The Hole Argument and Putnam's Paradox

Bryan Cheng* and James Read†

Abstract

We discuss affinities and differences between (i) the hole argument in general relativity and (ii) Putnam's model-theoretic argument against metaphysical realism ('Putnam's paradox'). Following Pooley (2002), we maintain that the hole argument is not a special case of Putnam's paradox. This notwithstanding, both of these arguments have been responded to through meta-linguistic means. While van Fraassen (1997) claims that Putnam's paradox dissolves due to our inability to identify a function mapping our theories to objects in the world independent of our total language, Bradley and Weatherall (2022) maintain that the language of general relativity does not allow for the hole argument to be formulated. We compare these responses and assess the extent to which either is successful, concluding that we find van Fraassen's argument more persuasive precisely because of the greater generality of Putnam's paradox.

Contents

*bryan.cheng@philosophy.ox.ac.uk
†james.read@philosophy.ox.ac.uk

1 Introduction

The hole argument in its modern guise challenges 'manifold substantivalism', which is the position that the manifold in general relativity (GR) should be interpreted as existing independently of the material fields that lie upon it.[1] The hole argument claims that two models of GR related by diffeomorphism, differing non-trivially only in a 'hole region', are both empirically equivalent solutions of the theory, thereby leading it to suffer from both underdetermination and indeterminism problems—notions we will explain below. Putnam's model-theoretic argument—sometimes known as Putnam's paradox—targets metaphysical realism, which (at least in part) is the position that there exist mind-independent objects in the world such that a correspondence relation between these objects and words in our language constitutes truth, by suggesting the radical failure of our theories' referential terms to latch onto such mind-independent objects.[2]

There are clear affinities between these two arguments: in both cases, terms or structures in theories fail to correspond uniquely to entities in the world, motivating the rejection of certain ontological commitments—the spacetime manifold (*qua* entity independent of material fields) in the case of the hole argument, and a mind-independent world in the case of Putnam's paradox (subject to other assumptions, which we note in §3). However, as Pooley (2002) has noted and as we will expand on below in significant detail, there are important differences between the two arguments, with Putnam's paradox underwriting a more pervasive failure of correspondence than the hole argument. In any case, both of these arguments have been responded to in similar ways through meta-linguistic means, and it is upon such meta-linguistic responses which we focus in the later sections of this article.

While Weatherall's response to the hole argument (Weatherall 2018)—which can be understood as being meta-linguistic in nature, as elaborated upon by Bradley and Weatherall (2022)—claims that the language of GR does not have the resources to generate the hole argument, van Fraassen (1997) argues that Putnam's paradox dissolves due to our inability to identify a function mapping the terms in our theories to objects in the world in a manner independent of our own language.[3] We argue that these solutions possess essential dissimilarities because of differences in the original arguments; this is also why we consider the argument from Bradley and Weatherall to be more question-

[1]See Earman and Norton (1987) for the canonical statement of the hole argument in the modern philosophical literature, and e.g. Norton et al. (2023) for a survey of responses to the problem which have since been offered. Of course, at least some of the substantivalism/relationalism debate consists in establishing exactly what 'exists independently' amounts to—we won't discuss this further here, but see e.g. Pooley (2013) for further details.

[2]For more on metaphysical realism, see Button (2013, ch. 1).

[3]To be clear: Bradley and Weatherall (2022) don't explicitly compare their response to the hole argument with van Fraassen's response to Putnam's paradox; however, we take it that the clear parallels between the two responses to their respective arguments suggest that a philosophical comparison would be illuminating.

able than that from van Fraassen: the former invokes a stricture that GR be interpreted 'in the language of Lorentzian manifolds,' while the latter necessarily involves the total language of a community.

In §2 and §3, we introduce the hole argument and Putnam's paradox, respectively. In §4, we compare the two arguments. In §5 and §6, we introduce respectively Weatherall's response to the hole argument and van Fraassen's response to Putnam's paradox, before contrasting and evaluating the arguments in §7. In §8 we conclude.

2 The hole argument

The hole argument in its original form was introduced by Einstein during his development of GR, as an argument against generally covariant theories, which are theories invariant under general coordinate transformations (see e.g. Norton et al. (2023) or Stachel (2014) for the history). Motivated by a failure to reconcile his initial theory (which was generally covariant) with Newtonian gravitation theory in the weak field limit, Einstein sought to reject wholesale all generally covariant theories; the hole argument was a crucial tool for illustrating the supposedly rebarbative consequences of such theories. Eventually, however, Einstein would come to accept general covariance as integral to the formulation of GR, and would go on to argue that the physical content of a theory is exhausted by the catalogue of the spacetime coincidences which it licenses (his famous 'point coincidence argument'). According to this later version of Einstein (but speaking somewhat anachronistically), two models of GR related by hole diffeomorphism (more on which below) are such that if one represents a certain possible world then the other represents that world equally well, since spacetime coincidences are preserved by diffeomorphisms. Over 70 years later, however, Earman and Norton (1987) would revive the hole argument in order to attack the position of manifold substantivalism, introduced above.

We now present the hole argument a little more rigorously. The models of GR are tuples $\langle M, g_{ab}, \Phi \rangle$, consisting of a differentiable manifold M, a Lorentzian metric field g_{ab} on M, and fields Φ on M representing matter. These objects possess different mathematical properties, especially in terms of their respective isomorphisms. For differentiable manifolds, the standard of isomorphism is diffeomorphism: smooth transformations (with smooth inverses) preserving differential structure. The addition of a metric to a bare manifold results in a Lorentzian manifold, which implements further structure; the standard of isomorphism for Lorentzian manifolds is isometry, which in addition preserves distances and angles.[4]

To formulate the hole argument, define a diffeomorphism on M itself, $h : M \to M$. This mapping induces a transformation on tensor fields defined on the manifold,

[4]See e.g. Menon and Read (2023) for a discussion of different notions of isometry.

called a pullback, h^*. The new model generated by the diffemorphism is then written as $\langle M, \tilde{g}_{ab}, \tilde{\Phi}\rangle := \langle M, h^* g_{ab}, h^*\Phi\rangle$. The particular diffeomorphism considered is one for which the metrics g_{ab} and $\tilde{g}_{ab}$ are everywhere pointwise identical save for a 'hole', which is a subset of the manifold $H \subset M$ where the diffeomorphism is non-trivial.

