The Interpretation of String-Theoretic Dualities

James Read

Merton College, University of Oxford, OX1 4JD, UK

james.read@merton.ox.ac.uk

Abstract

We analyse the possibility that string-theoretic dualities present a genuine case of strong underdetermination of theory by evidence. Drawing on the parallel discussion of the hole argument, we assess the possible interpretations of dualities. We conclude that there exist at least two defensible interpretations on which dualities do not present a worrying case of underdetermination *per se*.

Contents

1

1 Introduction

"I dreamt of a... nightmarish café, brilliantly lit, but underground, with no way out. I'd been dead a long, long time. The waitresses all had the same face. The food was soap, the only drink was cups of lather. The music in the café was... this." [29]

Within contemporary philosophy of science, scientific realists are sometimes charged with accounting for the possibility of *strong underdetermination of theory by evidence*, i.e. the possibility of the existence of multiple competing scientific theories which make different ontological claims about the world, yet which are all empirically adequate. Such cases are supposed to be problematic for the realist, as there exists no principled means of determining *which* theory should be endorsed as correctly describing the world. Nevertheless, in the absence of any concrete and convincing cases of this phenomenon, the realist has often made light of the problem: if it is to be taking seriously, the burden is on her opponents to provide convincing examples.

In fact though, the dearth of cases of strong underdetermination may be reaching an end. This is in light of the phenomenon of *duality*, which has recently arisen in the context of string theory. Roughly, two theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ are *dual* when there is an exact equivalence (or, to use the terminology of [49], a "dictionary") between their models respecting all correlation functions, with both theories therefore making identical empirical predictions [42]. As stated in [1], dualities in quantum field theory and string theory are often very intricate: $\mathcal{T}$ and $\tilde{\mathcal{T}}$ generally have different microscopic degrees of freedom, and different actions. Having said that, we will also see that there are some cases of *self-duality*, where $\mathcal{T}$ and $\tilde{\mathcal{T}}$ turn out to be the very same theory.

One of the most characteristic features of string theory is the preponderance of such dualities. The existence of such ontologically distinct yet empirically equivalent theories can be read *prima facie* as a genuine case of strong underdetermination. Faced with this, many questions naturally arise. For example: do two respective models of $\mathcal{T}$ and $\tilde{\mathcal{T}}$ really describe two different possible worlds, or do they in fact merely describe the same world? Or: does there exist a principled means by which one can consider one of $\mathcal{T}$ and $\tilde{\mathcal{T}}$ to be more fundamental

than the other, and thereby to preference its ontological claims?[1] Faced with such questions, it is incumbent on us to investigate the appropriate interpretation of dualities.

When assessing possible interpretations of dualities, it is helpful to make sure that we have fully explored the space of possible responses. To this end, it can be useful to draw from other debates in the philosophy of science and philosophy of physics. Recently, Rickles has suggested an analogy between the infamous *hole argument* in General Relativity and dualities [43]. In the case of the hole argument, we are supposed to face a troubling indeterminism at the heart of General Relativity; our task is either to find some way of interpreting the models of the theory in order to avoid this problem, or to find some means of accepting the indeterminism while diffusing its potency. Analogously, in the case of dualities, we must find some way of interpreting the dual theories such that either only one set of ontological claims is privileged, or an explanation is given as to how the existence of multiple models of theories making distinct ontological yet empirically equivalent claims is unproblematic.

Drawing on the responses to the hole argument, we find that there are four possible classes of response to string-theoretic dualities. First, one can accept this as a genuine case of underdetermination, and then try to argue either that (a) the underdetermination is unproblematic, (b) the two dual theories are complementary and their ontological claims are reconcilable, (c) the two theories are better interpreted as being embedded in a "deeper" theory, which makes a *unique* set of ontological claims. Second, one can argue that two models of the dual theories in fact only describe one possible world; any difference between the two is unphysical and can be "quotiented out", in a manner analogous to gauge redundancies in e.g. Yang-Mills theory. Third, one can try to interpret the manifolds on which the dual theories are defined in a way that removes the ontological underdetermination. Fourth, one can try to privilege the ontological claims of the models of just one of the theories in a dual pair. In fact, we will find that, by combining the above options, there are two *prima facie* viable routes available to us which diffuse the underdetermination.

The structure of this paper is as follows. First, in section 2 we provide a precise characterisation of the differences between symmetries, gauge redundancies, and dualities; this will be useful in what follows. In sections 3 and 4, we provide a brief summary of the relevant technical details pertaining to string-theoretic dualities and the hole argument, respectively. In

---

[1]Such claims are often made in the context of the so-called AdS/CFT correspondence, for example.

section 5, we undertake an in-depth analysis of the possible interpretations of string dualities. We close in section 6 with an assessment of these interpretations, and some outlook for the future.

2   Symmetries, Gauge Redundancies and Dualities

Before we begin our discussion of string dualities, we must make precise the distinction between *symmetries*, *gauge redundancies*, and *dualities*. In order to do this, we first follow Healey [15] in defining *empirical symmetries* and *theoretical symmetries*:

Definition 1. (Empirical symmetry) *A mapping $\phi : \mathcal{S} \to \mathcal{S}$ of a set of situations onto itself is an* empirical symmetry *if and only if no two situations related by ϕ can be distinguished by means of measurements confined to each situation.*

Definition 2. (Theoretical symmetry) *A mapping $f : \Lambda \to \Lambda$ of a set Λ of models of a theory Θ onto itself is a* theoretical symmetry of Θ *if and only if for every model $\mathcal{M} \in \Lambda$ of Θ that may be used to represent a situation $\mathcal{S}$ in a possible world w, $f(\mathcal{M})$ may also be used to represent $\mathcal{S}$ in w.*

Here, *situation* is taken to mean something akin to *configuration of a physical system*. With this in mind, in [15] Healey defends the thesis that the distinction between symmetry transformations (simpliciter) and gauge transformations (gauge redundancies) can be made as follows:

Definition 3. (Symmetry transformation) *A theoretical symmetry of a theory with a corresponding non-trivial empirical symmetry among the situations it represents.*

Definition 4. (Gauge transformation) *A theoretical symmetry of a theory with no corresponding non-trivial empirical symmetry among the situations it represents.*

Such a distinction seems to be a good explication of the *de facto* usage of these terms in the relevant physics literature. Symmetries (simpliciter) are structure-preserving transformations which relate to one another solutions of the equations of motion of some theory (or a physical system described by the theory); in each case the application of the symmetry to a solution

5

leads to another solution representing a distinct physically possible state of the system. Unlike symmetries, gauge transformations do not map between physically distinct possibilities, but rather map one representation of a state to another representation of the same state. In fact, *physical state* is really shorthand for *equivalence class of states under gauge symmetries*, so that physical states are represented by entire gauge orbits rather than their elements [42].

In contrast to symmetries and gauge transformations, there exists a class of transformations relating physical *theories*, rather than states within a theory; these are known as *dualities*. Roughly, two theories are dual whenever they determine the same physics; i.e. the same correlation functions, physical spectra, etc. By analogy with symmetries, one is faced with a space whose elements are theories, as opposed to states or configurations, so that dualities map one theory onto another in a way that preserves all 'physical' predictions [42, 51].

To be more precise, consider a theory $\mathcal{T}$, characterised by a number of parameters λ_i. The space of the λ_i is denoted $\mathcal{U}$, the so called *parameter space* or *moduli space* of the theory.[2] Fixing all the λ_i fixes a specific model $\mathcal{M}$ of $\mathcal{T}$, corresponding to a specific point $u \in \mathcal{U}$.[3] Now suppose that there exists a second theory $\tilde{\mathcal{T}}$, with parameters $\tilde{\lambda}_i$ and parameter space $\tilde{\mathcal{U}}$. We can construct a definition of duality as follows:

Definition 5. (Duality) *Theory $\mathcal{T}$ is* dual *to theory $\tilde{\mathcal{T}}$ if and only if for every $u \in \mathcal{U}$ with associated model $\mathcal{M}$, there exists an associated $\tilde{u} \in \tilde{\mathcal{U}}$ with model $\tilde{\mathcal{M}}$ such that $\mathcal{M}$ and $\tilde{\mathcal{M}}$ are empirically equivalent. This duality map is symmetric and one-to-one, and so amounts to an isomorphism between $\mathcal{U}$ and $\tilde{\mathcal{U}}$.*

Typically, the duality isomorphism involves pairs of models of theories with radically different structural properties, such that they would appear to be wholly unrelated. We can distinguish between two kinds of duality: those that relate models of pairs of *distinct* theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$, and those that relate (apparently quite distinct) models of one and the same theory $\mathcal{T}$. We call these *external duality* and *self-duality*, respectively. As Rickles states [42], since self-dualities state an equivalence between descriptions of one and the same system, they are often construed as gauge symmetries (though of a rather surprising nature), representing some interpretative ambiguity in the theory's formulation. We shall return to this point shortly.

---

[2]The λ_i are not free, but are in fact highly constrained. For example, the relevant λ_i parameterising the moduli space of vacua of a supersymmetric gauge theory often form an *affine algebraic variety*.

[3]This space might therefore be identified with the space of *kinematically possible models*, defined in e.g. [38].

