijection, π.

For each individual constant symbol "c", each n-place predicate "R" and each n-place function symbol "f" of $\mathscr{L}$, define:

$$\begin{aligned}
\iota^{\mathcal{B}}(\underline{c}) = c^{\mathcal{B}} &:= \pi(c^{\mathcal{A}}) \\
\iota^{\mathcal{B}}(\underline{R}) = R^{\mathcal{B}} &:= \{\langle \pi(x_1), \ldots, \pi(x_n)\rangle \mid \langle x_1, \ldots, x_n\rangle \in R^{\mathcal{A}}\} \\
\iota^{\mathcal{B}}(\underline{f}) = f^{\mathcal{B}} &:= \{\langle \pi(x_1), \ldots, \pi(x_n), \pi(y)\rangle \mid \langle x_1, \ldots, x_n, y\rangle \in f^{\mathcal{A}}\}
\end{aligned}$$

If these objects all exist, then $\iota^{\mathcal{B}}$ defines a model, $\mathcal{B}$, by setting $B = A$. Moreover, $\mathcal{B} \cong \mathcal{A}$, by definition. Finally, because $\mathcal{A}$ is non-trivial, an appropriate bijection π can easily be chosen so that $\iota^{\mathcal{B}} \neq \iota^{\mathcal{A}}$ and hence $\mathcal{B} \neq \mathcal{A}$.

It remains to prove that every $c^{\mathcal{B}}, R^{\mathcal{B}}$ and $f^{\mathcal{B}}$ exists. The case of the constants is trivial. To deal with relations (the same method works for functions), given $R^{\mathcal{A}}$, define a predicate of the model theory:

$$Gy := (\exists x_1)\ldots(\exists x_n)(\exists p_1)\ldots(\exists p_n)\big(\langle p_1, \ldots, p_n\rangle = y \wedge \langle x_1, \ldots, x_n\rangle \in R^{\mathcal{A}} \wedge \textstyle\bigwedge_{k=1}^{n} F\langle x_k, p_k\rangle\big)$$

By Separation, $(\forall x)(\{y \in x \mid Gy\}$ exists). Instantiate with $A^n = \overbrace{A \times \ldots \times A}^{n \text{ times}}$, using Proposition 1, and we have $\{y \in A^n \mid Gy\} = R^{\mathcal{B}}$. □

This immediately yields:

Permutation: Theorem 3 (ZF–IP and Z–I)**.** Let T be a theory with a non-trivial model. T has multiple distinct isomorphic models. □

Note that none of this requires an axiom stating the existence of any set; the model $\mathcal{A}$ itself supplies the basic sets we need to work with. Furthermore, to reach Theorem 3, the use of Replacement/Powersets was confined to the proof of Proposition 1. So, if we want to prove the Permutation Theorem in a *really* minimal model theory, we could do away with Replacement and Powersets, and simply take Proposition 1 *itself* as an axiom. That is, we could prove the Permutation Theorem in a model theory containing just: Proposition 1; the Axiom of Extensionality; appropriate Separation instances; and whatever other resources are required to develop standard model-theoretic notions (in particular, to define "$\models$").

I leave it to the reader to determine her favourite balance of strength against naturalness. My aim is merely to show that the Permutation Theorem can be offered in model theories which are both extremely weak and natural.

Finally, note that nothing in Lemma 2 depends upon $\mathcal{A}$ being a set-model rather than a class-model (see §4.3). All that changes is the notation: if the domain of $\mathcal{A}$ is a class $\mathbf{A}$, then we define a class-sized bijection $\boldsymbol{\pi}$, and each relation or function of $\mathcal{A}$ is a class $\mathbf{R}^{\mathcal{A}}$ or $\mathbf{f}^{\mathcal{A}}$.

8.2 The Skolem Hull Theorem

In the Permutation Theorem, we start with a model and construct an isomorphic model. In the various Skolemising theorems, we shall construct elementarily equivalent models. Accordingly, we first require a criterion for when two structures are elementarily equivalent.

Tarski–Vaught: Theorem 4 (Z)**.** Let $\mathcal{A}, \mathcal{B}$ be $\mathscr{L}$-structures such that $\mathcal{A} \subseteq \mathcal{B}$. $\mathcal{A} \equiv \mathcal{B}$ **iff** for every $\mathscr{L}$-formula ϕ and every $\overline{a} \in A$, if $\mathcal{B} \vDash (\exists y)\phi(\underline{\overline{a}}, y)$, then $(\exists b \in A)\mathcal{B} \vDash \phi(\underline{\overline{a}}, \underline{b})$. □

For a proof, see Hodges (1993, pp. 48, 55). Hodges' proof is officially given in ZFC, but invokes very little background theory.

