Erkenntnis
https://doi.org/10.1007/s10670-020-00340-2

ORIGINAL RESEARCH

On Putnam's Proof of the Impossibility of a Nominalistic Physics

Thomas William Barrett[1]

Received: 6 January 2020 / Accepted: 26 October 2020

Abstract
In his book *Philosophy of Logic*, Putnam (1971) presents a short argument which reads like—and indeed, can be reconstructed as—a formal proof that a nominalistic physics is impossible. The aim of this paper is to examine Putnam's proof and show that it is not compelling. The precise way in which the proof fails yields insight into the relation that a nominalistic physics should bear to standard physics and into Putnam's indispensability argument.

1 Introduction

The indispensability argument aims to establish the existence of mathematical entities by appealing to the fact that mathematics plays a crucial role in our best scientific theories. It has been called the only "non-question-begging" argument for mathematical realism (Field 2016, p. 4), and it has been one of the most discussed arguments in philosophy of mathematics for the last half century.

The argument is often put as follows (Colyvan 2001, 2019):

- We ought to be ontologically committed to all and only the entities that are indispensable to our best scientific theories.
- Mathematical entities are indispensable to our best scientific theories.

∴ We ought to be ontologically committed to mathematical entities.

Despite the significant attention that the indispensability argument has received, one rarely sees strong arguments put forward in support of the second premise. This is perhaps because the received view about the indispensability argument is that the first premise is the one "that is most obviously in need of support" (Colyvan 2019). Justification for the second premise is most often limited to pointing out that our best scientific theories *do* appeal to mathematical entities in their current formulation and

✉ Thomas William Barrett
thomaswbarrett@ucsb.edu

[1] Department of Philosophy, UC Santa Barbara, Santa Barbara, CA, USA

Published online: 24 November 2020

that it is hard to imagine how one could formulate these theories without appealing to such entities. Colyvan (2001) himself remarks that if mathematical entities were successfully dispensed with, then he would give up on his ontological commitment to them. This same attitude toward the second premise is suggested by Quine, who famously described himself as only a 'reluctant platonist'.

In his short book *Philosophy of Logic*, Putnam (1971) provides a positive argument for the second premise of the indispensability argument which has so far been overlooked in the literature on indispensability. Putnam's argument appears in the chapter entitled "The Inadequacy of Nominalistic Language". It reads like a formal proof that it is "impossible to 'do' physics in nominalistic language", or in other words, that physics cannot be done in such a way that does "not presuppose the existence of such entities as classes or numbers" (Putnam (1971), p. 35). In what follows, we will call the argument *Putnam's proof*. If Putnam's proof is sound, then it would conclusively establish the second premise of the indispensability argument. A nominalistic physics would be impossible, and so mathematical entities would be indispensable to our physical theories.

The aim of this paper is to examine Putnam's proof in detail. One begins with the suspicion that the proof is flawed in some way. If it were sound, then any attempt to nominalize physics would be a non-starter since the proof would imply that "a nominalistic language is *in principle* inadequate for physics" (Putnam (1971), p. 39). But in the decades since Putnam put forward this argument, there have been a number of attempts made to formulate a nominalistic physics.[1] And although there is room for debate about how successful these attempts have been, they are certainly not non-starters. As we will see, one's initial suspicion is correct: Putnam's proof is not sound.

A careful examination of the proof, however, does more than simply demonstrate its shortcomings. There are two additional reasons why this discussion of Putnam's proof is of broad philosophical interest. First, the discussion yields insight into the precise relationship that a nominalistic physics—or, for that matter, any theory that aims for economy by 'dispensing with' something—must bear to the theory that we began with. In brief, Putnam requires the two theories to bear *too close* a relation to one another, and this is what makes his proof unsound.

Second, there is also a historical reason that this examination of Putnam's proof is of interest. Although the indispensability argument as presented above is often attributed to Quine and Putnam, recent work has shown that the argument Putnam actually put forward was quite different. Indeed, Putnam intended a different conclusion to his indispensability argument than the one above; he wanted to demonstrate the objectivity of mathematics, rather than the existence of mathematical objects. This has already been remarked upon by Bueno (2018), Liggins (2008), and Putnam (2012) himself. But consideration of Putnam's proof illuminates another difference between Putnam's indispensability argument and the received one. The *kind* of indispensability that Putnam had in mind is different from the kind of

[1] The work of Field (2016) is, of course, the most famous. See also Balaguer (1996) and Arntzenius and Dorr (2011).

indispensability that has been adopted in other discussions of the indispensability argument. In addition, Putnam's proof appeals to the concept of *equivalent theories*. This concept links his indispensability argument closely to other parts of his philosophy, in particular his later famous arguments for conceptual relativity. One comes away from this proof with a better appreciation of how important the topic of equivalent theories was for Putnam.

2 Putnam's Proof

We begin with a reconstruction of Putnam's proof (Putnam (1971), pp. 37–39). Putnam considers our standard physics, whose formulation appeals to various abstract mathematical entities like numbers, functions, and sets. We will call this theory T_p. It is formulated in a language L_p, which Putnam calls the "language of physics", that is rich enough to formulate infinitely many sentences of the form "the mass of object A is r" where r is an arbitrary rational number.[2] According to the theory T_p, for different choices of r these statements are not equivalent to one another. They are, after all, saying different things about what the mass of A is. This observation yields the first premise of Putnam's proof.

P1. There are infinitely many sentences $\phi_1, \phi_2, \ldots$ in L_p that are not pairwise equivalent according to T_p, i.e. for any $i \neq j$ it is not the case that $T_p \vDash \phi_i \leftrightarrow \phi_j$.

One can think of the sentence ϕ_i as the statement "the mass of object A is i".

If it were possible, a nominalistic physics would be a theory T_n that is formulated in a nominalistic language L_n. Putnam (1971, p. 35) describes roughly what L_n must be like as follows:

> By a "nominalistic language" is meant a formalized language whose variables range over individual things, in some suitable sense, and whose predicate letters stand for adjectives and verbs applied to individual things (such as "hard", "bigger" than, "part of"). These adjectives and verbs need not correspond to observable properties and relations; e.g., the predicate "is an electron" is perfectly admissible, but they must not presuppose the existence of such entities as classes or numbers.

Putnam makes some further assumptions about what T_n and L_n would have to be like in order to count as suitably nominalistic. First, the nominalist physics must say that "there are only finitely many individuals" (Putnam (1971), p. 38). This provides the second premise in Putnam's proof.

[2] The assumption that r is rational has no real bearing on the proof, other than the fact that it allows for infinitely many distinct statements of the required form. Putnam does not allow r to be an arbitrary real number because he worries that we cannot have names for uncountably many objects.

P2. A nominalistic physics says that there are finitely many things, i.e. $T_n \vDash \exists_{=N}$ for some natural number N.

We use the notation $\exists_{=N}$ as an abbreviation for the sentence in L_n that says "there are exactly N things".

Putnam's idea here is the following. Since the nominalist thinks that there are no abstract objects like numbers, functions, and sets, Putnam claims that they must think that only physical objects exist. Indeed, when discussing the nominalist's position earlier in the book he says as much:

> Goodman denies that nominalism is a restriction to "physical" entities. However, while the view that only physical entities (or "mental particulars", in an idealistic version of nominalism; or mental particulars and physical things in a dualistic system) alone are real may not be what Goodman intends to defend, it is the view that most people understand by "nominalism", and there seems little motive for being a nominalist apart from some such view (Putnam (1971), pp. 15–16).

Since there are only finitely many physical objects, Putnam claims that a nominalistic physics will say that there are only N things, for some specific finite number N.[3]

Putnam also assumes that "the number of primitive predicates in the language is finite" (Putnam (1971), p. 38). So we have the third premise.

P3. The language L_n of the nominalistic physics T_n is finite.[4]

Putnam does not give an argument for P3, and in fact, he only mentions this premise in a footnote. But we can reconstruct what he might have been thinking. A nominalistic language cannot contain, for example, names for numbers or other abstract objects. So Putnam is supposing that L_n contains only names for and predicates that apply to concrete entities, of which there are plausibly only finitely many.

From the premises P1, P2, and P3, Putnam concludes that no "translation scheme" (Putnam (1971), p. 38) from the standard physics T_p to the nominalistic physics T_n can exist. He does not make precise exactly what notion of translation he is using here. But there is a standard kind of translation that is well known to logicians and does the work that Putnam wants it to do. It will take a moment to introduce this concept.

We begin with the idea of a reconstrual between languages L_1 and L_2. A **reconstrual** F of L_1 into L_2 is a map from elements of the language L_1 to L_2-formulas that satisfies the following condition.

[3] Putnam expresses a similar idea about nominalism later, writing that "[n]ominalists must at heart be materialists, or so it seems to me: otherwise their scruples are unintelligible" (Putnam (1971), p. 36).

[4] Note that we are following standard model theoretic practice and thinking of the languages L_n and L_p as just containing the basic predicates of the language. Infinitely many sentences can be formulated in L_n, although, as we will shortly see, many of them are logically equivalent to one another.

- For every n-ary predicate symbol $p \in L_1$, $Fp(x_1, \dots, x_n)$ is a L_2-formula with n free variables.

We assume here for simplicity that the languages L_1 and L_2 only contain predicate symbols.[5] The important fact about a reconstrual $F : L_1 \to L_2$ is that it naturally induces a map from arbitrary L_1-formulas to L_2-formulas. The map is defined in the standard recursive manner; one simply requires that F 'respects' the logical connectives. So, for example, F maps the L_1-sentence $\phi_1 \wedge \phi_2$ to the L_2-sentence $F\phi_1 \wedge F\phi_2$, the L_1-sentence $\forall x\phi(x)$ to the L_2-sentence $\forall xF\phi(x)$, and so on.