A problem now arises because the Einstein equation, which is meant to determine the dynamical content of GR, is diffeomorphism invariant: any model of GR that is a solution to the Einstein equation yields a set of diffeomorphic models that are also solutions. Moreover, because the observable quantities in a spacetime theory are invariant under diffeomorphisms, all empirical content is preserved; no physical data will differentiate between the two models, even though the metric and matter fields have different values at manifold points within the hole.[5] This leads to a problem of underdetermination; there is also a problem of indeterminism, because if the hole $H \subset M$ lies to the future of some spacelike Cauchy surface Σ, then the laws plus data at a given time would seem to fail to fix what will happen to the future—a failure of Laplacian determinism.[6]

To formulate their argument against manifold substantivalism, Earman and Norton (1987, p. 522) propose the principle of 'Leibniz Equivalence': "Diffeomorphic models represent the same physical situation." They see Leibniz Equivalence as an *acid test* of substantivalism, arguing that a substantivalist cannot take diffeomorphic models to represent the same physical situation and must therefore reject the principle. While they do not invoke a specific substantivalist position, they give examples of associated views that violate Leibniz Equivalence, such as the position that "each model *is* a physically possible world, one of them being our world" (Earman and Norton 1987, p. 521).[7] If only diffeomorphism-invariant content is treated as physically real, then the manifold, which apparently contributes nothing to the physical description, should be eliminated (at least *qua* entity independent of material fields). This contradicts the position of the manifold substantivalist.

The conclusion of Earman and Norton's argument is therefore the rejection of manifold substantivalism: since the models do not agree on the points of the manifold upon which the fields take their values despite their being empirically equivalent, we should eliminate the extraneous structure of the manifold. The aforementioned principles are

[5]Notably, this argument presented by Earman and Norton "applies to all local spacetime theories and that includes generally covariant formulations of virtually all known spacetime theories", making it more general than Einstein's initial version (Norton et al. 2023).

[6]For much more detailed recent discussion of the underdetermination and indeterminism problems, see Pooley and Read (2025).

[7]Modulo an apparent equivocation between models and worlds, Rynasiewicz (1994, p. 409) labels this view 'model literalism': "Each model represents a possible physical situation, and distinct models represent distinct situations"; he identifies this view as leading to Einstein's initial rejection of diffeomorphism invariance. Another alternative to Leibniz Equivalence which Rynasiewicz invokes is 'model selectivism', which is "the thesis that some models may fail to represent any situation whatsoever, although distinct models which do represent situations represent distinct situations". This latter position linked to the views of authors such as Maudlin (1988) and Butterfield (1989) (Rynasiewicz 1994, p. 414).

not uncontroversial; there are responses denying that one's position regarding Leibniz Equivalence can be used as a measure of one's commitment to substantivalism. These responses have often been labelled 'sophisticated substantivalism'; we will not explore these arguments in this article.[8]

3 Putnam's paradox

We now move onto Putnam's paradox, which is a more general argument targeting 'metaphysical realism'. Button (2013, ch. 1), drawing on Putnam's own characterisation of the view, defines metaphysical realism by way of the following three criteria:

1. objects, relations, and properties in the world exist independently of the human mind, and are not dependent upon our perceptions;

2. truth is characterised as a correspondence relation between words and these mind-independent entities; and

3. even an ideal theory can be false.

The key idea behind Putnam's argument is that, given a particular model satisfying a theory which has terms referring to objects in the world and predicates whose extensions reflect relationships between these objects, one can always find a distinct model which also satisfies the theory. The intuition is that "there is no semantic glue to stick our words onto their referents, and so reference is very much up for grabs" (Lewis 1984, p. 221).

There are several versions of the model-theoretic argument, but we focus on a particularly simple formulation: what Button (2013) calls the 'permutation argument'.[9] As an example, Putnam proposes the sentence, 'A cat is on a mat.' Given a possible world in which this statement is true, the sentence can nevertheless "be reinterpreted so that in the *actual* world 'cat' refers to *cherries* and 'mat' refers to *trees* without affecting the truth-value" of the sentence in any possible world (Putnam 1981, p. 33). Putnam (1981, p. 34) demonstrates this through a redefinition of terms under three cases:

(a) Some cat is on some mat, and some cherry is on some tree.

(b) Some cat is on some mat, and no cherry is on any tree.

(c) Neither of the foregoing.

[8] See e.g. Pooley (2002, §4.1.4) for more on sophisticated substantivalism.

[9] For a more mathematically sophisticated version of Putnam's paradox based on the Löwenheim–Skolem theorem, see Putnam (1980). Both versions seek to demonstrate that, given a particular model satisfying a theory, one can always construct another model satisfying the theory using methods provided in the respective arguments.

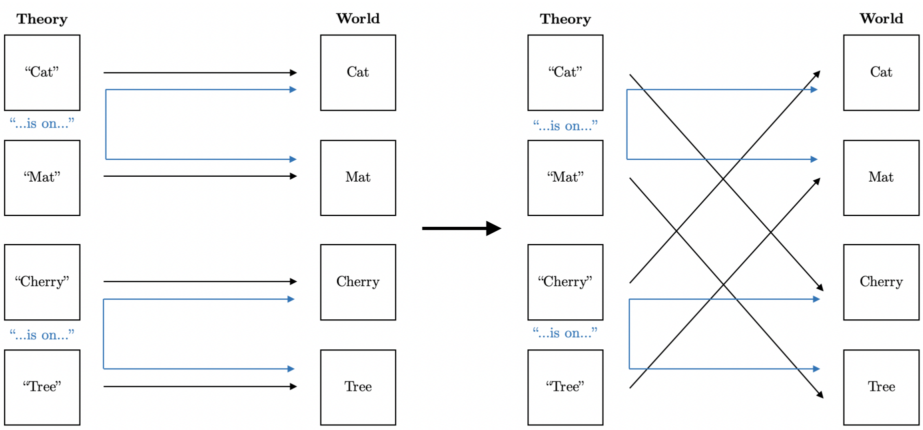

Figure 1: The permutation argument when predicates are the same.

Then consider the following definitions (Putnam 1981, p. 34):

- "x is a cat* if and only if case (a) holds and x is a cherry; or case (b) holds and x is a cat; or case (c) holds and x is a cherry."
- "x is a mat* if and only if case (a) holds and x is a tree; or case (b) holds and x is a mat; or case (c) holds and x is a quark."