Our first Skolemising theorem will be the Skolem Hull Theorem. Many standard proofs of the Skolem Hull Theorem employ Skolem functions (e.g. Hodges 1993, pp. 88–90). I shall follow a different strategy, though, since I want to emphasise the role of Countable Dependent Choice:

Definition 5. The following principle is *Countable Dependent Choice*: If $A \neq \varnothing$ and $(\forall x \in A)(\exists y \in A)xRy$, then for any $a \in A$, there is a sequence $\langle a_n \rangle$ such $a_0 = a$ and for all $n < \omega$, $a_n R a_{n+1}$. □

In my naming-conventions for theories, taking Countable Dependent Choice as an axiom is indicated by "D_ω". Thus, ZD_ω is Zermelo set theory with the Axiom of Countable Dependent Choice. Countable Dependent Choice has some immediately useful consequences:

Proposition 6 (ZD_ω)**.** Every countable set of non-empty sets has a choice set. Moreover, every countable union of countable sets is countable. □

For a proof, see Potter (2004, pp. 161–2, 243). (A *choice set* for A is any set $\{f(a) \mid a \in A\}$, for some function f such that $f(a) \in a$ for every $a \in A$.)

My strategy for proving the Skolem Hull Theorem is now as follows (cf. Hodges 1993, pp. 87–88). Given a model $\mathcal{A}$, we start by defining a partial ordering on countable subsets of A. We then use Countable Dependent Choice to select a sequence of these subsets. The union of this sequence forms the domain for our countable Hull, $\mathcal{H}$. We treat $\mathcal{H}$ as a submodel of

$\mathcal{A}$, and then demonstrate that $\mathcal{H} \equiv \mathcal{A}$ using Theorem 4. Here is the full proof (which silently invokes Proposition 6 several times):

Skolem Hull: Theorem 7 (ZD_ω)**.** For every structure $\mathcal{A}$ and any countable set $S \subseteq A$, there is a countable structure $\mathcal{H} \prec \mathcal{A}$ with $S \subseteq H$.

Proof. Define a relation on the set $\mathbb{A} = \{X \subseteq A \mid X \text{ is countable}\}$ as follows:

$$X \lhd Y \text{ iff } X \subseteq Y \wedge \\ (\forall \phi)(\forall \overline{x} \in X)((\exists a \in A)\mathcal{A} \vDash \phi(\overline{x}, \underline{a}) \rightarrow (\exists a \in Y)\mathcal{A} \vDash \phi(\overline{x}, \underline{a}))$$

To see that $\mathbb{A}$ and $\lhd$ have the property required by Countable Dependent Choice, fix $X \in \mathbb{A}$ and define:

$$D := \{Z \subseteq A \mid (\exists \phi)(\exists \overline{x} \in X)(\forall a \in A)(a \in Z \leftrightarrow \mathcal{A} \vDash \phi(\overline{x}, \underline{a}))\}$$

As X is countable, there are only countably many finite tuples $\overline{x} \in X$. Furthermore, there are only countably many formulæ ϕ in our language. So D is countable. Let C be a choice set for D; then:

$$(\forall \phi)(\forall \overline{x} \in X)((\exists a \in A)\mathcal{A} \vDash \phi(\overline{x}, \underline{a}) \rightarrow (\exists a \in C)\mathcal{A} \vDash \phi(\overline{x}, \underline{a}))$$

Let $Y = X \cup C$; then $Y \in \mathbb{A}$ and $X \lhd Y$. So, by Countable Dependent Choice, given any countable subset $S \subseteq A$ there is a sequence starting with $S = H_0$ such that $H_n \lhd H_{n+1}$, for each $n < \omega$. That is, each H_{n+1} contains witnesses (from A) for every true (in $\mathcal{A}$) existential formula with parameters drawn from H_n. Now define $\mathcal{H}$ explicitly as a submodel of $\mathcal{A}$ as follows:

$H := \bigcup_{n<\omega} H_n$	the domain
$c^{\mathcal{H}} := c^{\mathcal{A}}$	for each constant "c"
$R^{\mathcal{H}} := R^{\mathcal{A}} \cap H^n$	for each n-place relation symbol "R"
$f^{\mathcal{H}} := f^{\mathcal{A}} \cap H^{n+1}$	for each n-place function symbol "f"

H is a countable union of countable sets, so $\mathcal{H}$ is countable.