If T_1 and T_2 are theories in the languages L_1 and L_2, then a **translation** $F : T_1 \to T_2$ is a reconstrual $F : L_1 \to L_2$ such that $T_1 \vDash \phi$ implies that $T_2 \vDash F\phi$ for every L_1-sentence ϕ. Translations in this sense map theorems to theorems and they preserve the logical relations between sentences. We say that a translation $F : T_1 \to T_2$ is **conservative** if $T_2 \vDash F\phi$ implies that $T_1 \vDash \phi$ for every Σ_1-sentence ϕ. There is a strong sense in which a conservative translation $F : T_1 \to T_2$ preserves all of the logical relations between sentences of T_1. For example, with a conservative translation T_1 entails that one sentence implies another if and only if T_2 entails that the translation of the one implies the translation of the other. It is reasonable to assume that this is the precise notion of translation that Putnam has in mind, and that he means to draw the following conclusion from premises P1, P2, and P3.

C1. There is no conservative translation F from T_p to T_n.

There are three reasons why it is reasonable to believe that Putnam had this precise notion of translation in mind. First, in a paper published a few years after *Philosophy of Logic*, Putnam (1974, p. 29) gestures at a technical notion of translation, which he calls an "analytical hypothesis", that is very close to the standard notion of translation described here. Second, Putnam explicitly says that any translation between T_p and T_n "must disrupt logical relations" (Putnam (1971), p. 39), in the precise sense of mapping a "false 'theorem'" to a 'true theorem'. Conservative translations are precisely those translations that *preserve* logical relations and do not map 'non-theorems' to theorems, so we therefore have good reason to believe that this is the exact kind of translation whose existence Putnam is aiming to rule out.

The third reason is based on a principle of charity: It is indeed the case that C1 follows from P1, P2, and P3. This inference relies on the following Lemma, which Putnam proves in a footnote (Putnam (1971), fn. 3).

Lemma *If P2 and P3, then there is a finite collection of sentences $\psi_1, \dots, \psi_m$ in L_n such that for any sentence ψ in L_n, $T_n \vDash \psi \leftrightarrow \psi_i$ for some i.*

The Lemma is intuitive. Since the nominalistic physics T_n is formulated in a finite language and says that there are only finitely many things, it is only able to express

[5] If the signatures contain function or constant symbols, there are two more conditions that F must satisfy. For further details on reconstruals and translations see Barrett and Halvorson (2016a).

finitely many different statements. Now with the Lemma in hand we have the following straightforward result.

Proposition 1 *If* $P1$, $P2$, *and* $P3$, *then* $C1$.

Proof Suppose for contradiction that there is a conservative translation $F : T_p \rightarrow T_n$. This means that each of the pairwise non-equivalent L_p-sentences $\phi_1, \phi_2, \ldots$ (whose existence is guaranteed by P1) are translated into L_n-sentences $F\phi_1, F\phi_2, \ldots$. Now P2 and P3 imply, via the Lemma, that for each i

$$T_n \vDash F\phi_i \leftrightarrow \psi_j$$

for some j. Since there are infinitely many of the $F\phi_i$ and only finitely many of the ψ_j, it must be that there are ϕ_l and ϕ_k such that $l \neq k$, but $F\phi_l$ and $F\phi_k$ are equivalent to the same ψ_j. This implies that $T_n \vDash F\phi_l \leftrightarrow F\phi_k$. Since F is a reconstrual, this is just saying that $T_n \vDash F(\phi_l \leftrightarrow \phi_k)$. Since F is conservative, $T_p \vDash \phi_l \leftrightarrow \phi_k$, which contradicts P1. □

Putnam's argument from P1, P2, and P3 to C1 is therefore valid. So he has shown that it is not possible to provide a nominalistic physics T_n that satisfies P2 and P3 such that there is a conservative translation $F : T_p \rightarrow T_n$. From this he draws one last conclusion. He writes that "any 'translation' of 'the language of physics' into 'nominalistic language' must disrupt logical relations [...] Thus a nominalistic language is in principle inadequate for physics" (Putnam (1971), p. 39). He is here drawing the following conclusion from C1.

C2. There is no nominalistic physics T_n.

Unfortunately, there is a logical gap between C1 and C2. The fact that there is no conservative translation from T_p to T_n does not immediately imply that a nominalistic physics T_n does not exist. Putnam claims that the nominalist "wishes to [...] find a 'translation function'" from the standard physics T_p into their nominalistic physics T_n (Putnam (1971), p. 19). But it is not immediately clear why the nominalist wishes to do this.

The bottom line is this: Putnam must be implicitly committed to some additional premise which, in conjunction with C1, entails C2. There is a particularly natural way to fill this logical gap. Putnam's basic idea here is that a nominalistic physics T_n must be *equivalent* to the standard physics T_p. Equivalent theories 'say the same thing' or 'have the same content', despite perhaps being formulated using different language or formal apparatus. The most famous examples of equivalent theories in physics are the Hamiltonian and Lagragian formulations of classical mechanics and the Heisenberg and Schrödinger formulations of quantum mechanics.[6] If

[6] The former has recently been the subject of significant debate. See for example North (2009), Curiel (2014), and Barrett (2015, 2017).

two theories are equivalent, then there should be a way to translate back and forth between them. That suggests that if one requires T_n to be equivalent to T_p, then C1 will imply C2.

This gives us the following premise.

P4. A nominalistic physics T_n is equivalent to the standard physics T_p.

The idea behind P4 should be clear. Insofar as the nominalist wants to be able to *do physics* without appealing to abstract objects like numbers, sets, and functions, the nominalistic physics T_n that they formulate must bear a close relationship to the standard physics T_p. Intuitively, T_n must 'do the same work as' T_p. It cannot be a completely different theory that is unrelated to T_p. Otherwise the nominalist will merely have shown that *something* can be done without abstract objects, but they will not have shown that *physics* can be done without abstract objects. The idea behind P4 is that in order for T_n to 'do the same work as' T_p, T_n and T_p must be equivalent.

Although he does not explicitly commit to P4 in *Philosophy of Logic*, there is evidence that Putnam had it in mind when putting forward his proof. The first piece of evidence is again based on a principle of charity. We will see shortly that P4 makes Putnam's argument valid, as was certainly his intention. There is also a piece of direct textual evidence in *Philosophy of Logic* that suggests Putnam is committed to P4. At the outset of his proof he remarks that "Newton's law [of universal gravitation] has a content which [...] quite transcends what can be expressed in nominalistic language" (Putnam (1971), p. 37). This suggests that he is trying to show that it is impossible to capture the *same content* as T_p in a nominalistic theory T_n, and that is tantamount to endorsing the premise P4. Moreover, the topic of equivalent theories was a familiar one to Putnam. In both earlier and later work, Putnam makes it clear that he thinks that equivalence is a "profoundly significant" topic (Putnam 1983, p. 45).[7] And indeed, he mentions "equivalent descriptions" in the final chapter of *Philosophy of Logic* as one of the topics that he would have discussed if he had the space. This suggests that equivalence was on his mind while writing *Philosophy of Logic*. If he is committed to P4, that would explain why: Putnam is trying to demonstrate that we cannot reformulate standard physics in such a way that its content is preserved—i.e. so that the reformulation 'says the same thing' as standard physics—without appealing to abstract objects like numbers.

One can find another piece of evidence that Putnam (1971) is committed to P4 by looking back to his 1967 paper "Mathematics Without Foundations". There Putnam (1967) discusses how mathematics can be dispensed with using modal logic, i.e. how "mathematical proposition[s] can be treated as a statement[s] involving modalities, but not special objects" (Putnam (1967), p. 11).[8] After he shows how

[7] Putnam (1983) is his most detailed discussion of equivalence, but the topic comes up in a number of other papers as well.

[8] He did not call this theory nominalistic because it included appeal to modalities, which he believed the nominalist must reject. Furthermore, he does not discuss whether "mixed predicates" like "the mass of object A is r" might be dispensable, which is his main concern in Putnam (1971). See (Burgess and

mathematical entities can be dispensed with using modalities, he calls his resulting modal picture an "equivalent description" to the standard picture of mathematics that involves "special objects" (Putnam (1967), p. 9). This suggests that his understanding of the nominalist's endeavor is the following. The nominalist is trying to provide a theory T_n that is equivalent to the standard theory T_p, but which does not appeal to abstract objects like numbers, functions, and sets. Just as Putnam (1967) was aiming for an equivalent description of classical mathematics that does not employ abstract objects, Putnam (1971) assumes that the nominalist is aiming for an equivalent description of standard physics that does not employ abstract objects. In understanding the nominalist's aim this way, he is committing to P4.

But in order to make the inference from C1 to C2 valid, one has to be clear about exactly what *kind* of equivalence is at play in P4. There are a number of different standards of equivalence that have been proposed over the years.[9] It seems that the one that best fits Putnam's purpose is definitional equivalence. Definitional equivalence is a particularly well known standard of equivalence that was being discussed by both logicians and philosophers of science at the time that Putnam was writing *Philosophy of Logic*. Glymour (1971, 1977, 1980) was the first to apply definitional equivalence to cases of real interest in physics, but the concept was already being discussed by logicians in the 1960s. Artigue et al. (1978) and de Bouvére (1965) attribute the concept to Montague (1957). And it was certainly familiar by the late 1960s, as is evident through the work of de Bouvére (1965), Shoenfield (1967), and Kanger (1968).