Finally, it is important to keep clear the distinction between *theories*, *models*, and *possible worlds*. We denote such objects by $\mathcal{T}$, $\mathcal{M}$, and W, respectively. The apparent underdetermination in the case of dualities can be expressed as follows: we have two theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$, with respective models $\mathcal{M}$ and $\tilde{\mathcal{M}}$. Both these models are empirically adequate to the actual world[4], yet appear to represent *distinct* possible worlds W_1 and W_2. Hence, we face a problem of determining which world of W_1 and W_2 is the actual world. In the case where $\mathcal{T}$ and $\tilde{\mathcal{T}}$ coincide (i.e. in the case of self-dualities), we *still* have two distinct models corresponding to two possible worlds, and so *still* face a *prima facie* case of underdetermination.

3 String Theory and its Dualities

With precise distinctions between symmetries, gauge redundancies, and dualities in hand, we can now move on to consider dualities in string theory. In this section, we first provide a very brief *précis* of perturbative string theory, before proceeding to discuss T-duality, mirror symmetry, S-duality, and the AdS/CFT correspondence.

3.1 Perturbative String Theory

A string can be regarded as a special case of a p-brane, which is an object with p spatial dimensions and tension $T_p = 1/(2\pi\alpha')$, where α' is the *Regge slope parameter*. In fact, various p-branes do appear in superstring theory as non-perturbative excitations. The classical motion of a p-brane extremises the $(p+1)$-dimensional volume V that it sweeps out in space-time. Thus there is a p-brane action that is given by $S_p = -T_p V$. In the case of the fundamental string, which has $p = 1$, V is the area of the string world sheet and the action is called the *Nambu-Goto action* [50]. Classically, the Nambu-Goto action is equivalent to the string

---

[4]Throughout this paper, we consider a hypothetical scenario in which the dual theories under consideration *do* give realistic possible models of the world. Examples such as that of the AdS/CFT correspondence presented in section 3 do not relate empirically viable theories.

sigma-model action (also known as the *Polyakov action*):

$$S_\sigma = -\frac{1}{4\pi\alpha'} \int_\Sigma d^2\sigma \sqrt{-h} h^{\alpha\beta} \eta_{\mu\nu} \partial_\alpha X^\mu \partial_\beta X^\nu \tag{3.1}$$

where σ and τ are coordinates on the world sheet, $h_{\alpha\beta}(\sigma,\tau)$ is an auxiliary[5] world sheet metric, $h = \det h_{\alpha\beta}$, and $h^{\alpha\beta}$ is the inverse of $h_{\alpha\beta}$. Σ denotes the world sheet, and $d^2\sigma = d\sigma d\tau$. The functions $X^\mu(\sigma,\tau)$ describe the spacetime embedding of the string world sheet. The Euler-Lagrange equation for $h^{\alpha\beta}$ can be used to eliminate it from the action and recover the Nambu-Goto action. Quantum mechanically, instead of eliminating h via its classical field equations, one should perform a path integral, using standard machinery to deal with the local symmetries and gauge fixing. Doing this, one finds that there is a conformal anomaly unless the space-time dimension is $D = 26$. An analogous analysis for superstrings (i.e. strings for which supersymmetry is added - either on the world sheet as in the so-called *RNS sector*, or to the background spacetime as in the *GS sector*) gives the critical dimension $D = 10$ [1].

When one uses the superstring formalism for both left-moving modes and right-moving modes on the string, the fermions associated with the left-movers and the right-movers can have either the same or opposite chirality. These two possibilities give different theories called the type IIA and type IIB superstring theories, respectively. A third possibility, called type I superstring theory, can be derived from the type IIB theory by quotienting out by its left-right symmetry, a procedure called *orientifold projection*. The strings that survive this projection are unoriented. A more intriguing possibility is to use the formalism of the 26-dimensional bosonic string for the left-movers and the formalism of the 10-dimensional superstring for the right-movers. The string theories constructed in this way are called *heterotic*. The extra sixteen left-moving dimensions must describe a torus with very special properties to give a consistent theory. There are precisely two distinct tori that have the required properties, and they correspond to the Lie algebras of $SO(32)$ and $E_8 \times E_8$. Thus, there are five consistent superstring theories: types I, IIA/B, and heterotic $SO(32)$ or $E_8 \times E_8$.

The parameter τ in the embedding functions $X_\mu(\sigma,\tau)$ is the world sheet time coordinate; σ parametrizes the string at a given world sheet time. For a closed string, which is topologically a

---

[5]In the sense that $h_{\alpha\beta}$ is a *new* variable, *a priori* independent of the pullback of the spacetime metric to the world sheet.

circle, one must impose periodicity in the spatial parameter σ. Choosing its range to be π, one identifies both ends of the string $X_\mu (\sigma, \tau) = X_\mu (\sigma + \pi, \tau)$. All string theories contain closed strings, and the graviton always appears as a massless mode in the closed-string spectrum of string theories in the critical dimension.

We can see the appearance of the graviton in the string spectrum as follows. After quantising the string and defining suitable ladder operators, one can act on the ground state of the string with raising operators to study its spectrum. One finds that for the closed bosonic string and the RNS type II closed superstring in the so-called *NS-NS sector* (which is one way of imposing boundary conditions on worldsheet fermions), the first excited states correspond to g_{ij}, B_{ij}, and Φ. g_{ij} is symmetric and traceless in i and j and transforms under $SO(D-2)$ as a massless, spin-two particle, the *graviton* (here, D is the critical dimension of the string). B_{ij} transforms under $SO(D-2)$ as an antisymmetric, second-rank tensor, with associated three-form field strength H_{ijk}. The trace term Φ is a massless scalar, which is called the *dilation*. We can add (for reasons that will become clear shortly) the dilation field to the string action as a background field in a term of the form:

$$S_\Phi = \frac{1}{4\pi} \int_\Sigma \Phi(X) R^{(2)}(h) \, d^2 z \tag{3.2}$$

where $R^{(2)}(h)$ is the scalar curvature of the two-dimensional string world sheet computed from the world-sheet metric $h^{\alpha\beta}$. The dilaton plays a crucial role in defining the string perturbation expansion. This can be best understood by considering the case in which Φ is a constant. (More generally, if Φ approaches a constant at infinity, it is possible to separate this constant mode from the rest of Φ and focus on its contribution. [1]) The key observation is that, when Φ is a constant, the integrand in (3.2) is a total derivative. This means that the value of the integral is determined by the global topology of the world sheet, and this term does not contribute to the classical field equations. The topological invariant that arises here is the famous *Euler characteristic* of Σ:

$$\chi(M) = \frac{1}{4\pi} \int_\Sigma R^{(2)}(h) \, d^2 z \tag{3.3}$$

Hence we see that if $\Phi = \Phi_0$ for Φ_0 a constant, then $S_\Phi = \Phi_0 \chi$. Now, when calculating

string scattering amplitudes, one must (by extension of summing over all possible Feynman diagram topologies in quantum field theory) sum over all possible world sheet topologies. How should we weight this sum over different topologies? To answer this question, suppose that $\Phi = \Phi_0 + \phi(X)$, where $\phi(X)$ is zero in the vacuum, i.e. Φ_0 is the vacuum expectation value of $\Phi(X)$. Then there will appear a factor in the path integral of the form $e^{-\Phi_0 \chi} = g_s^{2(g-1)}$, with $g_s \equiv e^{\Phi_0}$ and g the genus of Σ. In other words, the vacuum expectation value of the dilation field determines the string coupling constant! This is a special case of a more general lesson in string theory: all dimensionless numbers (e.g. coupling constants) that characterise specific string vacua are determined as the vacuum expectation values of scalar fields. [1]

As a point of interest, note that the dilaton field acts like a Brans-Dicke scalar. In fact, the effective gravity equations in string theory include a graviton-dilaton part that looks very similar to the Brans-Dicke scalar-tensor theory of gravity.

3.2 T-Duality

As discussed, the massless fields that appear in the closed bosonic string spectrum or the NS-NS sector of either of the type II superstrings consist of the spacetime metric $g_{\mu\nu}$, the two-form $B_{\mu\nu}$, and the dilation Φ. So far, we have only considered a flat background with vanishing $B_{\mu\nu}$. One can analyse more general possibilities by introducing the background fields into the world sheet action. The appropriate generalisation of the world sheet action that includes the above background fields[6] is: [2]

$$S = -\frac{1}{4\pi\alpha'} \int_\Sigma d^2\sigma \sqrt{-h} \Big(h^{\alpha\beta} \partial_\alpha X^\mu \partial_\beta X^\nu g_{\mu\nu}(X)$$
$$+ \epsilon^{\alpha\beta} \partial_\alpha X^\mu \partial_\beta X^\nu B_{\mu\nu}(X) + \alpha' \Phi R(h) \Big) \quad (3.4)$$

---

[6]One might reasonably ask: in what sense are $g_{\mu\nu}$, $B_{\mu\nu}$ and Φ *background* fields? All we have shown so far is that they represent excited states of strings. The typical answer here runs along the following lines: when we introduce these fields as "background fields" in spacetime, we envisage them as coherent states of strings at all points in spacetime, at low energy so "stringy" effects can be ignored, so that they behave as typical quantum fields. There is much room for conceptual clarification and expansion here; the author hopes to explore such issues in a future paper.