We must check that $\mathcal{H} \subseteq \mathcal{A}$. Given $\overline{x} \in H$, there is some least n such that $\overline{x} \in H_n$. Fix a function symbol "f" of $\mathscr{L}$. Where $a \in A$ is the unique object such that $\mathcal{A} \vDash f(\overline{x}) = \underline{a}$, by construction, $a \in H_{n+1} \subseteq H$. Hence by definition of $f^{\mathcal{H}}$, $f^{\mathcal{H}}(\overline{x}) = f^{\mathcal{A}}(\overline{x}) = a$. Generalising, $\mathcal{H} \subseteq \mathcal{A}$.

Suppose now that $\mathcal{A} \vDash (\exists y)\phi(\overline{x}, y)$, for some $\overline{x} \in H$. Again, there is some least n such that $\overline{x} \in H_n$, and so some $h \in H_{n+1} \subseteq H$ such that $\mathcal{A} \vDash \phi(\overline{x}, \underline{h})$. Hence, for any formula ϕ and any sequence $\overline{x} \in H$:

$$\text{if } \mathcal{A} \vDash (\exists y)\phi(\overline{x}, y), \text{ then } (\exists h \in H)\mathcal{A} \vDash \phi(\overline{x}, \underline{h})$$

So $\mathcal{H} \equiv \mathcal{A}$, by Theorem 4. □

8.3 The Transitive Skolem Theorem

The Skolem Hull Theorem is sufficient to run a Skolemisation argument against any theory. However, if we are interested in Skolemising set theory, we can go even further, as discussed in §5. We first need some definitions:

Definition 8. $\mathcal{A}$ is a *pure-set-structure* iff the only symbol in $\mathcal{A}$'s signature is a binary relational predicate, "$\in$", which is extensional in $\mathcal{A}$; that is, $(\forall x \in A)(\forall y \in A)((\forall z \in A)(z \in^{\mathcal{A}} y \leftrightarrow z \in^{\mathcal{A}} x) \rightarrow x = y)$.

A pure-set-structure $\mathcal{A}$ is *transitive* iff both $(\forall x \in A)(\forall y \in x)y \in A$ and $(\forall x \in A)(\forall y \in A)(y \in x \leftrightarrow y \in^{\mathcal{A}} x)$.

A pure-set-structure $\mathcal{A}$ is *wellfounded* iff every subset of A has an $\in^{\mathcal{A}}$-minimal member; that is, $(\forall X \subseteq A)(\exists x \in X)(\forall y \in X)y \notin^{\mathcal{A}} x$. □

When $\mathcal{A}$ is transitive, it is easy to see that we can can replace "$\in^{\mathcal{A}}$" with "$\in$". Indeed, everything that $\mathcal{A}$ "says" about membership relations is "true"; membership-in-$\mathcal{A}$ is "really" membership. (Scare quotes are needed, since a sceptic might claim that transitivity only ensures that *according to the model theory*, membership-in-$\mathcal{A}$ is really membership.) It follows that:

Proposition 9 (Z)**.** If $\mathcal{A}$ is a transitive pure-set-structure, $\mathcal{A}$ is wellfounded.

Proof. The Axiom of Regularity (in the model theory) guarantees that there is no infinite descending $\in$-chain. Since $\mathcal{A}$ is transitive, there is no infinite descending $\in^{\mathcal{A}}$-chain either. □

Our strategy now follows McIntosh (1979, pp. 321–2; though see comments below). We start with a transitive pure-set-structure, and take its Skolem Hull, which is a countable submodel of the initial model. We then apply a fresh result, the Mostowski Collapse, which yields a countable, transitive model, as desired. So the only significant task ahead of us is to prove (a special case of) the Mostowski Collapse. This first requires the set-theoretic notion of *rank*. (For a definition of this notion which makes use of Replacement, see Kunen 1980, ch.3; for a definition which does without Replacement, see Potter 2004, ch.3.)

Definition 10 (Z)**.** The *rank* of a set is defined by recursion on $\in$:

$$\varrho(x) := \begin{cases} 0 & \text{if } x \text{ is } \in\text{-minimal} \\ \sup\{\varrho(y) + 1 \mid y \in x\} & \text{otherwise} \end{cases}$$

Since $\in$ is wellfounded (by the Axiom of Regularity), if there is any x such that $\varrho(x)$ is not defined, then there is some t such that, for all $y \in t$, $\varrho(y)$ is defined, but such that $\varrho(t)$ is not defined. This is absurd; so *rank* is a well-defined notion.

