CONVENTIONALITY, UNDERDETERMINATION AND THE INTER-THEORY TRANSPORT OF CONCEPTS

by

Andreas Kristoffer von Achen

Submitted to the department for Philosophy, Logic and Scientific Method for the degree of Doctor of Philosophy

The London School of Economics and Political Science

October 2023

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

Abstract

This thesis investigates two kinds of conventionalism in the context of two issues in the philosophy of spacetime: the Einstein Algebra formulation of General Relativity (GR) and the status of simultaneity in special relativity. The outcome of the analysis is that these two cases pull in different directions: I take a step back and analyse the strategy of breaking underdetermination by the invocation of what is often thought of as "non-epistemic" virtues. I argue that certain such virtues are more epistemically relevant than previously thought, in particular where these virtues have to do with the ability of a theory to "point ahead" towards new theories. This conclusion is that the underdetermination between the two formulations of GR only prima facie requires breaking by convention. On the other hand, a careful appraisal of the relativistic limit of Minkowski spacetime leads to the conclusion that relativistic simulaneity is in a precise sense so conventional so as to be devoid of content.

Acknowledgements

I would like to thank my main supervisor Bryan Roberts for patiently providing feedback on written work, for taking the time, many times, to discuss ideas, and for supporting me when I needed support. I would also like to thank Miklos Redei for valuable feedback in the finishing stages and Roman Frigg for helpful discussions in the beginning of my PhD journey. I owe James Wills and Grace Field my thanks for helping to uphold some degree of contact with the outside world during Covid through our reading group. My friend Rasmus Jaksland was kind enough to provide feedback on multiple drafts in various stages of completion and my friend Kevin Brix helped me wrap my head around tensors and functors. Every philosopher needs a mathematician friend. And last, my heartfelt gratitude for my wife Anjila and my son Samuel for their continuous and loving support, especially when the going got tough.

Contents

List of Figures

CHAPTER 1

Introduction

This thesis consists of three substantial chapters all centring on the topic of conventionalism. I distinguish between two kinds of conventionality: conventionality as a response to putative underdetermination and conventionality as arising when concepts from one theoretical context are transported to another. Chapter 2 provides a general analysis of the first kind while chapter 3 illustrates the approach from chapter 2 through a case study. Chapter 4 analyses the second type of conventionality in the context of the historic debate over relativistic simultaneity.

1. Geometry and Theory

Before the advances by such notable geometers as Gauss, Riemann and Lobachevsky in the 19th and early 20th centuries, the Kantian insistence on the pre-eminence of Euclidian geometry reigned unquestioned. Even after the proof of the co-consistency between Euclidian and various non-Euclidian geometries became well-known, the possibility of basing a theory of space and time on a non-Euclidian geometry remained purely theoretical. It was not until Einstein's 1915 publication "The Field Equations of Gravitation" that the practical relevance of non-Euclidian geometries for physics was established beyond doubt. The final version of the field equations contained in Einstein's paper formed the centrepiece of his theory of gravitation and pointed to two fundamental novelties: first, the spacetime continuum was modelled as a non-Euclidian geometry, and second, the dividing line between geometry and physics was blurred as a new link between the mathematical formalism and the physical interpretation was established by the so-called "geodesic principle" by which massive test particles would traverse along geodesics. Further, on Einstein's picture, spacetime geometry was dynamic and therefore the spatial geometry would change over time. This marked the departure from the modelling heuristic whereby physical laws were formulated on the basis of a passive spacetime container: Newton's spacetime had been a passive container whereas Einstein's was in a direct sense actively involved in the dynamic. The deviation from the Kantian insistence on the pre-eminence of Euclidian geometry raised the question of whether Einstein had realised a fact hitherto unbeknownst to humanity, namely that the physical spacetime continuum was non-Euclidian, or whether ascriptions of any particular geometry to physical spacetime should rather be seen as the result of Einstein exercising some sort of semantic freedom akin to the selection between German and French. The blurring of the dividing line between physics and geometry, on the other hand, drew into sharp focus arguments made by Poincaré 10 years earlier in his landmark "Science and Hypothesis". There, Poincaré had emphasised the inter-translatability of different geometries and made the case that one could compensate for a difference in geometry through a carefully designed difference in the physical laws. Poincaré had illustrated this with a famous thought experiment in which a disk-formed world allowed two different, yet both empirically adequate, descriptions. Either the disk was Euclidian while measuring sticks and light rays were disturbed by a carefully chosen temperature gradient and refractive index, or the world was hyperbolic while the physics were constructed without these delicate optical and thermal properties. In other words, a geometric preference for Euclidian geometry could in principle be honoured by making suitable adjustments to the physical laws[1]. Later in the 20th century, Reichenbach would continue this train of thought through his insistence on so-called[2] "universal forces".