Our conventions are such that $\epsilon^{\alpha\beta} = \pm 1/\sqrt{-h}$. We can now introduce the phenomenon of *T-duality* in the following way. First, assume that for some spacetime vector k^μ, the background fields satisfy $\mathcal{L}_k g = \mathcal{L}_k \Phi = 0$ and $\mathcal{L}_k H = 0$, i.e. $\mathcal{L}_k B = d\omega$. (Recall: $H = dB$ is the field strength of B). Here, $\mathcal{L}_k$ is the Lie derivative in the direction k; $\mathcal{L}_k g = 0$ clearly denotes an isometry of the spacetime metric. We choose coordinates, called *adapted coordinates*, $X^\mu = (\theta, X^i)$ such that $k^\mu = (1, 0)$ and $\mathcal{L}_k = \partial_\theta$, i.e. the isometry acts in the θ direction. Using the invariance of the action under $B \to B + d\omega$, we can choose all background fields to be independent of the coordinate θ. In these coordinates, the action becomes:

$$S = -\frac{1}{4\pi\alpha'} \int_\Sigma d^2\sigma \Big(\big(g_{\theta\theta}\partial_\alpha\theta\partial_\beta\theta + 2g_{\theta i}\partial_\alpha\theta\partial_\beta X^i + g_{ij}\partial_\alpha X^i\partial_\beta X^j \big) \eta^{\alpha\beta}$$
$$+ \big(2B_{\theta i}\partial_\alpha\theta\partial_\beta X^i + B_{ij}\partial_\alpha X^i\partial_\beta X^j \big) \epsilon^{\alpha\beta} \Big) \quad (3.5)$$

We now rewrite this action in first order form:

$$S' = -\frac{1}{4\pi\alpha'} \int_\Sigma d^2\sigma \Big(\big(g_{\theta\theta}V_\alpha V_\beta + 2g_{\theta i}V_\alpha\partial_\beta X^i + g_{ij}\partial_\alpha X^i\partial_\beta X^j \big) \eta^{\alpha\beta}$$
$$+ \big(2B_{\theta i}V_\alpha\partial_\beta X^i + B_{ij}\partial_\alpha X^i\partial_\beta X^j \big) \epsilon^{\alpha\beta} + 2\tilde{\theta}\epsilon^{\alpha\beta}\partial_\alpha V_\beta \Big) \quad (3.6)$$

where in the functional integral, we now integrate over X^i, θ and $\tilde{\theta}$. The V_α do not have physical meaning and are used purely as mathematical tools: integrating over the Lagrange multiplier field $\tilde{\theta}$ enforces $dV = 0$ with solution $V_\alpha = \partial_\alpha\theta$; inserting this into the action (3.6) leads back to the action (3.5). Alternatively, we can integrate out the V_α, i.e. eliminate them via their equations of motion. If we define:

$$L^\alpha = \frac{1}{g_{\theta\theta}} \Big(g_{\theta i}\partial_\beta X^i\eta^{\alpha\beta} + \big(B_{\theta i}\partial_\beta X^i + \partial_\beta\tilde{\theta} \big) \epsilon^{\alpha\beta} \Big) \quad (3.7)$$

we can rewrite (3.6) in the form:

11

$$S' = -\frac{1}{4\pi\alpha'} \int_\Sigma d^2\sigma \left(\left(g_{\theta\theta} \tilde{V}_\alpha \tilde{V}_\beta - g_{\theta\theta} L_\alpha L_\beta + g_{ij} \partial_\alpha X^i \partial_\beta X^j \right) \eta^{\alpha\beta} \right.$$
$$\left. + B_{ij} \partial_\alpha X^i \partial_\beta X^j \epsilon^{\alpha\beta} \right) \quad (3.8)$$

where $\tilde{V}_\alpha = V_\alpha + L_\alpha$ and $\tilde{V} = 0$ is the equation of motion for V_α. Making a change of variables and integrating over $\tilde{V}$, one obtains the dual action:

$$\tilde{S} = -\frac{1}{4\pi\alpha'} \int_\Sigma d^2\sigma \left(\left(\tilde{g}_{\theta\theta} \partial_\alpha \tilde{\theta} \partial_\beta \tilde{\theta} + 2\tilde{g}_{\theta i} \partial_\alpha \tilde{\theta} \partial_\beta X^i + \tilde{g}_{ij} \partial_\alpha X^i \partial_\beta X^j \right) \eta^{\alpha\beta} \right.$$
$$\left. + \left(2\tilde{B}_{\theta i} \partial_\alpha \tilde{\theta} \partial_\beta X^i + \tilde{B}_{ij} \partial_\alpha X^i \partial_\beta X^j \right) \epsilon^{\alpha\beta} \right) \quad (3.9)$$

with:

$$\tilde{g}_{\theta\theta} = \frac{1}{g_{\theta\theta}} \qquad \tilde{g}_{\theta i} = \frac{1}{g_{\theta\theta}} B_{\theta i} \qquad \tilde{B}_{\theta i} = \frac{1}{g_{\theta\theta}} g_{\theta i}$$

$$\tilde{g}_{ij} = g_{ij} - \frac{1}{g_{\theta\theta}} (g_{\theta i} g_{\theta j} - B_{\theta i} B_{\theta j}) \qquad \tilde{B}_{ij} = B_{ij} - \frac{1}{g_{\theta\theta}} (g_{\theta i} B_{\theta j} - B_{\theta i} g_{\theta j}) \qquad (3.10)$$

These are the *Buscher rules* for the string frame metric and antisymmetric tensor field [2]. The Jacobian of the change of variables in the path integral, which is necessary in the dualisation procedure, is non-trivial, and, after appropriate regularisation, can be shown to lead to a shift in the dilaton:

$$\tilde{\Phi} = \Phi - \frac{1}{2} \log |g_{\theta\theta}| = \Phi - \frac{1}{4} \log \left| \frac{g_{\theta\theta}}{\tilde{g}_{\theta\theta}} \right| \qquad (3.11)$$

Combining this with $g_{\theta\theta}^2 \det \tilde{g} = \det g$, one finds the T-duality invariant combination [2]:

$$e^{-2\tilde{\Phi}} \sqrt{\det \tilde{g}} = e^{-2\Phi} \sqrt{\det g} \tag{3.12}$$

Our derivation of the Buscher rules applies whenever the target space has an Abelian isometry. In superstring theories, the fermion fields also transform under T-duality; one finds that a T-duality transformation takes the type IIA theory to the type IIB theory and vice versa; and also relates the two heterotic theories.

T-duality is an exact symmetry, i.e. one that holds to all orders in perturbation theory [37]. By analogy with gauge theories, we can understand T-duality transformations for the bosonic string as effecting a self-duality that identifies points in the parameter space of the theory [42]. For the two type II superstrings and the two heterotic strings, the duality identifies points in the parameter space of the corresponding theory. Finally, we note that there exists an extension of T-duality, known as *mirror symmetry*, in which the two dual theories are defined on topologically *inequivalent* manifolds [41]. From a philosophical point of view, mirror symmetry presents the curious result that theories on very different manifolds are physically equivalent with respect to all observables. Mirror symmetry will prove important in our subsequent analysis of the interpretation of dualities.

3.3   S-Duality

Though the above characterisation of T-duality shall provide us with most of the tools necessary to draw substantive philosophical conclusions on the nature of dualities, it is worth briefly describing another well-known string-theoretic duality, called *S-duality*. S-duality relates the string coupling constant g_s to $1/g_s$. The two basic examples relate the type I superstring theory to the $SO(32)$ heterotic string theory and the type IIB superstring theory to itself. Thus, given our knowledge of the small g_s behaviour of these theories, given by perturbation theory, we learn how the dual theory behaves when $g_s \gg 1$. In other words, strongly coupled type I theory is equivalent to weakly coupled $SO(32)$ heterotic theory; the type IIB case the theory is related to itself [1].

13

3.4 Gauge-Gravity Duality

Some of the most widely discussed results in contemporary theoretical physics relate to the phenomenon of *gauge-gravity duality*, and in particular the so-called *AdS/CFT correspondence*, first proposed by Maldacena in 1997 [25]. The guiding idea underlying this duality is the *holographic principle*, which states that for some description of a system of particles interacting gravitationally, we are able to describe the physics via a theory involving only the boundary of the spacetime region containing the system [49].

In the AdS/CFT correspondence, the spacetime dimensions of the two theories must be radically different for the duality to establish itself [40]. For instance, type IIB string theory on the product space $AdS_5 \times S^5$ is equivalent to $\mathcal{N} = 4$ supersymmetric Yang-Mills theory on the four-dimensional boundary of the AdS_5 space [25]. The explanation for the mismatch in dimensionality is as follows: the spacetime dimensions corresponding to the space S^5 on the string theory side become internal degrees of freedom of particles on the gauge side (corresponding to its $SO\,(6)$ global symmetry), while the remaining dimension of the AdS_5 space becomes an energy scale in the dual gauge theory [40, 56].

Let us consider this specific case of the AdS/CFT duality in more detail (following the very clear exposition of [40]). $AdS_5 \times S^5$ is characterised by the metric:

$$ds^2 = \frac{r^2}{R^2}\eta_{\mu\nu}dx^\mu dx^\nu + \frac{R^2}{r^2}dr^2 + R^2 d\Omega_5^2 \tag{3.13}$$

Here, $\mu = 0\ldots3$, so the x^μ are four coordinates on AdS_5; r is the so-called *radial dimension*. Ω_5 gives the metric of the unit five-sphere S^5. $R \equiv l_{\text{string}}/\lambda^2_{\text{t'Hooft}}$ is the scale of spacetime relative to the string scale. $\lambda_{\text{t'Hooft}} \equiv g^2_{\text{YM}}N_c$ and $l_{\text{string}} \equiv \alpha'^{1/2}$, where g_{YM} is the Yang-Mills coupling constant and the physical meaning of N_c will be introduced shortly.