In all possible worlds, the sentence "A cat is on a mat" is true if and only if "A cat* is on a mat*" is true. Reinterpreting "cat" as "cat*" and "mat" as "mat*" results in the failure of determination that Putnam seeks to highlight—despite the terms 'cat' and 'mat' referring to cherries and trees respectively in the *actual* world, the truth value of the sentence across possible worlds is preserved. Figure 1 demonstrates the relationships in the actual world between the names and referents (in black), as well as predicates and extensions (in blue), in Putnam's example above.

Despite permuting the objects to which the terms in the theory refer, the truth values of the sentences in the theory remain constant. Putnam's conclusion is that, regardless of the constraints we place on the interpretations of our models, even ones which fix truth values across all possible worlds, we fail to determine a unique model: "settling the truth values of every sentence in some language is insufficient to pin down the reference" of the terms in our theory (Button 2013, p. 16). We are therefore unable to verify whether our theories accurately refer to objects and relations in the world.

Putnam's above example, however, obscures the full generality of his argument.[10] Because both sentences utilise the predicate "...is on...", it appears initially that the permutation argument applies only to specific cases where the objects in question share predicates. Putnam's paradox in fact extends to more general cases where the predicates are *different*. For example, given the spacetime points A, B, C, and D, consider the

[10] To be clear: we don't intend to imply by this that Putnam was somehow unaware of the generality of his argument!

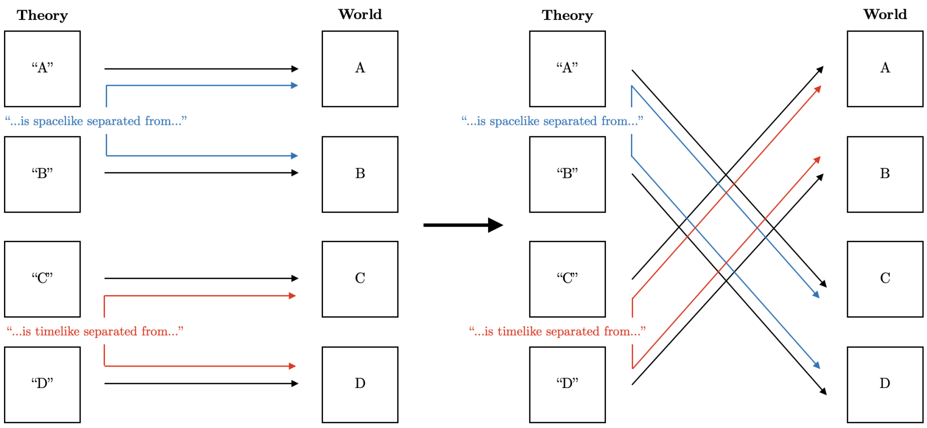

Figure 2: The permutation argument when predicates are different.

two sentences, "A is spacelike separated from B", and "C is timelike separated from D". Assume that the sentences are true in the actual world. To construct an equivalent scenario to above, one first swaps the names of spacetime points: A with C, and B with D. However, unlike the original scenario, one must *also* swap the extensions of the predicates: "...is spacelike separated from..." now has the pair {C, D} in its extension, while "...is timelike separated from..." has the pair {A, B} in its extension (see Figure 2).

Since the extensions and referents contained within each sentence remain matched, the truth values of the original sentences are constant under such a permutation, further demonstrating how Putnam's argument undermines our theories' relationship with the world. Note that these permutations describe the *same* physical scenario; the names and predicates associated with objects may change, but the physical properties in the actual world remain the same—an important point to which we return in §4. As such, in either case, at least some of the referents of names must change; otherwise, a sole change in the extension of predicates would suggest that the scenarios are just physically *different*.

One can link Putnam's views to Quine's arguments regarding the inscrutability of reference. Quine (1968) considers a situation in which one seeks to translate the term 'gavagai' in a foreign language to English. Assume that the term is used while gesturing at a rabbit. Because "a whole rabbit is present when and only when an undetached part of a rabbit is present; also when and only when a temporal stage of a rabbit is present", we cannot determine whether 'gavagai' refers to a rabbit, an undetached part of a rabbit, or a temporal stage of a rabbit (Quine 1968, p. 188). Quine argues that this failure of translation applies equally to the issue of reference in our own language, resulting in further inscrutability of reference. This reinforces Putnam's argument for the failure of metaphysical realism—our theories appear to be satisfied by any model with the correct cardinality.[11]

[11]Such arguments highlight the similarity with Newman's objection in the context of epistemic struc-

4 Comparison of the arguments

The affinities between Putnam's paradox and the hole argument have been noted by authors such as Rynasiewicz (1994) and Liu (1996), who indeed claim that the hole argument is a special case of Putnam's paradox. In this section, we focus first on Rynasiewicz's position, before turning our attention to Liu's alternative approach.

Rynasiewicz maintains that both the hole argument and Putnam's paradox have it that "given any complete description of a single state of affairs, there are hopelessly many distinct but indiscernible ways of construing the extensions of the descriptive terms of the language on the domain of discourse in question" (Rynasiewicz 1994, p. 419). In the case of the hole argument, the referents of manifold points in the theory are fixed. Furthermore, the extensions of predicates related to topological properties remain unchanged. However, quantities within the hole described by the metric and material fields are subject to a failure of determination under hole diffeomorphisms—and by identifying the diffeomorphisms with the permutations in Putnam's argument discussed above, there is a supposed parallel to Putnam's paradox.

But it's here that Pooley (2002) identifies a significant difference between Putnam's paradox and the hole argument—one which we endorse wholeheartedly. As Pooley (2002, p. 113) puts the point: "In Putnam's argument the permutation of the model's domain is used to reassign both extensions to predicates *and referents to names*, leaving the sentences held true fixed. In the hole argument, the diffeomorphism is used to reassign predicate extensions *without* reassigning the referents of names", which are points in spacetime. One particularly important consequence is that the models considered in the hole argument *prima facie* represent *distinct* physical possibilities, as opposed to the physically identical permutations postulated by Putnam's paradox. Figure 3 shows the effect of the hole diffeomorphism on models of the theory: because the referents of names and extensions of predicates are different in the two models, they appear to describe distinct physical possibilities.