For each von Neumann ordinal α, the α^{th} level of the set-hierarchy is:

$$V_\alpha := \{x \mid \varrho(x) < \alpha\}$$

The entire set-hierarchy is then the class $\mathbf{V} := [x \mid (\exists\alpha)x \in V_\alpha]$. □

Armed with this, we prove our new result:

Mostowski Collapse (instance): Theorem 11 (Z)**.** Let $\mathcal{B}$ be a transitive pure-set-structure. For any $\mathcal{A} \subseteq \mathcal{B}$, there is a transitive pure-set-structure $\mathcal{M} \cong \mathcal{A}$.

Proof. We first define a notion of rank that is relative to a model, $\mathcal{X}$:

$$\varrho^{\mathcal{X}}(x) := \begin{cases} 0 & \text{if } x \text{ is } \in^{\mathcal{X}}\text{-minimal} \\ \sup\{\varrho^{\mathcal{X}}(y) + 1 \mid y \in^{\mathcal{X}} x\} & \text{otherwise} \end{cases}$$

Since $\mathcal{B}$ is transitive, $\mathcal{B}$ is wellfounded, by Proposition 9. Since $\mathcal{A} \subseteq \mathcal{B}$, $\mathcal{A}$ is wellfounded too. So $\varrho^{\mathcal{A}}(x)$ is well-defined for every $x \in \mathcal{A}$, by the argument given in Definition 10 (applied to $\in^{\mathcal{A}}$ instead of $\in$).

We next define a *collapse function*, Φ, such that $\varrho^{\mathcal{A}}(y) = \varrho(\Phi(y))$:

$$\Phi(x) := \begin{cases} \varnothing & \text{if } \varrho^{\mathcal{A}}(x) = 0 \\ \{\Phi(y) \mid y \in^{\mathcal{A}} x\} & \text{if } \Phi(y) \text{ is defined whenever } \varrho^{\mathcal{A}}(y) < \varrho^{\mathcal{A}}(x) \end{cases}$$

This definition seems to require Replacement, but this is only for clarity. To see this note the following, for any $y \in A$:

$$\begin{aligned} \varrho(\Phi(y)) = \varrho^{\mathcal{A}}(y) & \quad \text{by definition} \\ \leq \varrho^{\mathcal{B}}(y) & \quad \text{since } \mathcal{A} \subseteq \mathcal{B} \\ = \varrho(y) & \quad \text{since } \mathcal{B} \text{ is transitive} \\ < \varrho(A) & \quad \text{since } \in \text{ is wellfounded (in the model theory)} \end{aligned}$$

It follows we could have defined $\Phi(x)$ using Separation, instead of Replacement, with the following recursion clause:

$$\Phi(x) := \{z \in V_{\varrho(A)} \mid (\exists y \in^{\mathcal{A}} x)\Phi(y) = z\}$$

We now use Φ to define a model $\mathcal{M}$, the *Mostowski collapse* of $\mathcal{A}$:

$$\begin{aligned} M &:= \{\Phi(x) \mid x \in A\} && \mathcal{M}\text{'s domain} \\ \in^{\mathcal{M}} &:= \{\langle x, y\rangle \mid x \in y\} && \mathcal{M}\text{'s single relation} \end{aligned}$$

Again, the use of Replacement is easily eliminable. To see that $\mathcal{M}$ is transitive, suppose $y \in x \in M$. Then $x = \Phi(a) = \{\Phi(z) \mid z \in^{\mathcal{A}} a\}$, for some a. So $y = \Phi(b)$, for some $b \in^{\mathcal{A}} a$; hence $y \in M$.

To prove that $\mathcal{A} \cong \mathcal{M}$, it suffices to prove that Φ is an isomorphism. Φ is obviously a surjection; to check that it is an injection we proceed by induction on $\varrho^{\mathcal{A}}$. For induction, suppose that whenever $\max(\varrho^{\mathcal{A}}(x), \varrho^{\mathcal{A}}(y)) < \gamma$,

we have $\Phi(x) = \Phi(y) \rightarrow x = y$. Let $\max(\varrho^{\mathcal{A}}(x), \varrho^{\mathcal{A}}(y)) = \gamma$, and suppose $\Phi(x) = \Phi(y)$. For any $s \in^{\mathcal{A}} x$, $\Phi(s) \in \Phi(x) = \Phi(y) = \{\Phi(z) \mid z \in^{\mathcal{A}} y\}$. So there is some $t \in^{\mathcal{A}} y$ such that $\Phi(s) = \Phi(t)$; but $\max(\varrho^{\mathcal{A}}(s), \varrho^{\mathcal{A}}(t)) < \gamma$, so by the induction hypothesis, $s = t \in^{\mathcal{A}} y$. Similarly, if $t \in^{\mathcal{A}} y$, then $t \in^{\mathcal{A}} x$. So $x = y$, by extensionality in $\mathcal{A}$. Thus Φ is injective, and hence bijective. Moreover, Φ preserves the structure of $\in^{\mathcal{A}}$ by definition, so Φ is an isomorphism. □

The proof of this *instance* of Mostowski's Collapse follows Mostowksi's (1969, pp. 20–1) own *general* proof very closely. The only real difference is that I eliminate Replacement in this special case. The pay-off is that Transitive Skolem's Theorem does not require Replacement.