This metric corresponds to the horizon (in a sense defined below) of a D3-brane (a Dp-brane is a p-brane on which an open string can end; the 'D' stands for *Dirichlet boundary conditions*). The intuitive picture behind the AdS/CFT correspondence in this case is then as follows. First, imagine a stack of N_c parallel D-branes. It is known that D-branes gravitate, with gravitational coupling strength determined by g_s. The warping of spacetime geometry

away from Minkowski is then determined by $N_c g_s$ [40]. Given fixed N_c we can tune the parameter g_s across a range of values from very small to very large. This warping will be minimal, and the spacetime near flat, when $N_c g_s \ll 1$. In this case, there can be both opened and closed strings, but with low coupling strength they will be virtually decoupled from each other. The closed strings that decouple from the open strings give a picture of linearised, perturbative gravity. The open strings stuck to the D-branes are described by a gauge theory with N_c colours restricted to the D-brane stack [40].

On increasing the coupling strength so that $N_c g_s \gg 1$, the gravitational effect of the D-branes on the spacetime metric becomes non-negligible, leading to a curved geometry; in fact, a black hole geometry (or *black brane* geometry). By analogy with a Reissner-Nordström black hole, this geometry is $AdS_5 \times S^5$ [18]. Strings near the event horizon will be red shifted from the point of view of distant observers, and so will appear to have low energies. In the limit of low energies (ignoring massive states) the strings near the event horizon will decouple from the strings on the (flat) conformal boundary [40].

Putting these two scenarios together, at weak coupling the physics is described by a gauge theory on flat space, while at strong coupling it is described by a closed string theory on $AdS_5 \times S^5$. Maldacena conjectured that there was a duality linking these two descriptions together, by varying the t'Hooft parameter $\lambda_{\text{t'Hooft}}$, so that it was really one theory being viewed from different regions of parameter space [25, 40].

The holographic nature of the duality is evident from the fact that one is dealing with boundary data in the string theory. It is the boundary data that uniquely determines the gauge field theory. The gauge theory lives on the $r \to \infty$ conformal boundary of AdS_5, with the string theory defined throughout the $r < \infty$ interior, i.e. the bulk. The above equivalence[7] implies that the Hilbert spaces of these two quantum theories will be equivalent, which is curious since they are built up from very different types of object: strings, branes, gravitons, etc. on the string theory side and local gauge-invariant functionals of the gauge fields and their derivatives on the other [40]. If correct though, string theory (with quantum gravity) can be translated into the language of a fairly ordinary conformal field theory!

---

[7]It is worth noting that this equivalence has not been *proven*. Nevertheless, a vast number of non-trivial correspondences between theories linked by such dualities gives physicists confidence that the equivalence is correct.

4 The Hole Argument

In this section, we briefly recap the setup of, and potential responses to, the hole argument of General Relativity. Exploring the possible solutions to this problem will prove useful when it comes to discussing the possible interpretations of dualities.

4.1 Setup

The hole argument is an infamous problem case in the foundations of General Relativity (GR), originally raised by Einstein during the development of the theory (in the ultimately misguided attempt to demonstrate that no generally covariant theory is admissible as a theory of gravity [32]), and subsequently resuscitated by Earman and Norton [10], this time as a means to arguing against the reification of the manifold in GR. In other words, the argument was deployed by Earman and Norton to argue in favour of *relationism* (**REL**) about spacetime and against *substantivalism* (**SUB**), where these are defined as follows [39]:

Definition 6. (Substantivalism) *A complete catalogue of the fundamental objects in the universe lists, in addition to the elementary constituents of material entities, the basic parts of spacetime.*

Definition 7. (Relationism) *Claims apparently about spacetime itself are ultimately to be understood as claims about* material *entities and the possible patterns of spatiotemporal relations they can instantiate.*

We take it that **SUB** $\leftrightarrow$ ¬**REL**. Let us recall how the hole argument proceeds, following [38, 39]. Let $\mathcal{M}_1 = \langle M, g_{ab}, T_{ab} \rangle$ be a model of a generally relativistic theory. Here, M represents the spacetime manifold[8], g_{ab} the metric on this manifold, and T_{ab} the stress-energy tensor of the theory[9]. It follows from the diffeomorphism invariance of GR that, for an arbitrary diffeomorphism d, $\mathcal{M}_2 = \langle M, d^* g_{ab}, d^* T_{ab} \rangle$ also satisfies the theory's equations. The natural (though not ineluctable) conclusion is that $\mathcal{M}_1$ and $\mathcal{M}_2$ *jointly* represent spacetimes (call them W_1 and W_2) that are physically possible according to the theory.

---

[8]In discussions of the hole argument, the focus is on *manifold substantivalism*, according to which spacetime is identified with the manifold M. The hole argument does not necessarily speak against *metric substantivalism*, which states that the metric field g_{ab} forms an essential part of spacetime. See also footnote 12.

[9]Note that we are now using abstract (i.e. coordinate-free) indices, denoted by Roman letters.

In $\mathcal{M}_1$, each $p \in M$ is assigned certain properties encoded by $\langle g_{ab}(p), T_{ab}(p) \rangle$; in $\mathcal{M}_2$, p is assigned the generically distinct properties encoded by $\langle d^* g_{ab}(p), d^* T_{ab}(p) \rangle$. But, according to the substantivalist, M represents physical spacetime. This means that (on one natural understanding of how the points of M represent physical spacetime points), $\mathcal{M}_1$ and $\mathcal{M}_2$ represent one and the same spacetime point as having different properties. This gives us the next ingredient in the argument: the claim that the substantivalist is committed to regarding W_1 and W_2 as *distinct* possible worlds[10]. [38]

The problem is that, if this interpretation of spacetime models is permitted, GR is radically indeterministic. Let d be a *hole diffeomorphism*, i.e. a map that is only non-trivial within a restricted region of M (the so-called *hole*). Suppose that, relative to the metric of $\mathcal{M}_1$, d is only non-trivial to the future of some spacelike hypersurface Ψ. $\mathcal{M}_1$ and $\mathcal{M}_2$ will then be identical structures up to and including this surface but differ to its future. On the proposed interpretation of $\mathcal{M}_1$ and $\mathcal{M}_2$, they represent spacetimes that are identical up to the space like surface represented by Ψ but that differ to its future. It follows that the equations of GR, together with a complete specification of the history of the world up to some spacelike hypersurface, fail to fix the future.

The essence of the hole indeterminism, then, is a follows: Let W_1 and W_2 be distinct possible world that are physically possible according to some theory $\mathcal{T}$[11]. Suppose that W_1 and W_2 are identical up to some time t but differ after t. A complete specification of a possible world up to t does not distinguish between W_1 and W_2. Hence the history of a world up to t, together with $\mathcal{T}$, can fail to fix the future of that world. It is important to emphasise that Earman and Norton do not see this as a problem for substantivalism because they think indeterminism is objectionable *per se* [39]. Their claim, rather, is that determinism should fail only for reasons of physics and not as the result of a metaphysical commitment and in a theory-independent way:

> Our argument does not stem from a conviction that determinism is or ought to be true... Rather our point is this. If a metaphysics which forces all our theories to be deterministic is unacceptable, then equally a metaphysics which automatically decided in favour of indeterminism, is unacceptable. Determinism may fail, but

---

[10]This is Earman and Norton's so-called *acid test* of substantivalism.

[11]Not necessarily GR, for the same problem may arise in any theory defined on a manifold, with diffeomorphism invariance of the fields.

if it fails, it should fail for a reason of physics, not because of commitment to substantival properties which can be eradicated without affecting the empirical consequences of the theory. [10]

Note that $\mathcal{M}_1$ and $\mathcal{M}_2$ are isomorphic structures (something which is *not* true in many cases of duality: a point to which we shall return shortly). The possibilities they represent, therefore, involve exactly the same patterns of qualitative features. If W_1 and W_2 are distinct possibilities, they differ only over which spacetime points instantiate which of the particular features common to both worlds. In the terminology of modal metaphysics, the difference between the possibilities is merely *haecceitistic* [22].

Finally, it is worth pointing out that Earman and Norton do not make explicit the definition of determinism used in the above setup for the hole argument. As Brighouse states [3], clearly the failure of some brand of Laplacian determinism is what they have in mind, whereby agreement (of some kind) on a given region is insufficient to ensure agreement (of some kind) throughout spacetime. With this in mind, we can follow Brighouse in defining determinism as follows: [3]

Definition 8. (Determinism) *A spacetime theory $\mathcal{T}$ is deterministic if and only if for any given models $\mathcal{M}_1$ and $\mathcal{M}_2$ of $\mathcal{T}$, if those models are physically equivalent before time t (or, at time t) then they are physically equivalent for all times.*

Most will agree that definition (8) adequately expresses the idea of determinism. The issue as to whether a theory satisfies definition (8) will depend on which relation between models defines when two models are physically equivalent.

4.2   Responses to the Hole Argument

Following Pooley [39], when situating possible classes of response to the hole argument, it is helpful to lay out the following claims:

HAE : $\mathcal{M}_1$ and $\mathcal{M}_2$ (can be taken) to represent distinct physically possible worlds.

18

LZE : $\mathcal{M}_1$ and $\mathcal{M}_2$ (must be taken to) represent the same possible world.

ONE : If $\mathcal{M}_1$ is taken to represent a possible world, then $\mathcal{M}_2$ does not represent a possible world.