Pooley makes further pertinent remarks regarding the notion of 'models' in GR. On the most standard understanding, models in the context of the hole argument refer to the solutions of the Einstein equation—and are thus pieces of mathematics.[12] This identification is standard in discussions of theories in the philosophy of physics, and as Pooley notes,

tural realism (on which see Ainsworth (2009)). Epistemic structural realism is the view that we can only learn about the *structure* of the unobservable world, but not its content. One approach *Ramseyfies* the sentences of a theory by replacing unobservable terms with existentially quantified variables. Newman's objection states that any set of objects with the right cardinality will fulfil such a theory, rendering the Ramseyfied theory unsatisfactory as a description of the world. However, Newman's objection applies only to unobservable terms within a Ramseyfied theory, while Putnam's paradox expands the objection beyond just Ramseyfied unobservable terms to cover all terms in a theory.

[12] As such, the notion precedes the act of interpretation, contrasting with views such as those expressed by Weisberg (2013), in which models are treated as *interpreted* structures.

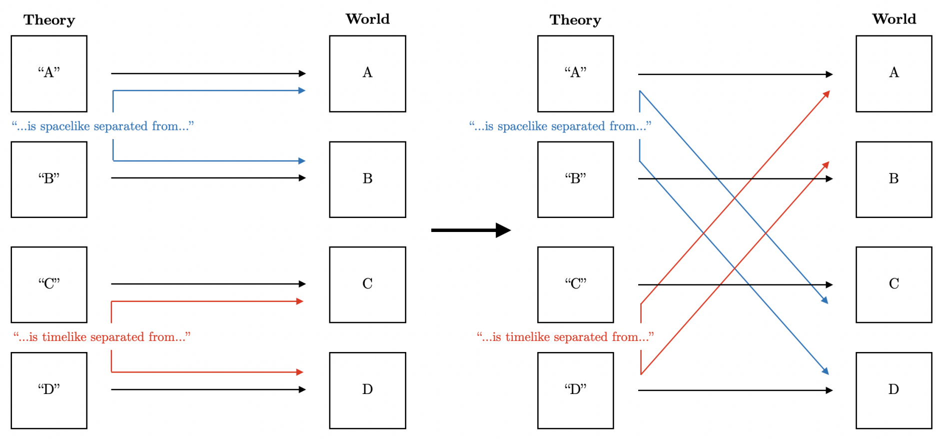

Figure 3: Permutations resulting from the hole argument.

results in models being descriptions of (potentially different) possible worlds. Addressing Rynasiewicz, who adopts this standard view of models, Pooley argues that manifold points must be understood as *bound variables* in order to ensure that diffeomorphism-related models are descriptions of the same world, resulting in purely qualitative descriptions of spacetimes.[13] A Putnam-type argument *is* possible on this way of understanding the models of GR, because one not only reassigns predicate extensions which describe the properties of the manifold points, but *also* reassigns the referents of manifold points, which are spacetime points (and this is mandatory given the understanding of manifold points as bound variables). By doing so, the truth values of the theory's sentences are preserved. But note that the hole argument is just different from this: in the hole argument, the predicates associated with spacetime points change within the hole, but the map between manifold points and physical spacetime points remains *fixed*. It is instead a property of GR *itself* which permits specific permutations in the form of hole diffeomorphisms on the same manifold.

The hole argument is most straightforwardly articulated by treating manifold points not as bound variables (as above) but rather as names. Then, hole diffeomorphic models can be understood as supplying descriptions of spacetimes which are qualitatively identical but physically distinct—for example, in one model, two spacetime points associated with the manifold points p and q are spacelike separated, while in another diffeomorphism-related model, the points are timelike separated. The fact that the referents of names are preserved across models suggests that the hole argument operates differently from the permutation argument, which permutes *all* such relationships within a given model. Therefore, on the standard understanding of models, the two arguments are inequivalent.

[13]Pooley comments that this reading is closely linked to the process of Ramseyfication described in footnote 11 (and discussed in the context of the hole argument by Maudlin (1988, p. 83)). In particular, a model of general relativity in which manifold points are understood in this way will represent a world which is qualitatively identical to that represented by some model in which quantifiers are removed and bound variables are replaced with names, more on which below.

An alternative, non-standard approach to the hole argument is taken by Liu (1996); on Pooley's reading of Liu (which we again endorse), models are "intermediaries between a "theory" and a possible world that makes the theory true" where the term "theory" is understood as the *description* of a single possibility (Pooley 2002, p. 112). Liu (1996, p. 249) states that a spacetime theory consists of names $\alpha, \beta, \ldots$, referring to spacetime points, and predicates $P_1, \ldots, P_s$, referring to the properties of these spacetime points. Given an interpretation $\mathbf{I}$ which maps the names to a "manifold of spacetime points" M, and provides extensions O_i to predicates, Liu denotes models m as tuples $m = \langle M, O_i \mid i = 1, \ldots, s\rangle$. Noting that hole diffeomorphisms h map m to $m' = \langle M, h^*O_i|i = 1, ..., s\rangle$, he claims that such mappings are "are nothing but 'C^∞ differentiable' permutations, clearly a subset of the permutations used in the Putnam theorem" (Liu 1996, p. 249). He concludes that the hole argument is not a problem about indeterminism between possible worlds, but instead one about the inscrutability of reference as in the case of Putnam's paradox.

To expand on Pooley's analysis, what Liu does is construe the hole argument in a non-standard fashion. To construct a direct analogue with Putnam's paradox, he reformulates the hole argument in first-order model-theoretic terms. What Liu calls 'models' differ from the models standardly referred to in discussions of the hole argument, and as we construed them above. Liu does in fact register this distinction: "We now see that in the original hole argument which supposedly engenders the radical indeterminism crisis, models as interpretations are confused with models as possible worlds or states of affairs. In the gauge theorem, models are model-theoretic entities which specify interpretations (or reference relations) between a theory and the items they refer; but that is not the way they are understood in the hole argument debate. There, they represent possible worlds (or states of affairs)" (Liu 1996, p. 250). But why this should be the case for Earman and Norton's gauge theorem (Earman and Norton 1987, p. 520), which clearly relies on the *standard* conception of models, is unclear, and seems to represent a distortion of their presentation of the hole argument.

As a separate point, recall that Liu characterises the manifold M as *directly* consisting of (rather than just representing) spacetime points.[14] While there is nothing manifestly wrong with such a conception, the question arises as to whether Liu's formulation in particular accurately depicts the hole argument when combined with his understanding of the gauge theorem, given that standard presentations of the hole argument focus on the relationship between manifold and spacetime points—something which appears to be bypassed entirely by Liu's discussion. In Liu's version of the hole argument, it is rather the *names* of spacetime points, α, β, ..., that are subject to a Putnam-style argument.