Transitive Skolem: Theorem 12 (ZD_ω)**.** For any transitive pure-set-structure $\mathcal{B}$, there is a countable transitive pure-set-structure $\mathcal{A} \equiv \mathcal{B}$.

Proof. Using Theorem 7, define a countable Skolem Hull $\mathcal{H} \equiv \mathcal{B}$. Since $\mathcal{H} \subseteq \mathcal{B}$, let $\mathcal{A}$ be the Mostowski Collapse of $\mathcal{H}$, using Theorem 11. □

To repeat myself: I have followed McIntosh's proof-strategy in proving the Transitive Skolem Theorem (1979, pp. 321–2). However, two points are worth noting. First, McIntosh is not clear that $\mathcal{A} \equiv \mathcal{B}$, so sells his result somewhat short. Second, McIntosh claims that $\mathcal{A} \subseteq \mathcal{B}$. This last claim has not been proved in *general* (so far as I can tell), but it does hold in the special instance that McIntosh considers (where $\mathcal{B}$ is to be a 'standard' model of ZF, i.e. a model of 'full' second-order ZF).

8.4 The Submodel Skolem Theorem

The proof of Theorem 12 employs the Skolem Hull Theorem. Sadly:

Proposition 13 (Z)**.** Countable Dependent Choice and the Skolem Hull Theorem are equivalent.

Proof. Theorem 7 shows that Countable Dependent Choice entails the Skolem Hull Theorem, so we now prove the opposite entailment. Let A be a set where $(\forall x \in A)(\exists y \in A)xRy$; treat this as a structure $\mathcal{A} \vDash (\forall x)(\exists y)xRy$. Fix an element $a \in A$; by the Skolem Hull Theorem, there is a countable substructure $\mathcal{H} \prec \mathcal{A}$ such that $a \in H$. As $\mathcal{H}$ is countable, we can enumerate all the elements $h_i \in H$, with $h_0 = a = a_0$. Now we define our sequence $\langle a_n \rangle$ by recursion:

$$a_{n+1} := \text{the } h_i \text{ such that } \mathcal{H} \vDash (\underline{a_n} R \underline{h_i}) \text{ and } (\forall k < i)\mathcal{H} \nvDash (\underline{a_n} R \underline{h_k}) \qquad \square$$

This proof is from Boolos & Jeffrey 1989, pp. 155–6. I repeat it here only because it has vanished from more recent editions of that book.

To obtain a Skolemising theorem that does *not* require any choice principle, we need a special (choice-free) instance of the Skolem Hull Theorem. This makes use of an idea from Hodges (1993, pp. 91–2):

Definition 14. $\mathcal{A}$ is a *Skolem-defined* structure **iff** for every $\mathscr{L}$-formula $\phi(\overline{x}, y)$ with $\overline{x}$ not empty, there is an $\mathscr{L}$-formula $\phi^*(\overline{x}, y)$ such that:

$$\mathcal{A} \vDash (\forall \overline{x})\left((\exists y)\phi(\overline{x}, y) \rightarrow ((\exists! y)\phi^*(\overline{x}, y) \wedge (\forall y)\left(\phi^*(\overline{x}, y) \rightarrow \phi(\overline{x}, y)\right))\right)$$

□

The idea is that, given a Skolem-defined structure $\mathcal{A}$, we can construct a Skolem Hull from $\mathcal{A}$ by selecting witnesses for each existential formula *explicitly*. Formally:

Skolem-Defined Hull: Theorem 15 (Z)**.** Let $\mathcal{A}$ be Skolem-defined. For any finite subset $S \subseteq A$, there is a countable structure $\mathcal{H} \prec \mathcal{A}$ with $S \subseteq H$.