Putnam himself comes close to endorsing definitional equivalence in "Mathematics Without Foundations", though he does not call it by that name. There he describes equivalent theories as follows:

> [T]he primitive terms of each [theory] admit of definition by means of the primitive terms of the other theory, and then each theory is a deductive consequence of the other. (Putnam 1967, p. 8)

This reads like an informal description of definitional equivalence. All of this evidence together suggests that Putnam was implicitly committed to P4 *and* thought that definitional equivalence (or something close to it) was a necessary condition in order for two theories to be equivalent.[10] From here on, therefore, we will assume

Footnote 8 (continued)
Rosen (1997), III.B.2.d) for a brief discussion of how to square Putnam (1967) with Putnam (1971), and Burgess (2018) for a critical discussion of Putnam (1967).

[9] See Weatherall (2019) for a survey of recent work.

[10] The situation in (Putnam 1983) is slightly more complicated. Putnam (1983, p. 40) suggests that "one expects some type of translation" to exist between two theories if they are to be considered equivalent. This is the case, as we will shortly discuss, with definitional equivalence, but Putnam here explicitly mentions a standard of equivalence called mutual relative interpretability, which is weaker than definitional equivalence (Barrett and Halvorson 2019). He claims that mutual relative interpretability plus the "informal requirement" that the interpretations preserve explanations will suffice for full equivalence of theories. The problem with this idea is that mutual relative interpretability is known to be a poor formal standard of equivalence, in that it considers *too many* theories to be equivalent; for an example see Barrett and Halvorson (2019). One therefore suspects that he would be happy instead endorsing definitional

that T_n and T_p must be definitionally equivalent in order for P4 to be true. This means that if C1 implies that T_n and T_p *cannot* be definitionally equivalent, then C2 immediately follows.

That is, in fact, exactly how P4 bridges the logical gap from C1 to C2. In order to show this we need to describe definitional equivalence precisely. We first need to formalize the notion of a definition. Let $L \subset L^+$ be languages and let $p \in L^+ - L$ be an n-ary predicate symbol. An **explicit definition of** p **in terms of** L is a L^+-sentence of the form

$$\forall x_1 \dots \forall x_n \big(p(x_1, \dots, x_n) \leftrightarrow \phi(x_1, \dots, x_n)\big)$$

where $\phi(x_1, \dots, x_n)$ is a L-formula. One describes explicit definitions of function and constant symbols in a similar manner (Barrett and Halvorson 2016a). A **definitional extension** of an L-theory T to the language L^+ is a L^+-theory

$$T^+ = T \cup \{\delta_s : s \in L^+ - L\},$$

where for each symbol $s \in L^+ - L$ the sentence δ_s is an explicit definition of s in terms of L.

Definition Let T_1 be an L_1-theory and T_2 be an L_2-theory. T_1 and T_2 are **definitionally equivalent** if

- there is a definitional extension T_1^+ of T_1 to the language $L_1 \cup L_2$
- and a definitional extension T_2^+ of T_2 to the language $L_1 \cup L_2$
- such that T_1^+ and T_2^+ are logically equivalent, i.e. they have the same class of models.

One often says that T_1 and T_2 are definitionally equivalent if they have a 'common definitional extension.' With P4 in hand Putnam's proof is essentially complete. All that remains is the following simple argument.

Proposition 2 *If* C1 *and* P4, *then* C2.

Proof Suppose for contradiction that a nominalistic physics T_n exists. It follows from P4 that a conservative translation $F : T_p \to T_n$ must exist. The argument for this claim is simple. Suppose that T_p and T_n are definitionally equivalent as P4 requires. One then defines a reconstrual $F : L_p \to L_n$ in the following natural manner. For each predicate symbol $p \in L_p$, we let Fp be the L_n-formula ϕ that T_n uses to define p. Using the fact that T_p and T_n are definitionally equivalent, one can easily verify that this reconstrual is indeed a translation $F : T_p \to T_n$, and moreover, that it is

Footnote 10 (continued)
equivalence, which is a much more reasonable standard, as the formal requirement. In either case, the same argument goes through to establish C2 from C1 and P4.

conservative. (See Proposition 4.5.26 and Theorem 4.6.17 of Halvorson (2019) for a full proof of this claim.) This directly contradicts C1, so the theory T_n does not exist. □

In full, therefore, Putnam's proof of the impossibility of a nominalistic physics runs as follows.

P1. There are infinitely many sentences $\phi_1, \phi_2, \ldots$ in L_p that are not pairwise equivalent according to T_p, i.e. for any $i \neq j$ it is not the case that $T_p \vDash \phi_i \leftrightarrow \phi_j$.
P2. A nominalistic physics says that there are finitely many things, i.e. $T_n \vDash \exists_{=N}$ for some natural number N.
P3. The language L_n of the nominalistic physics T_n is finite.
∴ C1. There is no conservative translation F from T_p to T_n.
P4. A nominalistic physics T_n is equivalent to the standard physics T_p.
∴ C2. There is no nominalistic physics T_n.

P1 is an intuitive claim about the content of T_p. P2 and P3 are claims about what T_n would have to be like for it to count as nominalistic. And P4 is a claim about what the relationship between T_p and T_n must be. Insofar as one understands the notion of translation in the way defined above and takes definitional equivalence to be a necessary condition on equivalence of theories, Propositions 1 and 2 imply that Putnam's proof is valid.

3 The Structure of Putnam's Proof

Our task now is to evaluate whether or not Putnam's proof is sound. Since it is valid, he has ruled out some kinds of nominalistic physics. The question is then whether any of the kinds of theories that are ruled out by the proof are the kinds of theories that nominalists should be aiming for. I will argue that, especially for a particular kind of nominalist, they are not. Nominalists who formulate their nominalistic physics with the aim of genuinely dispensing with mathematical entities—or, for that matter, anyone proposing a theory that aims to dispense with something—should not want their theory to be equivalent to the theory that they began with. Indeed, insofar as they want to dispense with something from T_p nominalists should actually aim for T_n to be *inequivalent* to T_p and, moreover, for there to be no translation from T_p to T_n. Our main attention will be devoted to showing that P4 is false in this sense, but in this section we will begin by briefly discussing the other premises.

A consideration of the other premises is merited because there is a kind of nominalist who might aim to formulate a nominalistic physics T_n that *is* definitionally equivalent to T_p. Such a nominalist would not propose T_n with the aim of genuinely dispensing with abstract objects. Rather, they would aim for T_n to 'redescribe' or 'reinterpret' the abstract objects of T_p in concrete terms. They might then be content with T_n being definitionally equivalent to T_p. Since P1 imposes a fairly weak constraint on T_p and is hard to resist—it is the case that one can say infinitely many different things about the mass of some object in our standard physics, and that seems to immediately imply P1—Putnam's proof entails that this kind of nominalist needs to deny P2 or P3 in order to fashion themselves with a sufficiently rich ideology to accomplish their aim.[11] Fortunately, P2 and P3 are less difficult to resist than P1, and as we will show in what follows Putnam's proof depends critically on them. If either of these premises are dropped, the argument to C2 is no longer valid.

We first consider P2. Depending on one's precise brand of nominalism, T_n need not entail that there are exactly N things. The theory T_n may no longer appeal to mathematical entities while nonetheless entailing that there are infinitely many things. A particularly famous example of a theory like this is the nominalistic theory of Newtonian gravitation proposed by Field (2016), which does not use numbers in its formulation but does quantify over an infinity of spacetime points.[12] For our purposes here, it will suffice to note that rejecting P2 allows one to resist Putnam's argument. Indeed, if one rejects P2, and has only P1, P3, and P4 in hand, then C2 does not follow. The following proposition illustrates this.

Proposition 3 *P*1, *P*3, *and P*4 *do not imply C*2.

Proof Let both languages L_n and L_p be the empty language, i.e. the language that has no non-logical vocabulary. Consider the following two theories.

$$T_n = \emptyset \qquad\qquad T_p = \emptyset$$

That is, both T_n and T_p have no axioms. It is easy to verify that P1, P3, and P4 hold of this pair of theories. Note first that the sentences $\exists_{=1}, \exists_{=2}, \exists_{=3}$, etc. in the language L_p are pairwise inequivalent according to T_p. This means that P1 holds. P3 holds trivially since L_n is empty. It is also trivial that T_p and T_n are definitionally equivalent (indeed, they are logically equivalent), so P4 holds too. And we have shown that C2 is false by exhibiting the theory T_n. □

One can also resist P3. Indeed, since Putnam gives no direct argument for this premise—he only mentions it in a footnote—this may be an easy way to avoid

[11] Note that a nominalist might instead respond to Putnam's proof by denying that the kind of translation that they are aiming for is the kind defined above. If they are aiming for a sufficiently weak kind of translation from T_p to T_n, then Putnam's proof may no longer go through. But the burden is then on them to make precise the kind of weak translation that they desire. The nominalist who does aim for T_n to be equivalent to T_p must also find a way to argue that T_n does not commit to abstract objects, despite the fact that it is equivalent to a theory T_p that *does* appear to make that commitment.

[12] This has been a contentious feature of Field's theory. The issue is whether or not Field should count as a 'genuine' nominalist given his willingness to quantify over spacetime points. It has been argued that these objects are not sufficiently concrete for the nominalist to admit them into their physics. But this debate is beyond the scope of this paper. P2 would also be violated if every model of T_n were finite, but some had a different finite number of elements than others. One suspects that Putnam's proof would still go through with a weakened version of P2 that stated only that T_n has only finite models. Even with a weakened P2, however, it is not clear that nominalists would be obligated to accept it, and Field's theory certainly would not satisfy it.

Putnam's conclusion. The argument for P3 that we presented earlier was that L_n cannot contain infinitely many names for numbers (or other abstract objects like sets or functions), since if it did it would not be a nominalistic language. But there may be some other way that L_n becomes infinite. There might, for example, be infinitely many properties that the finitely many concrete objects can have. In this case, L_n would contain infinitely many predicate symbols. For our purposes here it is most important to merely note that—as was the case with P2—Putnam's proof is not valid if we give up P3. The following proposition illustrates this.