[14] This understanding of manifold points as being numerically identical with physical spacetime points also appears to be deployed by Butterfield (1989); this counts as a 'Lagadonian language' in the sense of Lewis (1986, p. 45), because objects are now their own names.

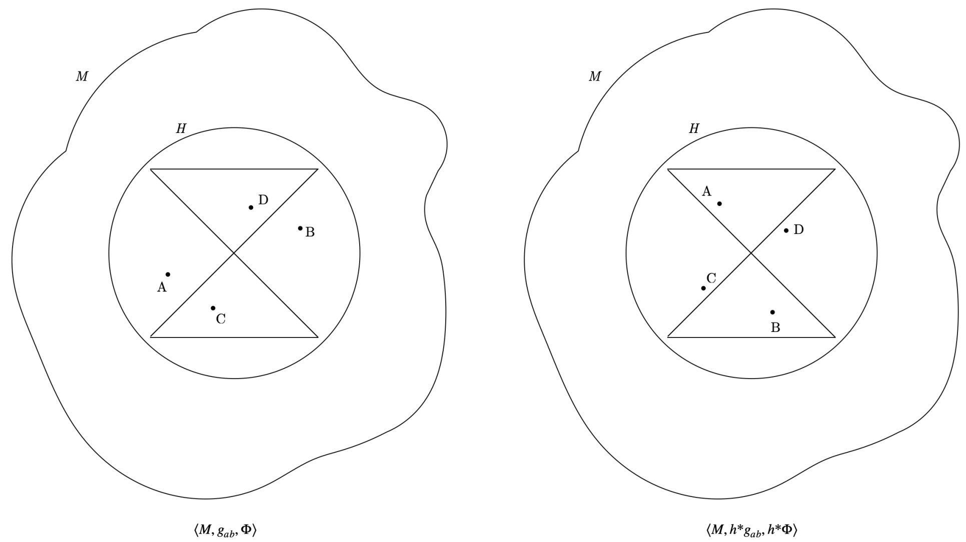

Figure 4: Two diffeomorphic models of spacetime.

We thus endorse Pooley's criticisms that this construal of models runs into the same problem as the standard formulation: "the sentences used to describe the scenarios contemplated in the hole argument typically come from sets that constitute *incompatible* descriptions" (Pooley 2002, p. 113). One might say that p and q are spacelike separated in one model but timelike separated in another; the two models do not correspond to equivalent descriptions, which Pooley argues is necessary to convert the hole argument into a case of Putnam's paradox.

Thus, the two arguments should be considered distinct. The hole argument changes the extensions of predicates within the theory, while maintaining the mapping from manifold points to spacetime points, which generates two *prima facie* distinct physical scenarios. In Figure 4, this corresponds to moving from the left to the right model through a hole diffeomorphism, precisely as described in §2. Putnam's paradox, on the other hand, changes both the extensions of predicates *and* the referents to names, resulting in different descriptions of the *same* physical scenario: the left model remains the relevant one even after the permutation.

One might maintain that the hole argument thus understood is nevertheless a special case of Putnam's model-theoretic argument, by identifying the identity map used to compare models of GR related by a hole diffeomorphism (what will be called 1_M below) as a trivial instance of permutation. However, there remains the same crucial disanalogy: to repeat, the key lesson of Putnam's paradox is that identical physical scenarios can be described by permuting simultaneously the names *and* extensions. On the other hand, it is the diffeomorphism invariance of GR that enables the hole argument to be constructed on a *given* manifold. It is true that both arguments rely on some form of permutation as a technical means to this goal, but as Pooley has argued, this is insufficient to reduce the hole argument to a special case of Putnam's paradox.

These differences between the hole argument and Putnam's paradox notwithstanding, there nevertheless remains an outstanding affinity between the two—one which will occupy us for the rest of this article. This affinity is that that both arguments have been responded to through meta-linguistic means. Through a comparison of these responses, we hope to emphasise why meta-linguistic responses to the hole argument are problematic, in a way which is directly related to its differences to Putnam's paradox.

5 Weatherall's response to the hole argument

While there have been numerous responses to the hole argument, often taking on certain metaphysical positions (again, see Norton et al. (2023) for a review), we focus now on Weatherall's proposed mathematical/formalist solution to the hole argument (Weatherall 2018).[15] Weatherall maintains that the hole argument originates from an improper interpretation of GR's mathematical formalism, and dissolves once the correct standard of sameness is applied to the models of GR in question. Weatherall (2018, p. 330) claims that the approach "is essentially neutral on the metaphysics of space and time", since (he claims) his solution through the mathematics of GR precedes any engagement with metaphysical considerations.

According to Weatherall, there are two relevant maps between models of GR involved in the hole argument: 1_M and ψ. First, $1_M : M \to M$ is an identity map, where all manifold points are mapped to themselves; ψ is a diffeomorphism similar to the hole diffeomorphism h in the original hole argument, and which witnesses the isometry $\tilde{\psi}$ of the two hole-diffeomorphic models under consideration. With respect to $\tilde{\psi}$, the metrics and therefore models do in fact agree, with $(g_{ab})|_p = (\tilde{g}_{ab})|_{\psi(p)}$ at any point on the manifold. Since (Weatherall argues) hole-diffeomorphic models must be compared using this map which witnesses the isometry of the models under consideration, the hole argument is thereby (he claims) blocked.[16]

Bradley and Weatherall (2022) develop this argument from Weatherall (2018). They state that "the hole argument, by invoking a privileged identification of spacetime points across models of general relativity, involves assertions, or invokes structure, that go beyond the theory of GR and should be viewed as representationally irrelevant" (Bradley and Weatherall 2022, p. 9). Bradley and Weatherall see 1_M as inducing exactly this kind

[15] Other mathematical/formalist responses to the hole argument include those of Mundy (1992) and Shulman (2017). Both of these, however, utilise different formalisms to standard GR: Mundy relies on a higher-order axiomatization of GR (or semi-Riemannian geometries more generally), and Shulman reformulates GR in the framework of Homotopy Type Theory with Univalent Foundations, also known as HoTT/UF. We return to this below. (For more on the background to such mathematical/formalist responses to the hole argument, see Bradley and Weatherall (2022).)