Proof. Enumerate the formulæ ϕ^*, giving each an index $m < \omega$. Now define:

$$\begin{aligned} H_0 &:= S \\ H_{n+1} &:= H_n \cup \{a \in A \mid (\exists m \leq n)(\exists \overline{x} \in H_n)\mathcal{A} \vDash \phi^*_m(\underline{\overline{x}}, \underline{a})\} \\ H &:= \bigcup_{n<\omega} H_n \end{aligned}$$

Note that each H_n is finite, so H is countable. Define the model $\mathcal{H}$ exactly as in Theorem 7; we must check that $\mathcal{H} \subseteq \mathcal{A}$. Given $\overline{x} \in H$, there is some least $n < \omega$ such that $\overline{x} \in H_n$. For each function symbol "f", the formula "$f(\overline{x}) = y$", or some logical equivalent, has some place in the enumeration of formulæ; let it be ϕ^*_m. So by construction, $f^{\mathcal{A}}(\overline{x}) = a \in H_{\max(m,n)+1} \subseteq H$. So by definition, for any $\overline{x} \in H$, $f^{\mathcal{H}}(\overline{x}) = f^{\mathcal{A}}(\overline{x}) = a$. Generalising, $\mathcal{H} \subseteq \mathcal{A}$.

To check that $\mathcal{H} \equiv \mathcal{A}$, suppose $\mathcal{A} \vDash (\exists y)\phi(\underline{\overline{x}}, y)$, for some ϕ and some $\overline{x} \in H_n$. As $\mathcal{A}$ is Skolem-defined, for some $m < \omega$:

$$\mathcal{A} \vDash ((\exists! y)\phi^*_m(\underline{\overline{x}}, y) \wedge (\forall y)(\phi^*_m(\underline{\overline{x}}, y) \rightarrow \phi(\underline{\overline{x}}, y)))$$

So there is some $h \in A$ such that $\mathcal{A} \vDash \phi^*_m(\underline{\overline{x}}, \underline{h})$; by construction, $h \in H_{\max(m,n)+1} \subseteq H$. Furthermore, $\mathcal{A} \vDash \phi(\underline{\overline{x}}, \underline{h})$. Thus $\mathcal{H} \equiv \mathcal{A}$, by Theorem 4. □

The explicit construction of $\mathcal{H}$ is, in one sense, easier than the construction of a Skolem Hull with Countable Dependent Choice: we did not need to define a partial ordering, $\lhd$, on the subsets of A. However, abandoning (all forms of) choice introduces a new complexity. Suppose we had started with a countably *infinite* set $S \subseteq A$. Then each H_n would be countably infinite, and H would be a countably infinite union of countably infinite sets. Without any form of choice, we cannot prove that such a set is countable, so we would have no guarantee that our model $\mathcal{H}$ is countable.

This is why S is finite in Theorem 15. The same thought motivates the indexing of each formula ϕ^*.

Skolem-defined models are relatively rare, in general, but every *constructible* pure-set-structure is Skolem-defined. In more detail:

Definition 16 (Z)**.** Where "$D_\alpha(x)$" formalises "x is definable by some formula whose quantifiers range only over L_α":

$$\begin{aligned} L_0 &:= \varnothing \\ L_{\alpha+1} &:= \{x \subseteq L_\alpha \mid D_\alpha(x)\} \\ L_\alpha &:= \textstyle\bigcup_{\beta<\alpha} L_\beta && \text{for limit ordinals } \alpha \\ \mathbf{L} &:= [x \mid (\exists\alpha) x \in L_\alpha] \end{aligned}$$

Since $\mathbf{V} = [x \mid (\exists\alpha)x \in V_\alpha]$, "$\mathbf{V}=\mathbf{L}$" abbreviates "$(\forall x)(\exists\alpha)x \in L_\alpha$" (see Definition 10). A model is *constructible* iff it satisfies the first-order sentence "$\mathbf{V}=\mathbf{L}$". □

Kunen (1980, ch.6) offers a thorough discussion of constructibility. The following Lemma is due to Hodges (1993, p. 92):

Lemma 17. If $\mathcal{A} \vDash \mathrm{ZF} + \mathbf{V}=\mathbf{L}$, then $\mathcal{A}$ is Skolem-defined.

Proof. ZF entails that there is a definable well-ordering, $<$, of $\mathbf{L}$ (see Kunen 1980, pp. 173–4). So define:

$$\phi^*(\overline{x}, y) := \phi(\overline{x}, y) \wedge (\forall z < y)\neg\phi(\overline{x}, z)$$

Uniqueness is immediate from the fact that $<$ is a well-ordering. □

We need just one more result:

Gödel Condensation: Lemma 18 (ZF)**.** If $\mathcal{M} \vDash \mathrm{ZF} + \mathbf{V}=\mathbf{L}$ is a transitive pure-set-structure, then $M = L_\gamma$, for some ordinal γ. □

For a proof, see Kunen (1980, p. 172). Putting all this together, we arrive at the desired choice-free result:

Submodel Skolem: Theorem 19 (ZF)**.** For any transitive pure-set-structure $\mathcal{B} \vDash \mathrm{ZF}$, there is a countable transitive model $\mathcal{A} \vDash \mathrm{ZF}$ such that $\mathcal{A} \subseteq \mathcal{B}$.