The sense that the gulf between geometry as a mathematical discipline and geometry as the study of the physical spacetime might be greater than Kant and Newton had imagined became all the more acute when reputable researchers started to suggest that the categories of space and time themselves might be emergent rather than the bedrock foundations upon which everything else had to be built[3]. This idea put into perspective Geroch's 1972 paper, titled "Einstein Algebras", in which he had

[1]Einstein was arguably influenced by Poincaré's conventionalism. See Einstein (1921).

[2]Conventionalism with regards to spacetime geometry also played central roles in the writings of Schlick (1920) and Carnap (1922).

[3]See e.g. Cao and Carroll (2018); Cao et al. (2017) for attempts at reconstructing spacetime from quantum states. See Butterfield and Isham (1999) for an overview of the philosophical literature.

pointed out that the Lorentzian Manifolds, in terms of which General Relativity finds its modern formulation, could be substituted for algebras of an appropriate kind. If we deny the special status of geometry, Poincaré's question of what to do in the face of a multitude of geometries resurfaces as the more general question of what to do in the face of a multitude of formalisms (geometric or not!). This interpretation of the question of conventionalism as a reaction to the question of theory choice in the face of underdetermination is supported by Ben-menahem (2006, 1990),

> "Conventionalism, we saw, thrives on the underdetermination of theory." (Ben-menahem 1990, p. 278)

and more recently by Dürr and Read (2023, p. 6),

> "Underdetermination of geometry is a sine qua non for conventionalism, as studied here."

As such, the recent debate over conventionalism is linked to the question of the underdetermination of theories by evidence. In chapters 2 and 3 of this thesis, I analyse the antecedent question: do we in fact face a plurality of theories between which only convention can discern? To answer this question we need to know when two theories are to be counted as genuinely different as opposed to merely notational variants, and this means we have to revisit the literature on theory equivalence.

2. Conventionalism and Theory Equivalence

Chapter 2 develops a framework through which to understand the debate on conventionalism as a reaction to putative underdetermination. The central idea is to conceive of conventionalism through the lens of equivalence relations on the space of theories: by declaring certain features of a theory conventional, the conventionalist employs a weak equivalence relation, in the sense that even prima facie inequivalent theories might emerge as equivalent because whatever seemed to distinguish them is deemed conventional. I interpret various responses to underdetermination as suggesting different equivalence relations and argue that the current literature on theory equivalence spearheaded by Halvorson, Weatherall, Rosenstock and Barrett can be understood as offering an alternative to conventionalism in the face of putative underdetermination. Like the conventionalists, authors like Halvorson and Weatherall argue for a weak equivalence relation under which prima facie dissimilar theories emerge as equivalent, but rather than explaining away differences between theories as mere conventional elements, these authors argue that apparent cases of underdetermination are eliminated once we realise that the theories involved share mathematical structure. Therefore, both conventionalism and this more modern approach focused on mathematical structure are species of what I will refer to as "Elimination". Ultimately, I think Elimination fails in a number of relevant cases for reasons I lay out in chapter 2. Taking a cue from Putnam's 1974 article "The Refutation of Conventionalism", I continue the chapter by developing another alternative to conventionalism. Rather than eliminating relevant cases of apparent underdetermination by pointing out that the relevant theories are merely notational variants, this alternative approach seeks to break apparent underdetermination by widening the range of factors considered relevant for theory choice. Since this approach aims to discriminate between prima facie underdetermined options, I refer to it as "Discrimination". This approach employs a strong equivalence relation under which even putatively similar theories can emerge as inequivalent. I suggest three factors on the basis of which discrimination can take place and offer a number of examples to show how these factors can guide theory choice.

The thorny question in this context is how epistemically relevant are the reasons leading one to choose one theory over another[4]. The worry here is that if I simply choose the theory I like best for non-epistemically relevant reasons, underdetermination will re-emerge on a deeper level. Specifically, the choice of which idiosyncratic criteria for theory choice I come up with will be underdetermined and presumably have to be fixed by convention. On the other hand, the feeling is that if the choice is guided by such epistemically hardhitting values as "the pursuit of truth", then in

[4]In chapter two, I will argue for the epistemic forcefulness of heuristic factors. It is, of course, also possible to accept only the negative part of my argument to the effect that conventionalism often cannot make sense of the kind of work theories actually do. If one took conventionalism to be the last hope for realism, this option might be taken as a reductio of realism.

an important sense the world constraints the choice rather than any convention made by the theoriser: throwing darts will enable a choice, but it is difficult to argue that one's hands are tied. I argue that the factors I have pointed out are relevantly tied to truth.