Earman and Norton's argument that substantivalism commits us to indeterminism then rules as follows: (A) **SUB** $\rightarrow$ **HAE**, (B) **HAE** $\rightarrow$ Indeterminism, (C) ¬Indeterminism; therefore (D) ¬**SUB**. There are four possible classes of response to this argument:

1. Accept both **SUB** and indeterminism, but argue that this indeterminism is acceptable.

2. Reject **SUB** and accept **REL**, arguing that the latter commits us only to **LZE**, thereby avoiding the indeterminism.

3. Question (A), i.e. **SUB** $\rightarrow$ **HAE**.

4. Question (B), i.e. **HAE** $\rightarrow$ Indeterminism.

Option (1) can be seen as saving **SUB** by rejecting (C) and thereby embracing indeterminism. Option (2) amounts to accepting the force of the hole argument, rejecting **SUB** accordingly. Options (3) and (4) attempt to save **SUB** by calling into question premises (A) and (B) of the hole argument, respectively. These options (1)-(4) provide a good means of situating all mainstream responses to the hole argument. We will now review some popular approaches by very briefly discussing each of these classes of response in turn.

(1) Accept the Indeterminism. A substantivalist might agree that accepting GR involves a commitment to such haecceitistic distinctions and accept that the theory is indeterministic. However, they might deny that this indeterminism is in any sense troublesome precisely because it is an indeterminism only about which objects instantiate which properties and not about which patterns of properties are instantiated. [38]

(2) Accept Relationism, Reject Substantivalism. This is Earman and Norton's preferred response [10]. There are many possible subsequent relationist moves. One has been to treat the metric field as just another material field [4, 45]. For example, for Brown, the metric field only gains its usual "chronometrical significance" (i.e. only corresponds to the practical

geometry manifest by the behaviour of material rods and clocks) in virtue of the particular way it dynamically couples to matter [4][12].

(3) Question the Commitment to Haecceitistic Differences. The most popular substantivalist response to the hole argument has been *sophisticated substantivalism*: a version of substantivalism that denies the existence of physically possible spacetimes that differ merely haecceitistically. The simplest way to secure this is to endorse *antihaecceitism*, i.e., the denial of merely haecceitistic distinctions between possible worlds. In this case, W_1 and W_2 are identical, and the indeterminism is removed. Thus, one can maintain a substantivalist position and avoid the force of the hole argument. [38]

(4) Question that Haecceitistic Differences Commit us to Indeterminism. One can accept **HAE** but reject indeterminism by rejecting premise (B) of hole argument. One way to do this is to identify some means of identifying one of the W_i as a possible world, while rejecting all others as possible worlds. Perhaps the most well-known such tactic is that of Maudlin, who endorses the doctrine of "metric essentialism". On this view, spacetime is an essentially metrical object and the points of spacetime bear their metrical relations *essentially*. This preferentially singles out of the of the worlds W_i as physical, thereby accepting **ONE** and so circumventing the indeterminism [27]. An alternative option is endorsed by Brighouse [3] and Pooley [38]: accept that GR is committed to haecceitistic distinctions but deny that it follows that GR is indeterministic by redefining determinism, so that the concept is only sensitive to qualitative differences.

5   The Interpretation of Dualities

With all this setup in hand, we are now in a position to assess possible interpretations of string-theoretic dualities. We will find that, while there is no perfect analogy between such dualities and the hole argument of GR, the wealth of literature on possible responses to the hole argument can provide inspiration when it comes to the interpretation of dualities. Indeed,

---

[12]Note, though, that one is potentially misrepresenting Brown by labelling his position as relationist, since he carefully avoids this terminology [4]. More generally, it is unclear whether this position is really faithful to all possible forms of relationism: while the existence of manifold points is denied, there is a strong association of the metric field with spacetime.

while the analogy between the two cases has been pointed out in [43], there has yet to be any systematic exploration in this direction.

Before we begin our analysis, it is worth making three general observations on the analogy between the hole argument and the problem of string-theoretic dualities. First, there is no direct analogy of *indeterminism* in the case of string-theoretic dualities. Nevertheless, there is an analogous problem of an underdetermination as to which of a set of possible worlds W_i represents the actual world. The underdetermination is problematic in the case of dualities since dual theories often make highly different ontological claims about the world. The hope is that there is a sufficiently strong analogy between the hole argument and the case of dualities to mean that possible solutions to the former may yield interesting insights in the case of the latter.

Second, unlike $\mathcal{M}_1$ and $\mathcal{M}_2$ in the case of the hole argument, the models of the two dual theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ are in general *not* isomorphic. This is easy to see in, for example, the case of T-duality for the bosonic string: if the first model is specified by $\mathcal{M} = \langle M, g_{ab}(p), \Phi(p), B_{ab}(p), \ldots \rangle$, then the second model of the theory[13] is specified by $\tilde{\mathcal{M}} = \langle M, \tilde{g}_{ab}(p), \tilde{\Phi}(p), \tilde{B}_{ab}(p), \ldots \rangle$. The fields for the second model are related to the fields of the first model via the Buscher rules, and this mixing is highly non-trivial, *unlike* the two models in the hole argument, where the fields on the manifold for the two models are related by a simple diffeomorphism. In other words, in the hole argument, *all* fields at $p \in M$ in one model are mapped to $q \in M$ in the second model, whereas for dualities fields at $p \in M$ may generically be mapped to *different* manifold points in the second model; this mixing results in a loss of isomorphism between the two models.

Third, it is worth noting that, in the case of e.g. mirror symmetry and gauge-gravity duality, even the manifolds M on which the dual field theories are defined are not necessarily the same (i.e. topologically equivalent). It will turn out that this closes off some options that were open to us in the case of the hole argument, where the manifold structure was the same for both $\mathcal{M}_1$ and $\mathcal{M}_2$.

With these points in mind, let us begin to consider possible responses to the problem of dualities. Teh [49], Knox [23] and Matsubara [26] all list possible responses to the apparent

---

[13]The *same* theory: recall that T-duality for the bosonic string is a self-duality.

underdetermination in this case. For example, Matsubara gives the following options:

I. The two different theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ give two models $\mathcal{M}$ and $\tilde{\mathcal{M}}$ which describe two genuinely distinct possible worlds W_1 and W_2, with it underdetermined which of these worlds is the actual world. Given this interpretation, there are two possible alternatives: [26]

    (a) Accept that we have a genuine and problematic case of underdetermination.

    (b) Identify "shared structure" between the two theories, and argue that only this shared structure can properly be taken to describe the actual world. This route naturally leads to weaker forms of realism, such as *structural realism*. [55]

II. Argue that $\mathcal{T}$ and $\tilde{\mathcal{T}}$ give models $\mathcal{M}$ and $\tilde{\mathcal{M}}$ which do not really describe different possible worlds after all.

While such classifications of possible responses to the problem of dualities are useful, we must ask whether these options exhaust the possible responses. This is especially so in light of the fact that the literature on dualities is abound with other apparently different responses; and also in light of the fact that other options seem available when one considers the analogy with the hole argument. In fact, by incorporating both of these considerations we are in a better position to situate all possible responses in the dualities case. With this in mind, we lay out below an extended range of responses to the case of dualities, labelled (1)-(4) in analogy with the (1)-(4) presented in our discussion of the hole argument in the previous section.[14]

1. Dual theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ give models $\mathcal{M}$ and $\tilde{\mathcal{M}}$ which describe two different possible worlds W_1 and W_2, only one of which could be the actual world. Within this response, there are three sub-responses:

    (a) Accept that models $\mathcal{M}$ and $\tilde{\mathcal{M}}$ describe two different possible worlds W_1 and W_2, either one of which could be the actual world. That is, accept that the dual theories give a genuine case of strong underdetermination.

---

[14]Though there are subtle differences, discussed below.

(b) Argue that models $\mathcal{M}$ and $\tilde{\mathcal{M}}$ do *globally* describe two distinct possible worlds W_1 and W_2, but that we can avoid the indeterminism by only applying each model in restricted, non-overlapping regions of the parameter space of the actual world.[15]

(c) Reconcile the two theories by embedding them in a unique overarching theory $\mathcal{T}^*$.

2. Argue that $\mathcal{M}$ and $\tilde{\mathcal{M}}$ only really describe one possible world, and that the difference between them is simply a gauge redundancy. (This is a rough analogue of some relationist approaches to the hole argument, such as [4].)

3. Change our metaphysical view of the manifold featuring in the models of the theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ in order to argue that the two models $\mathcal{M}$ and $\tilde{\mathcal{M}}$ only really describe one possible world. (This is the analogue of sophisticated substantivalist approaches to the hole argument.)

4. Privilege one of the two dual theories, arguing that only the models of this theory describe physically possible worlds. (This is the analogue of Maudlin's metric essentialist approach to the hole argument [27].)

Our (1a) corresponds to Matsubara's (Ia); his (Ib) we shall construe as a possible structuralist approach response to (1a). Our (2), (3) and (4) fit into Matsubara's (II), though he considers none of these options. It is unclear whether our (1b) and (1c) should or should not fall under Matsubara's (II), though in his paper it is implicit that this is so [26]. For the rest of this paper we shall not focus on the mundane task of comparing our list of responses with those of others; though we shall say that our list of options is more comprehensive than others presented in the literature up to this point.