[16] For a much more detailed evaluation of this 'argument from mathematical structuralism'—also discussed below—see Pooley and Read (2025).

of structure which goes beyond the theory of GR, and therefore should be rejected as a legitimate standard of comparison for models of the theory. Thus one should consider Weatherall's solution as not just mathematical but also *meta-linguistic*, insofar as it rejects interpretations of mathematical structures that (supposedly) transcend the language of GR.[17]

6 Van Fraassen's response to Putnam's paradox

Next, we consider van Fraassen's attempt at dissolving Putnam's model-theoretic argument, which rests on meta-linguistic arguments similar to those seen above from Bradley and Weatherall (2022) in the context of the hole argument. Van Fraassen's argument is based on Putnam's description of the act of choosing the mapping between the terms of the theory and the objects in the world. Van Fraassen claims that this procedure is "couched in the discourse of physical manipulation", raising the question of how it operates with reference to our own capabilities (van Fraassen 2008, p. 20).

To establish the relationship between one's theory and the world, Putnam instructs: "Pick a model M of the same cardinality as The World. Map the individuals of M one-to-one into the pieces of The World, and use the mapping to define relations of M directly in The World" (Putnam 1977, p. 126). Van Fraassen argues that the act of selecting such a function is not possible for our *own* language. In order to make this choice, he maintains that we need "an independent description of both the domain and range of an interpretation", such that it is possible to construct meaningful connections between the terms in the theory and referents in the world (van Fraassen 2008, p. 234).

For an arbitrary language other than our own, one uses our own language to describe "an assignment of extensions to its singular and general terms", in the form of a function (van Fraassen 2008, p. 233). However, when considering theories constructed from our own language, we can *only* use the language of the theory itself to describe the world, and therefore cannot identify the requisite function between the theory and the world. The key point is that "we can grasp an interpretation—i.e. function linking words to parts of The World—only if we can identify and describe that function. But we cannot do that unless we can independently describe The World" (van Fraassen 2008, p. 235). If such an independent description were in fact possible, then we could be certain of the extensions to the predicates in our theory, preventing Putnam's paradox from arising.

Van Fraassen claims that the paradox is therefore dissolved: no indeterminism of reference arises, because we only have *one* choice of how our terms refer to objects in the world—the one we have. This solution "does not transpose to those anthropologists studying recordings of an alien language", because we would be able to identify the func-

[17] Cf. Cudek (forthcoming).

tion using our own language, and hence provide a genuine interpretation of the theory in question (van Fraassen 1997, p. 21). Nonetheless, van Fraassen argues that this issue simply cannot arise when considering theories written in our own language, where discussion of such a function is meaningless.[18]

Bradley and Weatherall make a largely parallel claim in the context of discussing Rynasiewicz's views connecting Putnam's paradox and the hole argument. They argue that "the advocate for the mathematical response would presumably deny that there is any problem of reference in the first place, at least for theories that one takes to fully characterise their subject matter. After all, to generate the problem for any given theory, one must move from the formal theory under consideration to the metatheory. And the metatheory is representationally irrelevant" (Bradley and Weatherall 2022, p. 14).

Just as GR according to Bradley and Weatherall (2022) disallows the use of further metatheory for making comparisons between models, van Fraassen claims that we cannot use a language independent of our own to map between terms of our theory and objects in the world. Note, however, that Bradley and Weatherall claim such metatheory is *irrelevant*, not *impossible* in the way that van Fraassen claims with regard to finding a function for connecting our own language to the world. We expand on this difference (and others) in the next section.

7 Evaluation

Before returning to the above issues, in this section we first explore some other differences between van Fraassen (1997) on Putnam's paradox and Bradley and Weatherall (2022) on the hole argument. To this end, following the lead of Pooley and Read (2025), first recall that we can isolate two arguments from Weatherall (2018) against the validity of the hole argument: (i) the *equivocation argument*, which maintains that the hole argument relies on an illegitimate equivocation between two maps used to compare isometric manifolds: the identity map 1_M, and the map which witnesses the isometry $\tilde{\psi}$; and (ii) the *argument from mathematical structuralism*, which has already been presented above, and which concludes that it is only legitimate to compare said pair of isometric manifolds using $\tilde{\psi}$.

With these two distinct arguments in mind, note first that an analogue to the equivocation argument cannot be applied to Putnam's paradox.[19] In that case, there simply

[18]Helpful for understanding this argument is van Fraassen's account of scientific representation, which is explicitly indexical and relativised to human capabilities: "*For us* the claim (A) that the theory is adequate to the phenomena and the claim (B) that it is adequate to the phenomena as represented, that is, *as represented by us*, are indeed the same!" (van Fraassen 2006, p. 545). Van Fraassen sees this as an example of a 'pragmatic tautology', which is a logically contingent but practically undeniable statement. Practically speaking, therefore, we have no choice but to use the language that we have to represent the world. (To our knowledge, the only other author to bring these aspects of van Fraassen into contact with the hole argument—albeit without explicit reference to Putnam's paradox—is Landsman (2023, §5).)

[19]For much more detailed critical assessment of the equivocation argument, see Pooley and Read

is no pair of mappings between models between which one might equivocate; note that these are not the same as the maps that van Fraassen denies the possibility of identifying, which are maps between models and the world, rather than inter-model mappings.

On the other hand, Weatherall's argument from mathematical structuralism is strategically similar to van Fraassen's solution to Putnam's paradox, where methodological and meta-linguistic considerations supposedly prevent the problem from arising in the first place. While in Weatherall's case, the correct language of GR supposedly bars the interpretation of isomorphic models as inequivalent, van Fraassen argues that the language we use prevents us from grasping the function which would pick an alternative referential theory. The lack of linguistic machinery in both cases is meant to prevent the possibility that our theories refer to a radically different possible world than its intended target.