Proposition 4 *P*1, *P*2, *and P*4 *do not imply C*2.

Proof Let the languages L_n and L_p both consist of a countable infinite of unary predicate symbols $p_1, p_2, p_3, \ldots$. Consider the following two theories.

$$T_n = \{\exists_{=1}\} \qquad\qquad T_p = \{\exists_{=1}\}$$

That is, both T_n and T_p say merely that there is one thing, and both are silent as to whether that one thing has any of the properties p_i. It is easy to verify that P1, P2, and P4 hold of this pair of theories. Any two of the sentences $\forall x p_1(x), \forall x p_2(x), \forall x p_3(x), \ldots$ in the language L_p are inequivalent according to T_p, so P1 holds. P3 trivially holds and P4 does too, since the two theories are once again identical. And again we have shown that C2 is false by exhibiting the theory T_n. □

Propositions 3 and 4 shed light on the structure of Putnam's proof. It depends crucially on premises P2 and P3. Rejecting either one is a viable option for resisting the conclusion C2, particularly since both impose quite strong constraints on T_n. But both of these routes require one to consider the details of one's proposed nominalistic physics. This is a conceptual difference between P2 and P3, on the one hand, and P4 on the other. P2 and P3 are claims about exactly what the details of one's nominalistic physics T_n are. P4, on the other hand, is best understood as a claim about what one should be aiming for when one presents a nominalistic physics. Even if one sees good reason to reject P2 or P3, therefore, one should still be interested in whether or not P4 is true. We would like to know what the exact relationship should be between a theory that purports to 'dispense with' something, like a nominalistic physics does, and the theory that we began with. By committing to P4 Putnam is claiming that nominalists should be aiming for their nominalistic physics to have the same content as—i.e. be equivalent to, in the precise sense given by definitional equivalence—our standard physics. There is good reason, however, to disagree with this claim.

4 Euclidean Geometry

My aim in this section is to argue that P4 is, in general, incorrect by considering the case of two dimensional Euclidean geometry. Nominalists often cite this example as providing an illustration of exactly the kind of thing that they are trying to do. Euclidean geometry is sometimes formulated using numbers. These numbers are used to lay down coordinates on Euclidean space, so that each point in the space can be represented by a pair of real numbers. One then uses these coordinates to define various geometrical notions like distance between points. This way of formulating Euclidean geometry is called *analytic geometry*. But Euclidean geometry can also be formulated in a *synthetic* manner which does not employ the apparatus of numbers. Instead of using numbers to lay down coordinates, these synthetic formulations appeal directly to various relations, like 'betweenness' and 'congruence', that might hold between points in the space.

Synthetic Euclidean geometry is often thought of as a canonical example of a theory that dispenses with numbers (Field 2016). The aim of this section is to show that according to most reasonable standards of equivalence—including definitional equivalence—analytic and synthetic geometry are not equivalent theories. So if nominalists provide a theory that bears the same relationship to standard physics as synthetic geometry bears to analytic geometry, then the theories they put forward will not be equivalent to standard physics. They will therefore be denying P4.

In order to discuss this case carefully, we first need to state these two formulations of Euclidean geometry and record a small collection of facts about them. We begin with a formulation of analytic geometry, which we will call *AG*. The theory *AG* is formulated in a two-sorted language with a sort σ_p of points and a sort σ_n of numbers.[13] We will use letters $p, q, r,$ etc. to denote variables of sort σ_p and letters from the end of the alphabet like x, y, z to denote variables of sort σ_n. In addition to these two sort symbols, the language of *AG* contains the following predicate, function, and constant symbols.

- The relation symbol $\leq$ of arity $\sigma_n \times \sigma_n$, function symbols $+$ and $\cdot$ of arity $\sigma_n \times \sigma_n \to \sigma_n$, and constant symbols 0 and 1 of arity σ_n are all understood in the standard way: $\leq$ is the 'less than or equal to' relation on numbers, $+$ and $\cdot$ are the operations of addition and multiplication on numbers, and 0 and 1 are the additive and multiplicative identities.
- The 'betweenness' relation $b(p, q, r)$ of arity $\sigma_p \times \sigma_p \times \sigma_p$ indicates that the point q lies on the line between the points p and r, and the 'congruence' relation $c(p, q, r, s)$ of arity $\sigma_p \times \sigma_p \times \sigma_p \times \sigma_p$ indicates that the point p is as distant from q as r is from s.
- The function symbols $-_1$ and $-_2$ of arity $\sigma_p \to \sigma_n$ are the coordinate functions that assign to a point p its first coordinate p_1 and its second coordinate p_2.

The theory *AG* is then given by the following axioms.

- The elements of sort σ_n form a real-closed field with ordering $\leq$, operations $+$ and $\cdot$, and additive and multiplicative identities 0 and 1. For a precise statement of this collection of axioms, see Hodges (2008, p. 38) or Tarski (1959, p. 20).

[13] For details on many-sorted logic, see Barrett and Halvorson (2016b). See Burgess and Rosen (1997) and Tarski (1959) for further details on the theory *AG*, but note that neither calls it by that name.

- For any numbers x and y, there is a unique point p such that $p_1 = x$ and $p_2 = y$.
- For any points p, q, r, $b(p, q, r)$ if and only if the following three conditions hold: $(p_1 - q_1) \cdot (q_2 - r_2) = (p_2 - q_2) \cdot (q_1 - r_1)$, $0 \leq (p_1 - q_1) \cdot (q_1 - r_1)$, and $0 \leq (p_2 - q_2) \cdot (q_2 - r_2)$.
- For any points p, q, r, s, $c(p, q, r, s)$ if and only if $(p_1 - q_1)^2 + (p_2 - q_2)^2 = (r_1 - s_1)^2 + (r_2 - s_2)^2$.

The first collection of axioms guarantee that the things of sort σ_n behave like the real numbers. The second axiom says that the functions $-_1$ and $-_2$ give us a preferred coordinate system on our space; a point is uniquely determined by its two coordinates. The third and fourth axioms define the betweenness relation b and congruence relation c in terms of these preferred coordinates. (One could also add an axiom defining the Euclidean metric, but we will have no need for that in what follows.)

The theory AG, as given by these axioms, provides a particularly straightforward way of doing two-dimensional Euclidean geometry. It uses numbers—things of sort σ_n—to lay down a preferred set of coordinates on the points in our two-dimensional space—the things of sort σ_p—and then defines various basic geometric notions in terms of these coordinates.

We now consider the well known synthetic formulation of Euclidean geometry called E_2, which was stated by Tarski (1959). It is formulated in the language containing the sort symbol σ_p of points, along with the two relation symbols b and c representing the same betweenness and congruence relations that AG employed. The theory E_2 does not employ the sort σ_n of numbers that AG did, and it does not lay down coordinates. This means that E_2 cannot define b and c in terms of coordinates like AG did. Instead, E_2 has a collection of axioms guaranteeing that b and c behave in precisely the way that one expects them to. The reader is invited to consult Tarski (1959, p. 19) for the list of axioms.

In order to understand the relationship between AG and E_2, we need to record the following two familiar facts about them. Both of these results are proven using the techniques demonstrated by Tarski (1959, Theorem 1).

Theorem 1 (Representation theorem) *For every model M of E_2 there is a model N of AG such that $N|_{\{\sigma_p,b,c\}} = M$.*

Here $N|_{\{\sigma_p,b,c\}}$ is the model that results from 'forgetting about' all of the structures on N other than the set N_{σ_p} with the extensions of the predicates b and c. Roughly, this representation theorem is guaranteeing that in moving from AG to E_2, we are not 'adding in' any new structure to the models. It shows that from a model of E_2 one can simply 'add in' structures and not 'take away' any structures and arrive at a model of AG. The following uniqueness theorem then shows us exactly what structure we have to 'add back in' to a model of E_2 in order to recover a model of AG.

Theorem 2 (Uniqueness theorem) *Let M be a model of E_2 containing points $p \neq q$. There is a unique (up to isomorphism) model N of AG that satisfies the following three conditions:*

- $N|_{\{\sigma_p,b,c\}} = M$
- $p_1^N = 0^N$ and $p_2^N = 0^N$
- $q_1^N = 1^N$ and $q_2^N = 0^N$

Recall that an **isomorphism** $h : A \to B$ between L-structures A and B is a family of bijections $h_\sigma : A_\sigma \to B_\sigma$ for each sort symbol $\sigma \in \Sigma$ that preserves the predicate, function, and constant symbols in L.[14] Theorem 2 shows us that in order to recover a model of AG from a model of E_2, one needs to pick out 'benchmark' points p and q to serve as the origin of the coordinate system and the 'unit' of the coordinate system. Once the structure of these two preferred points has been added to a model of E_2, one can define the sort of numbers with its field structure and the preferred coordinate system. Theorem 2 therefore tells us precisely what has been excised from models of AG when we move to E_2, and in doing so, it guarantees that we have not excised too much structure from AG. Indeed, in moving from AG to E_2, we have excised precisely what we intended to: the structure of preferred coordinates and the number sort with its field structure.

With these basic facts about AG and E_2 in hand, we can ask whether or not the two theories are equivalent. There are a number of different standards of equivalence on the table. We have seen that there is good evidence that Putnam thought of something like definitional equivalence being necessary for equivalence of theories. But in recent years, a number of other standards have been proposed. For example, Barrett and Halvorson (2016b) introduced *Morita equivalence*, Weatherall (2016) and Halvorson (2012) introduced *categorical equivalence*, and Hudetz (2017) introduced *definable categorical equivalence*.[15] It has been shown that the most liberal of these standards is categorical equivalence (Barrett and Halvorson 2016b). If two theories are definitionally equivalent—or, indeed, if they are equivalent according to any of these other standards—then they will be categorically equivalent. Moreover, it has been argued that standards that are more liberal than categorical equivalence are implausible, in the sense that they will judge theories to be equivalent that we have good reason to consider inequivalent (Barrett and Halvorson 2016b). This means that if we can show that E_2 and AG are *not* categorically equivalent, then we will have shown that they are inequivalent according to any reasonable standard of equivalence.