Importantly, the options presented in our list above are not maximally mutually exclusive, and in fact it may turn out that they are only coherent or defensible when taken in conjunction with others. Indeed, one can arguably read Teh [49], Knox [23] and Rickles [42] as all defending a combination of (1b), (1c) and (2); we shall ultimately join them in arguing that this is one of the most defensible approaches to string-theoretic dualities.

Finally, it is worth making some further comments on (2) and (3) above and the analogous approaches in the hole argument debate. Arguably, (3) can be viewed as a special case of (2):

---

[15]We shall discuss what is meant by *parameter space of the actual world*, and whether this notion is coherent.

both claim that $\mathcal{M}$ and $\tilde{\mathcal{M}}$ describe the same possible world, but (3) gives a more focussed reason in terms of the ontology of manifold points. In the case of the hole argument, for both the sophisticated substantivalist and the relationist who denies the existence of manifold points, the choice of one of a range of diffeomorphically-related models is a mere gauge choice; both of these approaches are, therefore, analogous to (2). However, we have treated the sophisticated substantivalist analogue as an approach in its own right in the case of string dualities to justify a more extensive evaluation of this option.

5.1 Dual Theories as Describing Distinct Possible Worlds

5.1.1 Underdetermination

As we have seen, string-theoretic dualities present a *prima facie* problem of strong underdetermination of theory by evidence, since we have two theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ with respective models $\mathcal{M}$ and $\tilde{\mathcal{M}}$ which correspond to two possible worlds W_1 and W_2 with distinct ontologies; only one of these worlds can be the actual world. The first possible response, as given in our list above, is simply to accept that this is a genuine case of underdetermination. If so, what follows? Unlike the case of the hole argument, where one option was to argue that the indeterminism is not worrying since it amounts only to a question of which objects instantiate which properties, it seems that such an option is less defensible in this instance. The reason is that, as mentioned, the worlds described by $\mathcal{T}$ and $\tilde{\mathcal{T}}$ are not isomorphic, as was the case for the hole argument. Moreover, the ontological claims of the two models are often very different.

If one accepts that the underdetermination here is real, one might be led to weaken or dismiss one's previous realist attitudes in the philosophy of science. As Rickles states [42], a positivist might well be no more fazed by dual descriptions than standard underdetermined cases (such as Poincaré's example of curved space versus distorting forces that have equivalent empirical geometrical consequences [35,36]). Alternatively, one might choose to endorse constructive empiricism in order to avoid the threat of underdetermination from string-theoretic dualities [7,12], or some form of structural realism [8,9,26].

These discussions lead into familiar territory in the philosophy of science, albeit with some interesting twists. For example, Rickles has argued that dualities might even present a

case of *structural* underdetermination, thereby potentially posing problems for the structural realist [42]. Moreover, it is worth noting that many (post-quantum mechanics) physicists are likely to take some form of instrumentalist line towards string dualities, without concern for ontological ramifications [5]. For the purposes of this essay, however, we shall continue to focus on the task of accounting for dualities from a realist point of view. As such, we now move on to consider the other possible responses to the case of string dualities.

5.1.2 Pluralism

One potential way to diffuse the underdetermination in the case of dualities is as follows: admit that *prima facie* the respective models $\mathcal{M}$ and $\tilde{\mathcal{M}}$ of the two theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ describe two different possible worlds if construed globally, but argue that they can both also concurrently describe the actual world, if we partition the "parameter space" of the actual world such that in some places in this space the world is to be described by $\mathcal{M}$, and in other places it is to be described by $\tilde{\mathcal{M}}$. Physicists have made many suggestive comments along these lines. For example, Susskind states:

> [I]f one listens carefully, string theory is telling us that in a deep way reductionism is wrong, at least beyond some point. ... In string theory the coupling constant is not a constant; it is a field that can vary in spacetime. The field is called the dilaton field. Thus, in a strong dilaton wave, there will be places where the fundamental strings are the simplest objects, and other places where the D-strings are simplest. In fact there will be places where the coupling is of order one where neither is more elementary than the other. [48]

It is likely that Susskind has some form of S-duality in mind here: at regions where $g_s \gg 1$, the physics is best described in terms of the ontology of the dual theory, with coupling $\tilde{g}_s \sim 1/g_s$[16]. In fact, such sentiments are also echoed by philosophers of physics such as Rickles (though, as will become clear later, pluralism in this sense does exhaust Rickles' position):

> Dual theories provide distinct but ultimately empirically equivalent representa-
> tions. Do they thereby amount to underdetermination? I would argue that there

---

[16]Although Susskind talks of spacetime, we can continue to talk of an abstract parameter space.

are crucial and subtle differences. The dual theories are not in competition: they are complementary. They are both true in a sense, and the practice of physics suggests, in many cases at least, a pluralistic stance with respect to the dual theories: one theory is better equipped to describe some portion of parameter space than another [in the sense that calculations performed using that theory are simpler than those performed using its dual.] [42]

The main problem with this approach is making sense of the notion of *parameter space of the actual world*. One possible response is to say that such talk is elliptical: in fact, we need only consider the parameter space of one of our theories, say $\mathcal{T}$. As long as $g_s < 1$ in the models of this theory, we should indeed describe the world using the models of $\mathcal{T}$. However, at the points in the parameter space of $\mathcal{T}$ where we have models with $g_s \geq 1$, we should switch to describing the world via the models of the dual theory $\tilde{\mathcal{T}}$. While this seems to make good sense, and is likely what Susskind has in mind, it does not avoid the problem of underdetermination in the case of dualities *per se*, since either theory *can* be used to describe the world, even when coupling is large (one need only look to quantum chromodynamics for an example of a strongly coupled theory which we take to be ontologically charged). Moreover, this does not account for cases such as T-duality, where there is no parameter analogous to g_s by way of which analogous arguments can go through.

Are there *any* compelling reasons that the pluralist can give to always preference just one of $\mathcal{M}$ and $\tilde{\mathcal{M}}$? If it can be demonstrated that one of $\mathcal{M}$ and $\tilde{\mathcal{M}}$ of the respective theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ is always pathological in some sense (for example, due to the presence of curvature singularities), and if it can be convincingly argued that such pathological features mean that such a model cannot represent the real world, then we might be able to establish metaphysical grounds for always preferencing one of $\mathcal{M}$ and $\tilde{\mathcal{M}}$ (for a given point in the moduli space of one of our two theories), and thereby advancing the pluralist position. Though this argument has yet to be fleshed out in a compelling form, and it is perhaps unlikely that one of the two dual models will *always* be pathological in this way, it is feasible that *if* such a task were accomplished, this might yield a viable form of pluralism.

Another option is to attempt to embed the two theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ into a deeper theory $\mathcal{T}^*$; the (enlarged) moduli space of this deeper theory is then partitioned into domains in which each of the two theories apply. This is akin to the idea behind modern talk of *M-theory*;

whether the proposal fares any better shall be evaluated in the following subsection.

5.1.3 Unificationism

In the context of string theory, the network of duality relations between the various consistent perturbative theories is taken as evidence that there is some deeper unifying theory of which the string theories are offering glimpses. For the five superstring theories, this suspected unifying theory is known as *M-theory*, and was first proposed by Witten in 1995 [21]. Witten offers the following summary of the situation:

> In uncompactified ten-dimensional Minkowski space, the strong-coupling limit of the type I superstring is the weakly coupled heterotic $SO(32)$ superstring; the strong-coupling limit of the type IIA superstring is related to eleven-dimensional supergravity; the strong coupling limit of type IIB theory is equivalent to the same theory at weak coupling; and the strong-coupling limit of the $E_8 \times E_8$ heterotic string involves eleven-dimensional supergravity again. From this list, and additional items that appear after compactifying some dimensions, we learn that the different theories are all one. The different superstring theories studied in different ways in the last generation are different manifestations of one underlying, and still mysterious theory, sometimes called M-theory.[17] [53]

It is typically said that the various consistent superstring theories are distinguished points in the moduli space of M-theory. Of course, the duality relationships between the theories will hold independently of the existence of an M-theory, but in order to achieve a computable scheme for the whole of the moduli space (including regions away from the distinguished perturbation-friendly points) such an underlying theory is required [42].