Proof. Take $\mathcal{B}$'s constructible inner model, $\mathcal{C} \subseteq \mathcal{B}$ (so $C = \mathbf{L}^{\mathcal{B}}$, as it were). Since $\mathcal{C} \vDash \mathrm{ZF} + \mathbf{V}=\mathbf{L}$ (see e.g. Kunen 1980, pp. 169–70), $\mathcal{C}$ is Skolem-defined, by Lemma 17. We can therefore use Theorem 15 to form a countable Hull $\mathcal{H} \prec \mathcal{C}$, and then use Theorem 11 to generate $\mathcal{H}$'s countable Mostowski Collapse $\mathcal{A} \equiv \mathcal{H} \equiv \mathcal{C}$.

As $\mathcal{B}$ is a transitive pure-set-structure, $\mathcal{C}$ is too. So by Lemma 18, $C = L_\gamma$ for some γ. Likewise, since $\mathcal{A} \vDash \mathrm{ZF} + \mathbf{V}=\mathbf{L}$ and $\mathcal{A}$ is transitive, $A = L_\delta$ for some δ. As $\mathcal{A}$ is countable, $\delta \leq \gamma$. So $\mathcal{A} \subseteq \mathcal{C} \subseteq \mathcal{B}$. □

Potter (2004, p. 241) states this can be proved without any choice principle and without Replacement, but he does not give a proof.

The role of the Submodel Skolem Theorem in discussions of Skolem's paradox is slightly unclear. Benacerraf (1985, p. 101) and Wright (1985, p. 118) discuss the result: "Any transitive model for ZF has a transitive countable submodel" (see also Bays 2009, §2.3). As stated, this is precisely the Submodel Skolem Theorem. However, we should be slightly cautious. The result is not proved in Benacerraf's paper and, given a footnote appearing later in that paper (1985, p. 103, n. 9), we might guess that Benacerraf is relying on McIntosh's 1979-argument. And McIntosh, too, seems to claim to have proved the Submodel Skolem Theorem. However, as shown at the end of §8.3, McIntosh's argument *actually* establishes the Transitive Skolem Theorem, rather than the Submodel Skolem Theorem.

Acknowledgements

I wish to thank Timothy Bays, Michael Potter and Peter Smith for patient suggestions, advice, and comments. I particularly want to thank Gerald Sacks, who taught me model theory, with whom I had many engaging discussions, and without whom I would probably have no proof of the Submodel Skolem Theorem.

Notes

[1]Note that Bays does not think that Bays' dilemma is the only problem that Putnam's arguments face; nor, perhaps, that it is the main problem.

[2]There are various ways to add classes "harmlessly" to set theories; for an excellent philosophical and technical summary, see Potter 2004, Appendix B. Our classes are what Potter calls "virtual classes". I follow Potter's typographical distinction between classes and sets: when talking about classes, I use square-brackets rather than curly-brackets, and "ε" rather than "$\in$".

[3]This core omits details about coding and ω-models. There is some exegetical controversy over whether Putnam himself had this fallacious "proof" in mind; see Bellotti 2005, pp. 404–5 and Bays 2007, pp. 123–4.

[4]In fact, Putnam has another use for the Completeness Theorem: to argue that the metaphysical realist cannot make sense of the claim that an *ideal* theory might be *false* (1978, p. 126; 1980, pp. 472–4; 1989, p. 215). In the interests of brevity, I do not discuss this use of the Completeness Theorem directly.

[5]WKL_0 is a subsystem of second-order arithmetic which contains Weak Kőnig's Lemma as an axiom, i.e.: "every infinite subtree of the full binary tree has an infinite path". The other axioms of WKL_0 are those of RCA_0. RCA_0 contains the basic axioms of arithmetic, i.e. the existence of 0, and axioms governing $+$ and $\times$. RCA_0 also has Σ^0_1-induction (i.e. induction for any Σ^0_1-formula), and Δ^0_1-comprehension (i.e. for any Δ^0_1-formula ϕ, "$\{n \in \mathbb{N} \mid \phi(n)\}$ exists" is an axiom). See Simpson 1999, pp. 23–4, 92–3.