In order to show that the two theories are not categorically equivalent, we need a few preliminaries. Categorical equivalence is motivated by the following simple observation: First-order theories have categories of models. A **category** C is a collection of objects with arrows between the objects that satisfy two basic properties. First, there is an associative composition operation $\circ$ defined on the arrows of C, and second, every object c in C has an identity arrow $1_c : c \to c$. If T is a Σ-theory, we will use the notation Mod(T) to denote the **category of models** of T. An object in Mod(T) is a model M of T, and an arrow $f : M \to N$ between objects in Mod(T) is

[14] See Barrett and Halvorson (2016b) for a precise definition.

[15] See Weatherall (2019) for a review of recent work on equivalence.

an elementary embedding $f : M \to N$ between the models M and N. One can easily verify that Mod(T) is a category.

Before describing categorical equivalence, we need some additional terminology. Let C and D be categories. A **functor** $F : C \to D$ is a map from objects and arrows of C to objects and arrows of D that satisfies

$$F(f : a \to b) = Ff : Fa \to Fb \qquad F(1_c) = 1_{Fc} \qquad F(g \circ h) = Fg \circ Fh$$

for every arrow $f : a \to b$ in C, every object c in C, and every composable pair of arrows g and h in C. Functors are the “structure-preserving maps” between categories; they preserve domains, codomains, identity arrows, and the composition operation. A functor $F : C \to D$ is **full** if for all objects c_1, c_2 in C and arrows $g : Fc_1 \to Fc_2$ in D there exists an arrow $f : c_1 \to c_2$ in C with $Ff = g$. F is **faithful** if $Ff = Fg$ implies that $f = g$ for all arrows $f : c_1 \to c_2$ and $g : c_1 \to c_2$ in C. F is **essentially surjective** if for every object d in D there exists an object c in C such that $Fc \cong d$. A functor $F : C \to D$ that is full, faithful, and essentially surjective is called an **equivalence of categories**. The categories C and D are **equivalent** if there exists an equivalence between them.[16]

We can now turn to our main result about AG and E_2.

Proposition 5 *AG and E_2 are not categorically equivalent theories.*

The basic idea behind the proof is simple. Recall that an **automorphism** of an L-structure M is just an isomorphism from M to itself. One first shows that there is a model M of AG that is **rigid**, in the sense that its only automorphism is the identity map. No model of E_2 has this property, and this immediately implies that the two theories cannot be categorically equivalent. An equivalence of categories $G : \text{Mod}(AG) \to \text{Mod}(E_2)$ must map rigid models to rigid models.

Proof We begin by defining a model M of AG as follows: $M_{\sigma_n} = \mathbb{R}$, $M_{\sigma_p} = \mathbb{R} \times \mathbb{R}$, with the coordinate functions $-_1$ and $-_2$ and predicate symbols b and c defined in the natural manner. One easily verifies that M satisfies the axioms of AG. Now suppose that $h : M \to M$ is an automorphism. It is well known that the identity map is the only automorphism of the ordered field $\mathbb{R}$, so h_{σ_n} must be the identity. Let $p \in M_{\sigma_p}$. Since h is an automorphism, it preserves the coordinate functions in the sense that

$$h_{\sigma_p}(p)_1 = h_{\sigma_n}(p_1) \qquad\qquad h_{\sigma_p}(p)_2 = h_{\sigma_n}(p_2)$$

Since h_{σ_n} is the identity map, this means that $h_{\sigma_p}(p)_1 = p_1$ and $h_{\sigma_p}(p)_2 = p_2$. The point p is the unique point with the coordinates p_1 and p_2, however, so it must be

[16] The concept of a “natural transformation” is often used to define when two categories are equivalent. C and D are equivalent if there are functors $F : C \to D$ and $G : D \to C$ such that FG is naturally isomorphic to the identity functor 1_D and GF is naturally isomorphic to 1_C. See Mac Lane (1971) for the definition of a natural transformation and for proof that these two characterizations of equivalence are the same.

that $h_{\sigma_p}(p) = p$. We have therefore shown that h is the identity. Since h was arbitrary, M is rigid.

Now let N be a model of E_2. Theorem 1 implies that there is a model M of AG such that $M|_{\{\sigma_p,b,c\}} = N$. Let $p \in N_{\sigma_p}$. We define a map $h : N_{\sigma_p} \to N_{\sigma_p}$ by $h(p) = (p_1 + 1, p_2)$, i.e. $h(p)$ is the unique point in N_{σ_p} with those coordinates. Note that this definition makes sense since $N_{\sigma_p} = M_{\sigma_p}$. One now uses the axioms of AG that define b and c to calculate that h preserves these predicate symbols, and is therefore an automorphism of N. Since N was arbitrary, no model of E_2 is rigid. This immediately implies that AG and E_2 are not categorically equivalent. □

5 What is it to be Dispensable?

If a nominalist is aiming to provide a theory that bears the relationship to standard physics that synthetic Euclidean geometry bears to analytic Euclidean geometry, then they are denying P4. The previous section shows us that these two theories are not equivalent. Proposition 5 implies that they are not definitionally equivalent, and moreover, that they are not equivalent according to any reasonable standard of equivalence. The reason why is intuitive: AG posits more structure—in the form of preferred coordinates and the sort σ_n—than E_2 does. If the move from analytic to synthetic geometry is a canonical example of how to go about dispensing with numbers from a theory, then a nominalistic physics should similarly be inequivalent from standard physics, insofar as its aim is to dispense with something.[17]

The purpose of this section is to discuss the reasons why one might have been drawn to P4 in the first place. The most charitable reconstruction of an argument for P4 appeals to a particular view about what it is for something to be dispensable. Although it is not often discussed, the standard notion of dispensability in the literature on the indispensability argument is something like the following.

[17] Instead of appealing to a single preferred coordinate system as AG does, one might formulate analytic geometry by laying down a family of 'generalized coordinates' (Burgess and Rosen 1997, II.A.3). This kind of formulation will affirm a wider class of symmetries than AG does, making it so that the argument in proof of Proposition 5 does not go through. Indeed, one suspects that this formulation of analytic geometry will be equivalent to E_2. It is natural to think, however, that moving to E_2 in this case does not actually dispense with anything. Rather, it simply elucidates what the actual commitments of this formulation of analytic geometry were in the first place. By affirming this wider class of symmetries there is a sense in which we 'take back' or 'weasel away', to use a phrase of Melia (2000), some of AG's commitments. For example, if we were to put forward the Newtonian theory of space, with its standard of absolute rest, while at the same time affirming that 'Galilean boost' symmetries preserved all of the spacetime structures structures that we took to be significant, then there would be a strong sense in which we were not actually committing to there being an absolute standard of rest. A full discussion of this idea is unfortunately beyond the scope of this paper. See Weatherall (2016) for arguments about classical electromagnetism and Newtonian gravitation that are closely related.

Definition An entity or piece of structure X is **dispensable** to a theory T just in case there is a theory T^- that is empirically equivalent to T, does not appeal to X, and is suitably attractive (Colyvan (2001), p. 77).

In other words, X is dispensable from T if there is a suitably attractive empirically equivalent reformulation of T that does not appeal to X. It is worth taking a moment to unravel this definition. It is requiring that the 'dispensing theory' T^- satisfy three conditions.

First, we need to guarantee that T^- 'does the work' that T set out to do in the first place. Consider again the case of the analytic formulation of Euclidean geometry AG that we discussed above. One could not, for example, put forward group theory to show that numbers are dispensable from the theory AG. Group theory is not formulated using the apparatus of numbers, but this is irrelevant since it is a *completely different* theory than AG. It describes groups, which are obviously a completely different kind of thing than the Euclidean space that AG describes. It does not 'do the work' of the theory we began with. The dispensing theory T^- must therefore be sufficiently similar to the theory we began with, or else it is simply irrelevant. This is usually made precise by requiring that T^- is empirically equivalent to T. For example, Colyvan writes that the dispensing theory T^- must have the "same observational consequences" as the theory that we began with (Colyvan (2001), p. 77).[18]

Second, the requirement that T^- does not appeal to X is supposed to guarantee that it 'gets rid of' the entity or piece of structure X. Although something like this should obviously be a necessary condition on a theory dispensing with X, it is rarely made precise what it means for T^- to appeal to X. Colyvan (2001, p. 77) writes that T^- must neither mention nor predict the entity or structure in question. This kind of characterization is common in the literature on indispensability, but it is fairly imprecise. In particular, one wonders how we can tell whether T^- mentions or predicts X. We will return to this requirement and attempt to sharpen it shortly.

Third, the requirement that T^- is suitably attractive is simply meant to guarantee that T^- is not 'ad hoc'. Both Field (2016) and Colyvan (2001) appeal to Craig's theorem to argue that if something like this requirement were not adopted, then every theoretical entity would be dispensable. It is notoriously hard to make precise what it is for a theory to be 'suitably attractive', but fortunately this requirement will not be a central concern in what follows.