The main problem with this approach is the following. Even if we embed $\mathcal{T}$ and $\tilde{\mathcal{T}}$ in $\mathcal{T}^*$ so that they describe limited regions of the moduli space of the deeper theory (see figure (1)), this does not remove the fact that we still have two dual descriptions of the same physical

---

[17]One might reasonably question whether such observations do indeed warrant the inference to the existence of M-theory: certainly, the existence of such a theory does not appear to be implied by such observations as a matter of necessity. This point is discussed in more detail below.
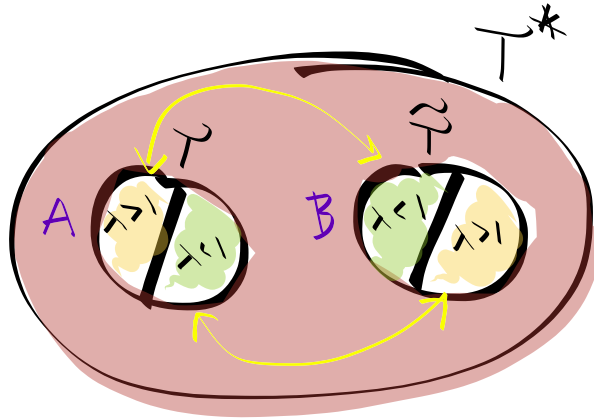
Figure 1: The embedding of the perturbative theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ into a more fundamental theory $\mathcal{T}^*$. The red circle indicates the parameter (moduli) space of $\mathcal{T}^*$. $\mathcal{T}$ and $\tilde{\mathcal{T}}$ individually describe patches of this parameter space, denoted A and B respectively. In this example, there exists a duality which takes points in the parameter space of $\mathcal{T}$ with coupling parameter $\lambda > 1$ to points in the parameter space of $\tilde{\mathcal{T}}$ with $\lambda < 1$, and vice versa. This could therefore be some version of superstring S-duality with $\lambda = g_s$, or AdS/CFT duality with $\lambda = \lambda_{\text{t'Hooft}}$. In the former case a candidate for $\mathcal{T}^*$ is M-theory; in the latter case a candidate for $\mathcal{T}^*$ is not known.

phenomena! So the unificationist strategy by itself cannot succeed. Indeed, even marrying unificationism with the pluralism of the previous subsection, we do not necessarily avoid the problem of dualities: pluralism provides us with a practical rule for when to use one model over another (e.g. use the model with the smallest relevant coupling parameter: g_s in the case of S-duality); this does not mitigate the fact that we still *could* use either theory to describe the same physics (unless our arguments based on pathology are independently justified). Hence, whether separately or united, pluralism and unificationism do not necessarily seem to be able to overcome the problem of underdetermination in the case of string dualities.

In fact, there is also another problem for this unificationist approach: the very notion that such a unifying theory must exist is an unjustified assumption. While there is some suggestive evidence to this end *apropos* superstring theory and M-theory, as yet even a partial formulation of this theory is not on the horizon [44, 47]. Moreover, in the case of gauge-gravity dualities, such a unifying theory is rarely spoken of, and there seems to be little consensus on whether such a theory should exist [49]. Absent the details, the demand that some deeper theory must exist in order to circumvent the underdetermination faces the charge of being *ad hoc*.

As an aside, it is worth commenting on the special case where the moduli space of $\mathcal{T}^*$ is identical with the moduli spaces of the dual theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$. Here, $\mathcal{T}$ and $\tilde{\mathcal{T}}$ are identified as the *same* theory; the duality in question then becomes a self-duality. Though this possibility is not widely countenanced in the case of string dualities, there are several plausible cases in the history of physics where this phenomenon may have occurred. To take two examples: (a) the demonstration of the empirical equivalence of Heisenberg's matrix mechanics and Schrödinger's wave mechanics (two *prima facie* profoundly ontologically distinct theories) [30, 31]; (b) the choice of Jordan frame versus Einstein frame in the Brans-Dicke scalar-tensor theory of gravity [11].[18]

5.2 Dual Theories as Describing the Same World: Gauge Reductionism

In response to the hole argument, some relationists argue that we should not treat the manifold as physical; therefore the indeterminism (which was a matter of which manifold points instantiate which properties) is removed, and we can treat the situation as a simple gauge redundancy. Is something analogous possible in the case of string-theoretic dualities, i.e. can we identify the models of the two dual theories, just as we would identify states on gauge orbits within a theory? As Rickles points out [42], the main problem with this approach is that gauge redundancy as usually construed amounts to the identification of certain states of *one specific* theory. While one can analogously identify models of self-dual theories, is not clear that this makes sense for dualities that are not self-dualities, where we would require some inter-theoretic notion of model identification.

Let us be more concrete. While it is possible to make sense of self-dualities from the perspective of gauge redundancy (for example, T-dual models for the bosonic string can be identified; the choice of one or the other is then equivalent to a choice of gauge [42] – and typically, as with gauge fixing, we will choose the model which is most convenient), what of inter-theoretic dualities, such as the AdS/CFT correspondence, or any of the dualities which relate two different superstring theories? The problem is that the notion of identifying physical states across theories is in general not well-defined.

---

[18]The author intends to investigate in detail such candidate cases of dualities in the history of physics in a future paper.

The most natural response to this problem is to endorse the unificationist standpoint detailed above. This solves the problem as it allows us to embed both dual theories within the parameter space of some deeper theory; then models of the theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$, which correspond to points in the parameter space of the unifying theory $\mathcal{T}^*$ can *then* be identified, since they now belong to the same parameter space. *Then*, one can identify models across these theories. Hence we see that the unificationist response and the gauge reductionist response go hand-in-hand.

In fact, this leads us to a defensible interpretation of string dualities: a combination of pluralism, unificationism, and gauge reductionism. The story runs as follows: we postulate some deeper theory $\mathcal{T}^*$; $\mathcal{T}$ and $\tilde{\mathcal{T}}$ describe respective patches A and B of the overarching parameter space of the theory as in figure (1). Each point in A corresponds to a model of $\mathcal{T}$, and each point in B corresponds to a model of $\tilde{\mathcal{T}}$. Embedding the parameter spaces of $\mathcal{T}$ and $\tilde{\mathcal{T}}$ within the overarching parameter space of $\mathcal{T}^*$ allows us to identify points in A and B which are related by a duality; choosing to describe the world via one than the other then amounts to a specific gauge choice. The principle via which one should choose a "gauge" (i.e. dual theory) to describe the world is one such as Susskind's, as presented in our discussion of pluralism. The ultimately ontology of the world should be described in terms of $\mathcal{T}^*$. A view akin to this is endorsed by Teh [49], Knox [23], Huggett [19] and Rickles [42]. Naturally, the biggest problem lies in the first step: finding an overarching theory $\mathcal{T}^*$ to begin with!

5.3 Sophisticated Substantivalism Analogies?

One possible approach to string theory dualities only becomes apparently after considering the sophisticated substantivalist solution to the hole argument. To recap, the sophisticated substantivalist adopts the antihaecceitistic thesis that two worlds cannot differ solely with regard to which objects instantiate which properties, so that GR models $\mathcal{M}_1$ and $\mathcal{M}_2$ only describe one possible world. The question is: could such an account also allow us to avoid the underdetermination which appears to arise in the context of string-theoretic dualities? Unfortunately, this response seems doomed to fail.

To see this, first note that while $\mathcal{M}_1$ and $\mathcal{M}_2$ are isomorphic models in the case of the hole argument, this is not in general true for the two models $\mathcal{M}$ and $\tilde{\mathcal{M}}$ in the case of string theoret-

ical dualities, as we have already seen. If there is no longer an isomorphism between the two dual descriptions, then saying that the models differ solely in regard of which objects instantiate which properties, and therefore only correspond to one possible world, is not so simple. Second, mirror symmetry and gauge-gravity duality make it apparent that two dual theories can have topologically inequivalent manifolds. Hence, this kind of approach to dualities seems fundamentally misguided, because we do not even have a uniquely defined manifold between the two models. [41, 42]

5.4   Preferencing One Dual Theory

Consider the specific case of the AdS/CFT correspondence. Oriti points out that most string theorists tend to speak as if the four-dimensional spacetime of the boundary gauge theory is real, with the bulk spacetime appearing as an auxiliary construction [33]. For example, Horowitz and Polchinski note that the AdS/CFT correspondence is a little different from other dualities in that the gauge theory side is exactly understood, while the string theory side is only approximately understood [18]. Building on this, they write:

> In the AdS/CFT case, the situation may not be so symmetric, in that for now the gauge side has an exact description and the string/gravity side only an approximate one: we might take the point of view that strings and spacetime are emergent and that the ultimate precise description of the theory will be in variables closer to the CFT form. [18]

Here, Horowitz and Polchinski appear to be using the explanatory primacy of the gauge theoretical description (the CFT side) in support of a thesis about its ontological primacy. In fact, such a position is presented in a wealth of popular literature on the AdS/CFT correspondence - see for example [6, 13, 24]. The question is: why does the *epistemic* fact about what human beings happen to currently know about two dual theories relative to one another warrant the *metaphysical* conclusion that the theory we currently know more about must give the correct description of the world? Such worries have been expressed by Teh [49], Knox [23], and Rickles [42]. What we need is some argument to the effect that one description of the world has metaphysical priority over its dual, given that the duality relation is formally sym-

31

metric [40]. Unfortunately, as Teh points out, "We have no good reason to think of the gravitational side of the duality as metaphysically emergent from the gauge theory side, or vice versa." [49]

Though this observation is reasonable at this point, it is at least worth considering one way in which the metaphysical primacy of one of the dual theories might be established. If it can be demonstrated that one of $\mathcal{M}$ and $\tilde{\mathcal{M}}$ of the respective theories $\mathcal{T}$ and $\tilde{\mathcal{T}}$ is pathological in the sense discussed in section 5.1.2, then we might be able to establish grounds for preferencing one of $\mathcal{M}$ and $\tilde{\mathcal{M}}$. If in every case the models of $\mathcal{T}$ end up being preferred over the models of $\tilde{\mathcal{T}}$ in this way, then we might obtain some principled reason for thinking that $\mathcal{T}$ rather than $\tilde{\mathcal{T}}$ correctly represents the ontology of the world.[19] Hence we see that it might be possible to defend this position on dualities, but we require defensible grounds for preferencing one of the two dual theories which are not obviously forthcoming.

6 Conclusions and Outlook

In this essay, we have used the literature on candidate solutions to the hole argument of GR as a springboard to categorising solutions to the apparent problem of strong underdetermination in the context of dualities. We have found that, in our ensuing categorisation of possible solutions to this problem, the options available are not mutually exclusive; and in fact a combination of pluralism, unificationism, and gauge reductionism seems to offer one of the most viable interpretations of the phenomenon. Though many of the other potential options do not appear promising, we have also found that the tactic of giving metaphysical priority to the models of one of the dual theories (whether on a local scale as with pluralism, or on a global scale as discussed in section 5.4) in order to privilege its ontological claims would be tenable, *if* a convincing metaphysical grounding could be found.