[6]For technical details, see Franzen 2004, pp. 172–6

[7]For technical details, see Franzen 2004, pp. 187–97.

[8]Bays 2007, pp. 126–7. I should emphasise that Bays does not commit himself to this thought; he merely suggests it as a possible response on behalf of the metaphysical

realist. Furthermore, Bays is not here considering iterated consistency sequences, but sequences of theories formed by adding increasingly large axioms of infinity of at each stage (in response to the constructivisation argument). So the response that I am here considering is an *adaptation* of a suggestion made by Bays.

[9]Bays suggested something like this to me in conversation on 7.xi.2008. I am not certain that he had exactly this in mind but, even if he did not, I wish to thank him for making me consider the idea.

[10]This is the gist of a remark made by Tarski (Skolem 1958, p. 638). Benacerraf (1985, pp. 101–4) endorses this response, as (perhaps) does Wright (1985, p. 118).

[11]Insisting that the intended interpretation is transitive would also undermine Putnam's constructivisation argument, since Putnam's constructible models are not well-founded, and so are not transitive (Putnam 1980, p. 467; Bellotti 2005, pp. 401–3).

References

Bays, Timothy (2001). 'On Putnam and His Models'. *The Journal of Philosophy*, **98**, pp. 331–50.

Bays, Timothy (2007). 'More on Putnam's Models: A Reply to Bellotti'. *Erkenntnis*, **67**, pp. 119–35.

Bays, Timothy (2009). 'Skolem's Paradox'. In Zalta, Edward N, editor: Stanford Encyclopedia of Philosophy. [http://plato.stanford.edu/entries/paradox-skolem/] – visited on 14.vi.2010.

Bellotti, Luca (2005). 'Putnam and Constructibility'. *Erkenntnis*, **62**, pp. 395–409.

Benacerraf, Paul (1985). 'Skolem and the Skeptic'. *Proceedings of the Aristotelian Society*, **59**, pp. 85–115.

Boolos, George S. & Jeffrey, Richard C. (1989). *Computability and Logic*. 3rd edition. Cambridge: Cambridge University Press.

Franzen, Torkel (2004). *Inexhaustibility: A non-exhaustive treatment*. Wellesley MA: Association for Symbolic Logic, Lecture Notes in Logic 16.

Hodges, Wilfred (1993). *Model Theory*. Cambridge: Cambridge University Press.

Kunen, Kenneth; K.J. Barwise, et al, editor (1980). *Set Theory: An Introduction to Independence Proofs*. Volume 102, Studies in Logic and the Foundations of Mathematics. Oxford: North-Holland.

Lewis, David (1984). 'Putnam's Paradox'. *Australasian Journal of Philosophy*, **62**, pp. 221–36.

McIntosh, Clifton (1979). 'Skolem's Criticisms of Set Theory'. *Noûs*, **13**, pp. 313–34.

Mostowski, A. (1969). *Constructible Sets with Applications*. Amsterdam: North-Holland, Studies in Logic and the Foundations of Mathematics.

Potter, Michael (2004). *Set Theory and its Philosophy*. Oxford: Oxford University Press.

Putnam, Hilary (1978). *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.

Putnam, Hilary (1980). 'Models and Reality'. *Journal of Symbolic Logic*, **45**, pp. 464–82.

Putnam, Hilary (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

Putnam, Hilary (1983). *Realism and Reason: Philosophical Papers*. Volume 3, Cambridge: Cambridge University Press.

Putnam, Hilary (1989). 'Model Theory and the 'Factuality' of Semantics'. In George, Alexander, editor: Reflections on Chomsky. Oxford: Basil Blackwell, pp. 213–232.

Putnam, Hilary (1992). 'Replies'. *Philosophical Topics*, **20.1**, pp. 347–408.

Simpson, Stephen G. (1999). *Subsystems of Second-Order Arithmetic*. Berlin: Springer-Verlag.

Skolem, Thoralf (1922). 'Some Remarks on Axiomatised Set Theory'. In Heijenoort, J van, editor: From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931 (1967). Cambridge MA: Harvard University Press, pp. 290–301.

Skolem, Thoralf (1958). 'Une Relativisation des Notions Mathématiques Fondamentales'. In Fenstad, E. J., editor: Selected Works in Logic (1970). Oslo: Universitetsforlaget, pp. 633–8.

Velleman, Daniel J. (1998). 'Review of Levin's 'Putnam on reference and constructible sets' (1997)'. *Mathematical Reviews*, **98c:03015**, p. 1364.

Wright, Crispin (1985). 'Skolem and the Skeptic'. *Proceedings of the Aristotelian Society*, **59**, pp. 117–137.