[18] It is worth considering whether or not the requirement that T and T^- are empirically equivalent is robust enough to capture what is going on in some cases of dispensability. In many cases it actually seems that the two theories bear a much closer relationship to one another than mere empirical equivalence. Indeed, since neither of the theories AG or E_2 have any empirical content, this suggests that perhaps the standard definition of dispensability merits revision. As we will see below, if we require that T and T^- bear *too* close a relationship to one another, the notion of dispensability we end up with is implausible. So the question is: exactly what relationship should we require T^- to bear to T? We will make a brief remark about this in the conclusion. Note also that some theory T^- might not postulate X but make *new or better* empirical predictions than the original theory T. While there is no doubt a sense in which T^- dispenses with X, this is not the kind of dispensability that Colyvan is after. He is trying to pin down the kind of dispensability at play in Field-style attempts to dispense with abstract objects, where one is trying to capture the same empirical content as the original theory without appeal to X.

Our examination of Putnam's proof provides us with evidence that Putnam's understanding of dispensability differs from this standard definition. Although Putnam does not explicitly discuss what it is for something to be dispensable, his endorsement of premise P4 points to how he must be thinking about it. In order for T^{-} to 'do the work' of T, one might think that the two theories have to bear a closer relationship to one another than mere empirical equivalence. Putnam's commitment to P4 suggests that he thinks that the two theories must actually be *fully equivalent*, as captured by some formal relation like definitional equivalence, rather than merely being empirically equivalent. The most charitable understanding of the notion of dispensability that Putnam has in mind—and moreover, the kind of dispensability that a proponent of P4 is implicitly adopting—is therefore the following.

Definition An entity or piece of structure X is **dispensable*** from a theory T if there is theory T^{-} that is equivalent to T, does not appeal to X, and is suitably attractive.

In other words, X is dispensable* from T if there is a suitably attractive equivalent reformulation of T that does not appeal to X. This way of understanding Putnam provides the most natural explanation of his endorsement of P4. He is simply requiring that in order for T^{-} to 'do the same work' as T, the relationship between the two theories must be much more robust than what is standardly required. If one thinks that dispensability* is a proper understanding of what it is for something to be dispensable, then one is committed to the equivalence of T_n and T_p being a necessary condition on T_n dispensing with anything from T_p. And that leads one immediately to endorse P4.

This provides us with a clearer picture of what is going on in Putnam's proof. The proof is showing that abstract objects are not dispensable* to standard physics. Insofar as a theory T_n is suitably nominalistic, it cannot appeal to abstract objects like numbers, functions and sets. Putnam claims that this means that T_n must satisfy the conditions given by premises P2 and P3. Then the fact that P1 holds implies that T_n cannot be equivalent to T_p. Since T_n must be equivalent to T_p in order to dispense* with something from T_p, this means that abstract objects are not dispensable*. The basic idea behind Putnam's proof is therefore the following: The theory T_n cannot be suitably nominalistic while at the same time 'doing the work' of T_p.

With two definitions of dispensability now on the table, one naturally wonders which one is more appropriate. I will argue that dispensability* is an unsatisfactory understanding of what it is for something to be dispensable. There are both cases of structures that are not dispensable* that we want to say are dispensable, and cases of structures that are dispensable* that we want to say are not dispensable. Indeed, the problem is that there is a tension inherent in the definition of dispensability*. If one requires that the dispensing theory T^{-} bear such a close relationship to the original theory T—that is, if one requires that they are equivalent—one actually precludes T^{-}

from ‘getting rid of’ X. Insofar as the two theories are equivalent, there is a strong sense in which T^- appeals to precisely the same entities and structures that T does.[19]

The following two examples illustrate precisely this point. The case of analytic and synthetic geometry discussed above shows that there are things that we want to say *are* dispensable that are *not* dispensable*.

Example 1 Consider the coordinate functions $-_1$ and $-_2$ and the sort σ_n of numbers from the theory AG. Tarski’s synthetic formulation of Euclidean geometry E_2 appeals to neither the coordinate functions nor the sort of numbers.

We have good reason to think that E_2 dispenses with numbers. Clearly, the theory does not employ a number sort, but it also does not *define* a sort of numbers. Indeed, we have seen this when we show that the two theories are not categorically equivalent. This implies that they also cannot be Morita equivalent, since Morita equivalence entails categorical equivalence (Barrett and Halvorson 2016b). The precise details of Morita equivalence are not important for our purposes here, but the basic idea is simple. One can define new sort symbols—just like one can define new predicate, function, and constant symbols—using some basic construction rules. Two theories are then said to be Morita equivalent if they have a ‘common Morita extension’, which is just like a common definitional extension except that it might define new sorts in addition to new predicates, functions, and constants. The fact that E_2 and AG are not Morita equivalent means that E_2 does not have the resources required to define the number sort σ_n that AG uses. This captures a strong sense in which E_2 does not appeal to numbers (or coordinate functions, for that matter).[20]

We want to say that E_2 dispenses with the structure given by the coordinate functions and the sort of numbers. But E_2 is not equivalent to AG, unless one adopts a standard of equivalence that is implausibly liberal. That means that E_2 does not dispense* with numbers from AG. Our second example shows that there are structures that are dispensable* that we actually want to say are *not* dispensable.

Example 2 Consider again the analytic formulation of Euclidean geometry AG. We define another theory AG^- in the language $\{\sigma_p, \sigma_n, \leq, +, \cdot, 0, 1, -_1, -_2\}$, or in other words, the language of AG without the symbols b and c denoting betweenness and congruence. All of these other symbols are thought of in the same way as before.

[19] This basic idea shares much in common with Alston (1958), whose target was the Quinean practice of ‘paraphrasing away’ ontological commitments. He pointed out that if the paraphrases preserve the content of the original statements, then they must have precisely the same commitments.

[20] If one has constants denoting two different points, then one can use the techniques of Tarski (1959) to define a real-closed field using E_2. (A similar result is true of the system of Field (2016). His discussion in chapter 4.1 is closely related to the following point.) One might worry that this means that E_2 *is* still committed to numbers. This thought is, however, a bit misleading. The definition of such a real-closed field relies on those two constant symbols, which label the ‘0’ and ‘1’ points. It is more appropriate to say, therefore, that the extension of E_2 to a signature containing these two additional constant symbols (with an axiom saying they are non-equal)—and not there mere theory E_2—can define a real-closed field. And it is natural to think that this extension of E_2 has more structure than E_2 itself did, since E_2 cannot itself define those two constant symbols.

The theory AG^- has the same axioms as AG minus the axioms that AG used to define the symbols b and c. So in particular, AG^- has axioms saying that the elements of sort σ_n form a real-closed field, along with the axioms guaranteeing that the function symbols $-_1$ and $-_2$ behave like coordinates.

Now it is easy to see that the pieces of structure b and c are dispensable*. We have exhibited a theory AG^- that is suitably attractive, at least insofar as AG itself is. Moreover, AG^- is definitionally equivalent to AG, as one can easily verify. The theory AG is itself a definitional extension of AG^-. And lastly, on the face of it AG^- does not appeal to the structures b or c, in the sense that neither appears in the language of the theory.

This example shows that b and c are dispensable* from AG. But this seems like the wrong verdict entirely. A closer consideration of AG^- shows why this is the case. There is a strong sense in which AG^- has dispensed with nothing from AG. Consider the concept of congruence that the predicate c denotes in AG. The theory AG^- still has the conceptual resources to talk about the distance between the points p and q being the same as the distance between the points r and s. Indeed, the following familiar formula in the language of AG^- says precisely that:

$$(p_1 - q_1)^2 + (p_2 - q_2)^2 = (r_1 - s_1)^2 + (r_2 - s_2)^2$$

One no doubt recognizes this formula as the formula that AG used to define the predicate c in the first place. The point here is a simple one. Even though AG^- does not explicitly appeal to the structures b and c—neither of those symbols appears in the language of AG—AG^- still implicitly appeals to them. AG^- has the resources to talk about the concepts of betweenness and congruence just like AG could originally, it is simply not as convenient for AG^- to do so since it does not contain the 'shorthand' or 'abbreviations' b and c for these concepts. This captures a sense in which nothing has been dispensed with when we move from AG to AG^-, and therefore, dispensability* seems to once again make the wrong verdict.

The basic idea here is almost trivial, but it is nonetheless important to appreciate. The concept of dispensability* is flawed for the following reason. If a theory T^- is actually equivalent to T, then there is a strong sense in which nothing has been dispensed with in the move from T to T^-. We can make this point a bit more carefully. Recall that neither Colyvan's standard conception of dispensability nor Putnam's alternative conception made perfectly precise what it is for a theory to *not appeal* to some entity or structure X. The requirement of Colyvan (2001) that the theory not mention or predict X is not as useful in practice as one would like. In Example 2, for instance, the structures b and c are not mentioned by the theory AG^-, since neither of those symbols appear among the vocabulary that the theory employs. But as we have seen, there is nonetheless a sense in which the theory still appeals to them. A clearer method for telling whether or not a theory appeals to X in its formulation is therefore needed.

Our examination of Euclidean geometry points to such a method. In the case of Euclidean geometry, one can tell from the form of the uniqueness theorem that some genuine dispensing has occurred in the move from AG to E_2. In particular, the uniqueness theorem does *not* say that one can recover a model of AG from a mere model of E_2. If that were the case, then there would be a sense in which nothing had been dispensed with in the move from AG to E_2; the models of E_2 would be able to recover, without the addition of any structure, the models of AG. Rather, the uniqueness theorem says that *once one supplements* a model of E_2 with the extra structure provided by a choice of two preferred points, one can recover a model of AG. Only after the structure of these two preferred points has been added to a model of E_2 can one can define the sort of numbers with its field structure and the preferred coordinate system.