As we see it, there are two main possible lines of objection to Weatherall's strategy. The first, as Pooley and Read (2025) have argued, is to grant that to use 1_M and other set-theoretic resources is to step outside of 'the language of GR' (i.e., the language of Lorentzian manifolds), but to maintain that such a meta-linguistic dissolution does not preclude the countenancing of a plurality of metaphysical possibilities, which are what lead to a problematic failure of determination: for more on this, see Pooley and Read (2025, §5). GR, defined narrowly as a theory about Lorentzian manifolds, might not have the linguistic resources to draw distinctions between possibilities which would generate the hole argument. However, GR is of course not a complete theory of nature, and besides (and partly for that very reason), it is not understood in real scientific practice in such a narrow sense.[20]

This brings us to our second possible line of objection to the strategy of Bradley and Weatherall (2022): to *deny* that set-theoretic resources such as 1_M are not part of language of GR (i.e., to maintain that they *are* part of the language of that theory, rather than merely its meta-language, as Bradley and Weatherall (2022) maintain quite explicitly when they write that "one can also step outside of the theory and, by invoking the further expressive resources of the "semantic metalanguage"—basically, Zermelo–Fraenkel (ZFC) set theory—one *can* express differences between set-theoretic representations of models of the theory" (Bradley and Weatherall 2022, p. 1266)). After all, on standard set-theoretic foundations for mathematics, Lorentzian manifolds *just are* structured sets, in which case set-theoretic resources are of course available, and the argument from mathematical structuralism begins to look more akin to a sociological stricture than to a mathematical one. While one could maintain that maps such as 1_M are indeed part of the

(2025).

[20]Now, Bradley and Weatherall (2022) might respond that none of the above is relevant to the issue of determinism *according to GR*. To some extent, we could be persuaded to agree—but on the other hand, as Cudek (forthcoming, §5), Gomes (2024), Gomes and Butterfield (2023, §3.2), and Landsman (2023, §1) argue, the narrow construal of GR to which Bradley and Weatherall (2022) appeal is inauthentic to how the theory is used in physics.

meta-language by moving to alternative foundations for mathematics (perhaps—but not necessarily—HoTT/UV, on which see also Dougherty (2020) and Ladyman and Presnell (2020)), Bradley and Weatherall (2022) explicitly do *not* make this move, in which case the position seems to rest upon a form of mathematical structuralism which is at least controversial—see Reck and Schiemer (2023).

So: it is not obvious that resources such as 1_M are part of the meta-language rather than language of GR, and even granting this, it is not obvious that one cannot avail oneself of them in the physical practice of GR. Neither of these questions arise in the case of van Fraassen (1997) on Putnam's paradox, where the resources needed to generate the problem certainly *are* part of the meta-language, and where—given the global nature of the problem—those resources certainly are *not* available. One further point of comparison of potential interest here is with Putnam (2000) on philosophical scepticism: the claim that there is no problem of philosophical scepticism for brains-in-vats because the *total language* of those brains-in-vats could never be used to articulate the problem might appear to be akin to van Fraassen (1997) on Putnam's paradox. However, if there are extra-vat linguistic resources of which those brains-in-vats can in principle avail themselves, then their insistence that such resources cannot be used looks akin to Bradley and Weatherall (2022) on the hole argument, and in turn looks correspondingly less convincing.[21]

This does, however, allow for changing the foundations of one's mathematics: see, for example, other formalist responses to the hole argument mentioned by Bradley and Weatherall, namely those due to Mundy (1992) and Shulman (2017). We concur with Gajic (2024) that switching to a different foundation of mathematics looks to be a more effective strategy for defusing the hole argument than what's offered by Weatherall (2018) and Bradley and Weatherall (2022), because (the thought—or at least aspiration—goes) isometric manifolds literally *are* the same manifold under suitable alternative foundations. So our criticisms are restricted to approaches with set-theoretic foundations, ZFC or otherwise—which was almost always the context for discussions of the hole argument in any case.

On the basis of set-theoretic foundations, there are further linguistic resources both available and necessary in the context of GR as a working, non-complete theory. This is not the case for van Fraassen's dissolution of Putnam's paradox. Precisely because of the generality of Putnam's paradox, the language to which van Fraassen refers is the *total* language of a particular community. Only by employing this strategy can Putnam's paradox be blocked on linguistic grounds: there is simply *no* other language to be used to form the permutation argument. This reinforces our above point in its appeal to a close attention to our use of language: because the *use* of GR involves reference to language

[21]For more on the analogy between the hole argument and Putnam on philosophical scepticism, see Cheng and Read (2022, §4.3).

(supposedly) beyond that of Lorentzian manifolds, Bradley and Weatherall's objection to the hole argument flounders.

The fact that the hole argument reemerges once one moves beyond the formalism of GR (granting that one need do this at all) initially suggests a flaw in van Fraassen's solution—one which is also revealed via the comparison with Putnam on scepticism. Van Fraassen's argument appears not to preclude the possibility of constructing a new language for the purpose of describing the function between terms in our current language and physical objects. This seems to recover the case of describing an alien language with our own language, such that the necessary function is no longer impossible to find. This meta-language would then be vulnerable to Putnam's paradox. Van Fraassen might argue, like Weatherall, that such meta-languages are *irrelevant*; but this claim is harder to maintain against the generality of Putnam's paradox.

However, van Fraassen has a better counterargument available: by once again claiming that this new meta-language cannot be subject to Putnam's paradox due to the lack of an independent language to describe the relevant function with, this criticism of van Fraassen has to resort to a *further* new meta-language. The result is an infinite regress, which is problematic given that the act of language and theory construction is limited by human capabilities.[22] Why this response might be less useful to Bradley and Weatherall is because the hole argument is a local rather than global issue like Putnam's paradox. The additional language of names (or otherwise) is an identifiable and constrained meta-language that is often found in the use of the theory; it does not require an infinite regress to further justify it.

To reiterate: while the hole argument and Putnam's paradox are structurally disanalogous, we have chosen to compare Bradley and Weatherall's response to the hole argument and van Fraassen's dissolution of Putnam's paradox because both rely on meta-linguistic strategies that aim to preclude the appeal to further meta-languages. However, it is precisely *because* of the greater generality of Putnam's paradox that we find van Fraassen's argument more persuasive.

8 Conclusion

Though the hole argument and Putnam's paradox are superficially similar, they are ultimately structurally different: Putnam's paradox is a general claim about the description of a single physical possibility within an arbitrary theory, while the hole argument relies on a particular feature of GR to link apparently distinct physical possibilities. Weather-

[22]van Fraassen (2010, p. 471) further criticises this Tarskian idea of an infinite hierarchy of languages and meta-languages; he claims that, while we can isolate proper parts of our language-in-use and construct a meta-language to describe these parts, we cannot just arbitrarily construct meta-languages so that we have "tools more precise than any tools we have, to determine the limits of precision of what we have."

all's solution to the hole argument fails to block the emergence of metaphysical plurality, insofar as there is the possibility of appealing to further language (even granting that GR itself does not have the linguistic resources to generate the problem, which is questionable). By contrast, van Fraassen's (1997) dissolution of Putnam's paradox argues that within our own language, we cannot identify a function which helps map our theoretical terms to objects in the world, so that no indeterminism between mappings arises; given its global nature, this argument appears more successful than that of Bradley and Weatherall (2022).