It is interesting to note that theoretical physicists tend to endorse both these approaches to dualities, but in different contexts. The attitude of Horowitz and Polchinski towards gauge-gravity duality - whereby the gauge theory is awarded ontological priority over the bulk theory (often, as we have seen, by fiat, or for philosophically unmotivated reasons) is ubiquitous in

---

[19]If, on the other hand, which of $\mathcal{M}$ or $\tilde{\mathcal{M}}$ is preferenced depends on the point in the moduli space of one of the two theories under consideration, then we are led to the defensible form of pluralism already discussed.

the literature on this subject [18]. On the other hand, the attitude towards superstring dualities is often roughly in line with our tripartite unificationist account above. In fact, authors often run these interpretations together, and switch between examples of superstring and gauge-gravity dualities in a way that is unclear and confused. Thus, we hope that this essay serves to clarify the most acceptable solutions to the problem of dualities, where these solutions are most reasonably applied, and the assumptions that go into them.

Of course, both the above-mentioned solutions rest on the need for future work, if they are to prove tenable. In the tripartite unificationist strategy, an underlying theory $\mathcal{T}^*$ needs to be found, both in the superstring duality case (where M-theory is currently the most plausible candidate), and the the gauge-gravity case (where no such candidate is forthcoming). On the approach which seeks to give metaphysical priority to one of a pair of dual theories, a principled philosophical reason for this choice needs to be given. Simply in virtue of the fact that (a) the duality relationship is formally symmetric, and (b) it does not appear that one of the dual models is always pathological in any obvious sense, we remain reserved about this latter approach. Thus, the unificationist strategy may offer the best hopes for making sense of dualities in a way that avoids any problematic underdetermination.

Acknowledgements

References

[1] Katrin Becker, Melanie Becker, and John Schwarz, "String Theory and M-Theory: A Modern Introduction", Cambridge University Press, 2007.

[2] Ralph Blumenhagen, Dieter Lüst, and Stefan Theisen, "Basic Concepts of String Theory", Springer Series in Theoretical and Mathematical Physics, 2013.

[3] Carolyn Brighouse, "Spacetime and Holes", in *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1994.

[4] Harvey R. Brown, "Physical Relativity: Space-Time Structure from a Dynamical Perspective", Oxford University Press, 2005.

[5] Jeremy Butterfield, "On Philosophy of Quantum Gravity", Seminar on the Philosophical Foundations of Quantum Gravity, Chicago, 2013.

[6] Ron Cowen, "Simulations Back Up Theory that Universe is a Hologram", Nature, December 2013.

[7] Richard Dawid, "Constructive Empiricism, Elementary Particle Physics and Scientific Motivation", 2005.

[8] Richard Dawid, "Scientific Realism in the Age of String Theory", Physics and Philosophy, 2007.

[9] Richard Dawid, "String Theory and the Scientific Method", Cambridge University Press, 2013.

[10] John Earman and John Norton, "What Price Spacetime Substantivalism?", British Journal for the Philosophy of Science 38, 1987.

[11] Valerio Faraoni and Edgard Gunzig, "Einstein Frame or Jordan Frame?", Int. J. Theor. Phys. 38, 1999.

[12] Bas van Fraassen, "The Scientific Image", Oxford University Press, 1980.

[13] Brian Greene, "The Hidden Reality", Alfred Knopf, 2011.

[14] Jeff Harvey, "Where is String Theory Today?", Seminar on the Philosophical Foundations of Quantum Gravity, Chicago, 2013.

[15] Richard Healey, "Gauging What's Real: The Conceptual Foundations of Contemporary Gauge Theories", Oxford University Press, 2007.

[16] Gerard t'Hooft, "On the Foundations of Superstring Theory", Foundations of Physics 43, 2013.

[17] Petr Horava and Edward Witten, "Heterotic and Type I String Dynamics from Eleven Dimensions", Nuclear Physics B 460, 1996.

[18] Leonard Horowitz and Joseph Polchinski, "Gauge/Gravity Duality", in *Approaches to Quantum Gravity: Toward a New Understanding of Space, Time and Matter*, D. Oriti, Cambridge University Press, 2009.

[19] Nick Huggett, "Philosophical Paths into String Theory", Seminar on the Philosophical Foundations of Quantum Gravity, Chicago, 2013.

[20] Nick Huggett, "The Emergence of Spacetime in Quantum Theories of Gravity" in Studies in History and Philosophy of Modern Physics 44, 2013.

[21] Nick Huggett and Christian Wüthrich, "Emergent spacetime and empirical (in)coherence", in Studies in History and Philosophy of Modern Physics 44, 2013.

[22] David Kaplan, "How to Russell a Frege-Church", Journal of Philosophy 72, 1975.

[23] Eleanor Knox, "The Dimensions of Duality", Seminar on the Philosophical Foundations of Quantum Gravity, Chicago, 2013.

[24] Juan Maldacena, "The Illusion of Gravity", Scientific American, November 2005.

[25] Juan Maldacena, "The Large N Limit of Superconformal Field Theories and Supergravity", Advances in Theoretical and Mathematical Physics 2, 1998.

[26] Keizo Matsubara, "Realism, Underdetermination and String Theory Dualities", Synthese, 2013.

[27] Tim Maudlin, "The Essence of Space-Time", in *Proceedings of the 1988 Biennial Meeting of the Philosophy of Science Association, Vol. 2*, A. Fine and J. Leplin (eds.), East Lansing, Michigan, Philosophy of Science Association, 1989.

[28] Charles Misner, Kip Thorne, and John Wheeler, "Gravitation", Freeman & Co., San Francisco, 1973.

[29] David Mitchell, "Cloud Atlas", Sceptre, 2004.

[30] F. A. Muller, "The Equivalence Myth of Quantum Mechanics – Part I", Studies in the History and Philosophy of Modern Physics 28, 1997.

[31] F. A. Muller, "The Equivalence Myth of Quantum Mechanics – Part II", Studies in the History and Philosophy of Modern Physics 28, 1997.

[32] John Norton, "The Hole Argument", in *The Stanford Encyclopaedia of Philosophy*, 2011.

[33] Daniele Oriti, "Approaches to Quantum gravity: Toward a New Understanding of Space, Time, and Matter", Cambridge University Press, 2009.

[34] Malcolm Perry, "Advanced String Theory: Part III Course Notes", 2014.

[35] Henri Poincaré, "Science and Hypothesis", New York (Dover), 1905 (1952).

[36] Henri Poincaré, "The Value of Science', translated by G.B. Halsted, 1914. Reprinted, New York: Dover, 1958.

[37] Joseph Polchinski, "String Theory, Vol. 1", Cambridge University Press, 1998.

[38] Oliver Pooley, "Substantivalist and Relationist Approaches to Spacetime", in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford University Press, 2013.

[39] Oliver Pooley, "CL 121: Advanced Philosophy of Physics", Advanced philosophy of physics lecture notes, 2012.

[40] Dean Rickles, "AdS/CFT Duality and the Emergence of Spacetime", Studies in History and Philosophy of Modern Physics 44, 2013.

[41] Dean Rickles, "Mirror Symmetry and Other Miracles in Superstring Theory", Foundations of Physics 43, 2013.

[42] Dean Rickles, "A Philosopher Looks at String Dualities", Studies in History and Philosophy of Modern Physics 42, 2011.

[43] Dean Rickles, "Quantum Gravity: A Primer for Philosophers", in *The Ashgate Companion to Contemporary Philosophy of Physics*, Ashgate, 2008.

[44] Carlo Rovelli, "A Critical Look at Strings", Foundations of Physics 43, 2013.

[45] Carlo Rovelli, "Halfway through the woods: Contemporary research on space and time", in *The Cosmos of Science*, J. Earman and J. Norton (eds.), University of Pittsburgh Press, Pittsburgh, 1997.

[46] Nathan Seiberg, "Emergent Spacetime", Rapporteur talk at the 23rd Solvay Conference in Physics, December 2005.

[47] Lee Smolin, "A Perspective on the Landscape Problem", Foundations of Physics 43, 2013.

[48] Leonard Susskind, "String Theory", Foundations of Physics 43, 2013.

[49] Nicholas Teh, "Holography and Emergence", in Studies in History and Philosophy of Modern Physics 44, 2013.

[50] Paul Townsend, "String Theory: Part III Course Notes", 2014.

[51] Cumrun Vafa, "Geometric Physics", in *Proceedings of the International Congress of Mathematics, Vol. 1*, G. Fischer and U. Rehmann (eds.), 1998.

[52] Tiziana Vistarini, "Emergent Spacetime in String Theory", PhD Thesis, Graduate College of the University of Illinois at Chicago, 2013.

[53] Edward Witten, "Duality, Spacetime and Quantum Mechanics", Physics Today, May 1997.

[54] Edward Witten, "Reflections on the Fate of Spacetime", Physics Today, April 1996.

[55] John Worrall, "Structural Realism: The Best of Both Worlds?", Dialectica 43, 1989.

[56] Alberto Zaffaroni, "Introduction to the AdS/CFT Correspondence", Class. Quantum Grav. 17, 2000.