We therefore have a positive proposal for how to tell what, if anything, has been dispensed with in the move from a theory T to a purported dispensing theory T^-. We simply examine the form of the uniqueness theorem that relates T^- to T. If one needs to 'add back in' some structure to models of T^- in order to recover models of T, then something has been dispensed with: precisely that structure that one needs to add back in. If one does not need to add back in any structure to models of T^- in order to recover models of T, then nothing has been dispensed with in the move to T^-. This sharpened understanding of dispensability allows us to better diagnose what is going wrong with dispensability* by appealing to the following well known fact (Hodges (2008), Theorem 2.6.3).

Proposition 6 *Suppose that $L_n \subset L_p$ are languages. If the L_p-theory T_p and the L_n-theory T_n are definitionally equivalent, then for every model M of T_n there is a unique model M^+ of T_p such that $M^+|_{L_n} = M$.*

This result is simply saying that if the nominalistic theory T_n is definitionally equivalent to T_p, as Putnam's premise P4 requires, then one can uniquely recover models of T_p from models of T_n without supplementing models of T_n with any additional structure. This helps isolate the exact sense in which P4 is false. If T_n and T_p are definitionally equivalent, then the form of the uniqueness theorem relating them shows that nothing has been dispensed with in the move from T_p to T_n. So insofar as one wants to dispense with some entity or structure that the theory T_p employs, one should actually be aiming, contra P4 and the concept of dispensability*, to formulate a theory T_n that is *not* equivalent to T_p. If the two theories are equivalent—that is, if they are the same theory presented to us in different guises—then they appeal to *precisely the same* entities and structures. Putnam's premise P4 and the related concept of dispensability* therefore have things the wrong way around. Insofar as T_n dispenses with abstract objects from T_p it must actually be *in*equivalent to T_p.

6 Conclusion

We conclude by discussing two further payoffs that this examination of Putnam's proof yields. First, the discussion allows us to draw some conclusions about what, in general, the relationship between two theories should be when one purports to 'dispense with' or 'excise' something from the other. And second, clarifying Putnam's understanding of dispensability yields a better appreciation of the role that equivalence plays in his thought more generally.

6.1 Dispensability and Translation

We have seen that in order for a theory T^- to dispense with something from a theory T, it is important that the two theories *not* be equivalent. This is the primary reason that Putnam's proof is unsuccessful. The premise P4 is false. This simple realization serves to tie the debate about the indispensability argument to the recent discussion of equivalence among logicians and philosophers of science. In order to be able to conclusively say whether or not something has been dispensed with, one first has to settle on a standard of equivalence.

If a dispensing theory T_n and the original theory T_p should not be equivalent, then a question remains: What relationship should they bear to one another? It is worth making one remark here on this point. There is good reason to think that in addition to being inequivalent, there should not even be a translation from T_p to T_n. Or in other words, the nominalist who aims to dispense with abstract objects should be perfectly comfortable accepting the conclusion C1 of Putnam's proof. Recall that Putnam concluded from P1, P2, and P3 that there is no conservative translation from T_p to T_n, and then he argued that this is a problem. Putnam thinks that what the nominalist "wishes to do is to find a 'translation function'" from the standard physics T_p into their nominalistic physics T_n.

The view that a nominalistic physics requires some kind of translation of physics into a nominalistic language is not unique to Putnam. It is, in fact, quite widespread. It goes back to the classic work of Goodman and Quine (1947, pp. 121–122), who attempted to address "the problem of translating into nominalistic language certain [...] sentences which had appeared to be explicable only in platonistic terms." Burgess and Rosen (1997, p. 5) call this tradition of trying to find a translation from platonistic to nominalistic language *reconstructive nominalism*, which they describe as "seeking accommodation through reconstrual or reinterpretation of those ways of speaking that appear to involve abstract entities, so as to render at least a large part of them compatible with overarching nominalistic scruples." Emphasizing just how widespread reconstructive nominalism became in the years following the initial work of Goodman and Quine, they mention that "[t]he stream of publications by later reconstructive nominalists began as a trickle in the 1960s, grew in the 1970s, and became a torrent by the 1980s" (Burgess and Rosen 1997, p. 6).

Despite this widespread desire to be able to translate from T_p to T_n, the conclusion C1 is actually one that a nominalist can be perfectly happy with. Indeed, a nominalist who aims to dispense with abstract objects—or for that matter, anyone proposing a theory that aims to dispense with or excise something—should *not want any* translation to exist from T_p to T_n. The following simple proposition helps us begin to see why.[21]

[21] This result is closely related to a well known fact about 'intertranslatability' and definitional equivalence. See Barrett and Halvorson (2016a).

Proposition 7 *Let $L_n \subset L_p$ be languages. Suppose that $F : T_p \to T_n$ is a translation from the L_p-theory T_p to the L_n-theory T_n and that the following conditions hold*:

(i) $T_p \vDash F\psi \leftrightarrow \psi$ *for every L_p-formula* ψ
(ii) $T_n \vDash F\phi \leftrightarrow \phi$ *for every L_n-formula* ϕ
(iii) T_p *is an extension of* T_n (*i.e. if* $T_n \vDash \phi$, *then* $T_p \vDash \phi$).

Then T_p and T_n are definitionally equivalent.

Proof Consider the L_p-theory $T_n \cup \{p \leftrightarrow Fp : p \in L_p - L_n\}$. We show that this theory is logically equivalent to T_p, which will imply that T_p is a definitional extension of T_n, and therefore the two are definitionally equivalent.

So suppose first that M is a model of $T_n \cup \{p \leftrightarrow Fp : p \in L_p - L_n\}$ and that $T_p \vDash \psi$ for some L_p-sentence ψ. We know that $M \vDash p \leftrightarrow Fp$ for each $p \in L_p - L_n$. In conjunction with ii) this implies that that $M \vDash \psi \leftrightarrow F\psi$. Since F is a translation, we know that $T_n \vDash F\psi$, and therefore $M \vDash F\psi$. So $M \vDash \psi$ and M is therefore a model of T_p. On the other hand, suppose that M is a model of T_p and that $T_n \vDash \phi$ for some L_n-sentence ϕ. It follows from iii) that $T_p \vDash \phi$, so $M \vDash \phi$. Condition i) implies that $M \vDash p \leftrightarrow Fp$ for each $p \in L_p$. So M is a model of the theory $T_n \cup \{p \leftrightarrow Fp : p \in L_p - L_n\}$. □

Proposition 7 is telling us that if a particularly natural kind of translation exists from T_p to T_n, then the two theories must be definitionally equivalent. As discussed above, this would be a problem for the nominalist whose aim in formulating T_n was to dispense with something from T_p. Insofar as the two theories are definitionally equivalent, nothing has been dispensed with. Proposition 6 guarantees that the form of the uniqueness theorem shows this.

The question is, therefore, whether the kind of translation employed in Proposition 7 is the kind of translation that Putnam and others think nominalists should require. There is good reason to suppose that it is. First, one can verify that F is conservative, so it is the kind of translation Putnam thinks nominalists require. And moreover, the three conditions that Proposition 7 imposes on F are natural requirements to impose on a translation from the standard physics T_p into nominalistic physics T_n, if one does indeed exist. Condition (i) requires that standard physics, which contains both nominalistic and 'non-nominalistic' vocabulary, says that every formula translates via F to a formula that 'says the same thing', in the weak sense of having the same extension in every model of T_p. A particularly famous translation, employed for instance by Goodman and Quine (1947, p. 108), satisfies this condition: The L_p-sentence "Class A has three members" and its translation in L_n "There are distinct objects x, y, and z such that anything is in A if and only if it is x or y or z" are equivalent by the lights of T_p. Similarly, condition (ii) requires that the nominalistic vocabulary $L_n \subset L_p$ is translated in such a way that each formula 'says the same thing' as its translation, again in the weak sense of having the same extension in all models of T_n. A translation that simply translates each piece of vocabulary in L_n *to itself* will satisfy this condition. And lastly, condition iii) is a natural requirement on the relationship between T_p and T_n: T_p must say everything that T_n says in nominalistically acceptable vocabulary, though possibly more. Intuitively, all three conditions should hold if F is to be considered an acceptable translation from T_p to T_n.

This means that there should not be a translation, conservative or otherwise, from T_p to T_n, at least insofar as T_n is supposed to dispense with something.[22] If there is a translation $F : T_p \to T_n$, this captures a sense in which T_n can define or 'build' all of the structures of T_p. Or in other words, it captures a sense in which T_n is as 'ideologically rich' as T_p (Quine 1951, p. 15). Any formula ϕ in the language of T_p would be expressible using the language of T_n. The theory T_n would be able to define all of the structures that T_p has. Or in other words, T_n could express all of the same concepts as T_p. Insofar as T_n dispenses with something from T_p, therefore, there should not be a translation in this direction. Indeed, if there is a translation $F : T_p \to T_n$, that captures a sense in which T_n dispenses with *nothing* from T_p. The nominalist who aims to dispense with abstract objects is therefore free to agree with Putnam's conclusion C1. There should not be a translation from T_p to T_n. If there were, that would mean that nominalistic theory T_n would have the resources required to express statements about numbers. Given that it was proposed with the aim of dispensing with numbers, this is exactly what it was *not* supposed to be able to do.[23]

6.2 Equivalence and Putnam

We conclude with a more scholarly remark on Putnam. As was mentioned above, the topic of equivalent theories was a familiar and "profoundly significant" one for Putnam (1983, p. 45). Indeed, Linnebo (2018, p. 249) recalls that a "recurring theme [of Putnam's lectures and seminars] was that of different but 'equivalent' descriptions of one and the same aspect of reality". Understanding Putnam's proof allows us to trace his interest in this topic back to his indispensability argument.