Acknowledgments

We're grateful to Clara Bradley, Jeremy Butterfield, Franciszek Cudek, Henrique Gomes, James Ladyman, Eleanor March, Oliver Pooley, and Jim Weatherall for helpful discussions and suggestions.

References

Ainsworth, Peter M. (2009). "Newman's Objection". In: *British Journal for the Philosophy of Science* 60.1, pp. 135–171.

Bradley, Clara and James Owen Weatherall (2022). "Mathematical Responses to the Hole Argument: Then and Now". In: *Philosophy of Science* 89.5, pp. 1223–1232.

Butterfield, Jeremy (1989). "The Hole Truth". In: *British Journal for the Philosophy of Science* 40.1, pp. 1–28.

Button, Tim (2013). *The Limits of Realism*. Oxford University Press UK.

Cheng, Bryan and James Read (2022). "Shifts and reference". In: The Foundations of Spacetime Physics: Philosophical Perspectives. New York: Routledge.

Cudek, Franciszek (forthcoming). "Counterparts, Determinism, and the Hole Argument". In: *British Journal for the Philosophy of Science*.

Dougherty, John (2020). "The Hole Argument, Take N". In: *Foundations of Physics* 50.4, pp. 330–347.

Earman, John and John D. Norton (1987). "What Price Spacetime Substantivalism? The Hole Story". In: *British Journal for the Philosophy of Science* 38.4, pp. 515–525.

Gajic, Gregor (2024). "On the Hole Argument under Nonstandard Approaches to the Foundations of Mathematics". MA thesis. University of Oxford.

Gomes, Henrique (2024). *The hole argument meets Noether's theorem*. arXiv: `2403.10970 [physics.hist-ph]`. URL: `https://arxiv.org/abs/2403.10970`.

Gomes, Henrique and Jeremy Butterfield (June 2023). "The Hole Argument and Beyond: Part I: The Story so Far". In: *Journal of Physics: Conference Series* 2533.1, p. 012002. URL: `https://dx.doi.org/10.1088/1742-6596/2533/1/012002`.

Ladyman, James and Stuart Presnell (2020). “The Hole Argument in Homotopy Type Theory”. In: *Foundations of Physics* 50.4, pp. 319–329.

Landsman, Klaas (2023). “Reopening the Hole Argument”. In: *Philosophy of Physics.*

Lewis, David (1984). “Putnam’s Paradox”. In: *Australasian Journal of Philosophy* 62.3, pp. 221–236.

— (1986). *On the Plurality of Worlds.* Malden, Mass.: Wiley-Blackwell.

Liu, Chuang (1996). “Realism and Spacetime: Of Arguments Against Metaphysical Realism and Manifold Realism”. In: *Philosophia Naturalis* 33.2, pp. 243–63.

Maudlin, Tim (1988). “The Essence of Space-Time”. In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1988, pp. 82–91.

Menon, Tushar and James Read (2023). “Some Remarks on Recent Formalist Responses to the Hole Argument”. In: *Foundations of Physics* 54.1, pp. 1–20.

Mundy, Brent (1992). “Space-Time and Isomorphism”. In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992.Volume One: Contributed Papers, pp. 515–527.

Norton, John D., Oliver Pooley, and James Read (2023). “The Hole Argument”. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Pooley, Oliver (2002). “The Reality of Spacetime”. PhD thesis. University of Oxford.

— (2013). “Substantivalist and Relationalist Approaches to Spacetime”. In: *The Oxford Handbook of Philosophy of Physics.* Ed. by Robert W. Batterman. Oxford University Press USA.

Pooley, Oliver and James Read (2025). “On the Mathematics and Metaphysics of the Hole Argument”. In: *The British Journal for the Philosophy of Science.*

Putnam, Hilary (1977). “Realism and Reason”. In: *Proceedings and Addresses of the American Philosophical Association* 50.6, pp. 483–498.

— (1980). “Models and Reality”. In: *Journal of Symbolic Logic* 45.3, pp. 464–482.

— (1981). *A Problem About Reference.* Cambridge University Press.

— (2000). “Brains in a Vat”. In: *Knowledge: readings in contemporary epistemology.* Ed. by Sven Bernecker and Fred I. Dretske. Oxford University Press, pp. 1–21.

Quine, W. V. (1968). “Ontological Relativity”. In: *Journal of Philosophy* 65.7, pp. 185–212.

Reck, Erich and Georg Schiemer (2023). “Structuralism in the Philosophy of Mathematics”. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta and Uri Nodelman. Spring 2023. Metaphysics Research Lab, Stanford University.

Rynasiewicz, Robert (1994). “The Lessons of the Hole Argument”. In: *British Journal for the Philosophy of Science* 45.2, pp. 407–436.

Shulman, Michael (2017). “Homotopy Type Theory: A Synthetic Approach to Higher Equalities”. In: Categories for the Working Philosopher. Oxford: Oxford University Press.

Stachel, John (2014). “The Hole Argument and Some Physical and Philosophical Implications”. In: *Living Reviews in Relativity* 17.1, p. 1. URL: https://doi.org/10.12942/lrr-2014-1.

Van Fraassen, Bas C. (1997). “Putnam’s Paradox: Metaphysical Realism Revamped and Evaded”. In: *Philosophical Perspectives* 11, pp. 17–42.

— (2006). “Representation: The Problem for Structuralism”. In: *Philosophy of Science* 73.5, pp. 536–547. URL: https://www.cambridge.org/core/product/2426010094DDB08407BABAE1B8D655D4.

— (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford University Press UK.

— (2010). “Reply to Belot, Elgin, and Horsten”. In: *Philosophical Studies* 150.3, pp. 461–472.

Weatherall, James Owen (2018). “Regarding the ’Hole Argument’”. In: *British Journal for the Philosophy of Science* 69.2, pp. 329–350.

Weisberg, Michael (2013). *Simulation and Similarity: Using Models to Understand the World*. New York, US: Oxford University Press.