In particular, there is a puzzle that has recently arisen with regard to Putnam's indispensability argument. In a paper published four decades after *Philosophy of Logic*, Putnam (2012) suggests that his indispensability argument has been misunderstood, and is considerably different from the one that is standardly attributed to him in the literature. In particular, he did not intend for it to be an argument for platonism. Rather, he intended for it to be an argument for the

[22] As mentioned at the beginning of Sect. 3, one can imagine a nominalist—for example, the hermeneutic nominalist of Burgess and Rosen (1997)—who wants to show that T_p and T_n are equivalent in order to show that T_p was not actually committing to abstract objects in the first place. Such a nominalist is not proposing T_n with the aim of dispensing with something from T_p, but rather in order to clarify the content of T_p. See, for example, the earlier discussion in footnote 19. This kind of nominalist would be fine with there being a translation from T_p to T_n and would have to deny P2 or P3.

[23] It is worth mentioning that one expects a similar argument to go through even if one moves to a more general notion of translation. Given some of the claims he made in later years about the equivalence of geometry with points and geometry with lines, one can imagine Putnam endorsing a more liberal standard of equivalence (like Morita equivalence), which corresponds to a 'looser' notion of translation than the one we have discussed here (Barrett and Halvorson 2017). The important point, however, is that if a good translation exists from T_p to T_n—regardless of how one makes it formally precise—that will allow one to express in T_n all of the statements about numbers that T_p is capable of formulating.

objectivity of mathematics (Liggins 2008; Bueno 2018). As Burgess (2018) puts it, Putnam thinks that "insufficient attention to what he has written about 'equivalent descriptions' in [(Putnam 1967)] and elsewhere" resulted in this general misunderstanding of his indispensability argument. Putnam did mention equivalent descriptions in the final chapter of *Philosophy of Logic* as one of the topics that he would have discussed if he had the space. So this suggests that equivalence plays some role in the arguments given in *Philosophy of Logic*, but that is puzzling since it is only explicitly mentioned in the book that one time. In brief, the puzzle is the following: Putnam clearly thinks that equivalence plays some important role in his indispensability argument, but it is unclear what this role is.

Our examination of Putnam's proof allows us to take a step towards resolving this puzzle. Putnam was committed to P4 in his proof of the impossibility of a nominalistic physics. He was therefore trying to demonstrate that we cannot reformulate standard physics in such a way that its content is preserved—i.e. so that the reformulation is equivalent to or 'says the same thing' as standard physics—without appealing to abstract objects like numbers. The concept of equivalence therefore plays a crucial role in Putnam's understanding of dispensability. This explains why Putnam believed that more attention needed to be paid to his remarks on equivalence in order to understand his indispensability argument. In order to understand the notion of dispensability—that is, dispensability*—that Putnam has in mind, one has to understand what it is for two theories to be equivalent.

This isolates a further difference between Putnam's indispensability argument and the indispensability arguments that came after it, like the one given on the first page of this paper. He has already explicitly stated that his conclusion was meant to be different. But our discussion here also implies that—since dispensability* is not the same as the concept of dispensability that later became standard—his premises involving indispensability are different too. He was employing a different, and unfortunately worse, notion of dispensability than what became standard in the literature that followed.

Now that we have Putnam's understanding of dispensability clearly on the table in the form of dispensability*, we can ask what kind of metaphysical mileage he is trying to get out of it. Many of Putnam's later arguments—most famously, those involving 'conceptual relativity' (Putnam 1977, 1992, 2001)—use facts about equivalent theories to draw conclusions about ontological matters. So it would be perfectly in character for him to use facts about equivalence or inequivalence of theories to draw metaphysical conclusions in the context of the indispensability argument as well. But it is entirely unclear that dispensability* will do any work for Putnam. As we have seen, it is a poor notion of dispensability—indeed, it's plagued by an internal tension—and so it is hard to imagine that one can draw any compelling conclusions whatsoever from facts about what is and what is not dispensable*.

Acknowledgements Thanks to Jeff Barrett, Alex LeBrun, Hannes Leitgeb, and two anonymous referees for their extremely thoughtful and helpful comments.

References

Alston, W. P. (1958). Ontological commitments. *Philosophical Studies*, *9*, 8–17.

Arntzenius, F., & Dorr, C. (2011). *Space, Time and Stuff, chapter 8*. Oxford: Oxford University Press.

Artigue, M., Isambert, E., Perrin, M., & Zalc, A. (1978). Some remarks on bicommutability. *Fundamenta Mathematicae*, *101*(3), 207–226.

Balaguer, M. (1996). Towards a nominalization of quantum mechanics. *Mind*, *105*(418), 209–226.

Barrett, T. W. (2015). On the structure of classical mechanics. *The British Journal for the Philosophy of Science*, *66*(4), 801–828.

Barrett, T. W. (2017). Equivalent and inequivalent formulations of classical mechanics. *the British Journal for the Philosophy of Science* (Forthcoming).

Barrett, T. W., & Halvorson, H. (2016a). Glymour and Quine on theoretical equivalence. *Journal of Philosophical Logic*, *45*(5), 467–483.

Barrett, T. W., & Halvorson, H. (2016b). Morita equivalence. *The Review of Symbolic Logic*, *9*(3), 556–582.

Barrett, T. W., & Halvorson, H. (2017). From geometry to conceptual relativity. *Erkenntnis*, *82*(5), 1043–1063.

Barrett, T. W., & Halvorson, H. (2019). Mutual translatability, equivalence, and the structure of theories. *Manuscipt*.

Bueno, O. (2018). Putnam's indispensability argument revisited, reassessed, revived. *Theoria*, *33*(2), 201–218.

Burgess, J. (2018). Putnam on foundations: Models, modals, muddles. In G. Hellman & R. T. Cook (Eds.), *Hilary putnam on logic and mathematics* (pp. 129–143). Berlin: Springer.

Burgess, J. P., & Rosen, G. (1997). *A subject with no object: Strategies for nominalistic interpretation of mathematics*. Oxford: Clarendon Press.

Colyvan, M. (2001). *The indispensability of mathematics*. Oxford: Oxford University Press.

Colyvan, M. (2019). Indispensability arguments in the philosophy of mathematics. *Stanford Encyclopedia of Philosophy*.

Curiel, E. (2014). Classical mechanics is Lagrangian; it is not Hamiltonian. *The British Journal for the Philosophy of Science*, *65*(2), 269–321.

de Bouvére, K. L. (1965). Synonymous theories. In *Symposium on the Theory of Models* (pp. 402–406). North-Holland Publishing Company.

Field, H. (2016). *Science without numbers* (2nd ed.). Oxford: Oxford University Press.

Glymour, C. (1971). Theoretical realism and theoretical equivalence. In *PSA 1970* (pp. 275–288). Berlin: Springer.

Glymour, C. (1977). The epistemology of geometry. *Noûs*, *11*, 227–251.

Glymour, C. (1980). *Theory and Evidence*. Princeton: Princeton University Press.

Goodman, N., & Quine, W. V. O. (1947). Steps toward a constructive nominalism. *The Journal of Symbolic Logic*, *12*(4), 105–122.

Halvorson, H. (2012). What scientific theories could not be. *Philosophy of Science*, *79*(2), 183–206.

Halvorson, H. (2019). *The Logic in Philosophy of Science*. Cambridge: Cambridge University Press.

Hodges, W. (2008). *Model Theory*. Cambridge: Cambridge University Press.

Hudetz, L. (2017). Definable categorical equivalence: Towards an adequate criterion of theoretical intertranslatability. *Philosophy of Science* (Forthcoming).

Kanger, S. (1968). Equivalent theories. *Theoria*, *34*(1), 1–6.

Liggins, D. (2008). Quine, Putnam, and the 'Quine-Putnam' indispensability argument. *Erkenntnis*, *68*(1), 113–127.

Linnebo, O. (2018). Putnam on mathematics as modal logic. *Hilary Putnam on Logic and Mathematics* (pp. 249–267). Berlin: Springer.

Mac Lane, S. (1971). *Categories for the working mathematician*. Berlin: Springer.

Melia, J. (2000). Weaseling away the indispensability argument. *Mind*, *109*(435), 455–480.

Montague, R. (1957). *Contributions to the axiomatic foundations of set theory*. PhD thesis, University of California, Berkeley.

North, J. (2009). The 'structure' of physics: A case study. *The Journal of Philosophy*, *106*, 57–88.

Putnam, H. (1967). Mathematics without foundations. *The Journal of Philosophy*, *64*(1), 5–22.

Putnam, H. (1971). *Philosophy of logic*. London: Routledge.

Putnam, H. (1974). The refutation of conventionalism. *Noûs*, *8*(1), 25–40.

Putnam, H. (1977). Realism and reason. *Proceedings and Addresses of the American Philosophical Association.*, *50*(6), 1.
Putnam, H. (1983). Equivalence. *Philosophical Papers: Realism and Reason* (Vol. 3, pp. 26–45). Cambridge: Cambridge University Press.
Putnam, H. (1992). *Renewing Philosophy*. Harvard: Harvard University Press.
Putnam, H. (2001). Reply to Jennifer case. *Revue internationale de philosophie*, *4*(218), 1.
Putnam, H. (2012). Indispensability arguments in the philosophy of mathematics. In *Philosophy in an age of science: Physics, mathematics, and skepticism* (pp. 181–201). Harvard: Harvard University Press.
Quine, W. V. O. (1951). Ontology and ideology. *Philosophical Studies*, *2*(1), 11–15.
Shoenfield, J. (1967). *Mathematical logic*. Boston: Addison-Wesley.
Tarski, A. (1959). What is elementary geometry? In *The axiomatic method with special reference to geometry and physics* (pp. 16–29). North-Holland.
Weatherall, J. O. (2016). Are Newtonian gravitation and geometrized Newtonian gravitation theoretically equivalent? *Erkenntnis*, *81*(5), 1073–1091.
Weatherall, J. O. (2019). Theoretical equivalence in physics. *Philosophy Compass* (Forthcoming).