3. Einstein Algebras

Whereas chapter 2 presents a general analysis of conventionalism as a reaction to putative underdetermination, chapter 3 proffers a case study of Lorentzian Manifolds and Einstein Algebras. The purpose of chapter 3 is to see the analysis from chapter 2 in action. The result of applying my analysis is that one should not count spacetime theories based on Lorentzian Manifolds as equivalent to spacetime theories based on Einstein Algebras despite the two frameworks being in a precise sense mathematically equivalent[5]. The chapter also contains a discussion of what one *can* reasonably conclude from a formal equivalence result and ties this back to the analysis from chapter 2.

4. The Conventionality of Simultaneity

While chapters 2 and 3 originate in the historic debate on the conventionality of geometry, chapter 4 visits another debate over conventionality: the conventionality of relativistic simultaneity. The dialectic in this literature consists of conventionalists proffering non-standard simultaneity relations on the presupposition that in the face of more than one option, only convention can decide. On the other hand, non-conventionalists provide arguments as to why each particular non-standard relation could not possibly be "simultaneity".

I take a step back and ask what "simultaneity" *could* mean in the context of Minkowski spacetime, and argue that, i) for the discussion to be meaningful, only certain answers can be acceptable and, ii) the only thing that can constrain the meaning of relativistic simultaneity is classical simultaneity. This leads to an analysis of the transport of concepts between theoretical contexts: presently, the transport of

[5]They are categorically dual (Rosenstock et al. 2015).

"simultaneity" from the context of Newtonian spacetime to the context of Minkowski spacetime. In the end, a careful appraisal of the classical limit of Minkowski spacetime results in the view that "simultaneity" is foreign to Minkowski spacetime.

The conventionality in play in chapters 2 and 3 is different from the conventionality in play in chapter 4. The question of the underdetermination of theory by evidence is premised on the availability of a number of different theories, each being the result of the free enterprise of a theorist. In particular, words such as "metric" are treated as freely available to be used however we please. We can call this "Poincaré-conventionality" or "P-conventionality" for short. On the other hand, in the case of transporting simultaneity from classical mechanics to special relativity, the term has baggage. Whereas P-conventionality refers to the connection between concept and world, this is a question about the connection between a name and a concept: can the old concept plausibly be transported to *that* concept in the new context? We can call this "transport-conventionality", or "t-conventionality" for short. In chapter 4, I offer a mechanism for the transport of concepts based on functionalism. Mathematics is ripe with useful examples, as follows.

4.1. An Example of the Transport of a Concept. The intuition undergirthing the modern concept of *continuity* is one of a function whose graph does not "jump". In the familar context of a function $f : (A, d_A) \to (B, d_B)$ between metric spaces, this idea is naturally expressed in terms of the well-known ϵ-δ formalism whereby f is continuous at $p \in A$ if $\forall \epsilon > 0\ \exists \delta > 0$ such that whenever $d_A(p, p') < \delta$ then $d_B(f(p), f(p')) < \epsilon$. Now, in an abstract topological space one abstracts away from the metric and retains only the idea of a collection of open sets. Does this mean that we cannot meaningfully talk of continuous functions between abstract topological spaces? Of course not! The idea of a collection of open sets is retained in the move from metric space to abstract topological space and so the mathematician readily observes that the definition of metric continuity is equivalent to the demand that $f^{-1}(U)$ be open in A whenever U is open in B. This definition makes sense in the general topological context where no metric structure is assumed, and it reduces to the metric notion of continuity when a metric is present. Is topological continuity conventional? That would depend on whether there are other realisers of continuity in the general topological case that, i) is equivalent to metric continuity in the metric case, but ii) is in-equivalent to the demand that the pre-images of open sets be open in the general case. As it happens, we can answer this in the positive: in the metric case, a function is continuous precisely if it takes convergent sequences to convergent sequences. Given a topology, we can make sense of convergence of sequences, and it is a theorem that a function $f : (X, \mathcal{T}_X) \to (Y, \mathcal{T}_Y)$ between topological spaces maps convergent sequences to convergent sequences if and only if the pre-image of each open set is open and $(X, \mathcal{T}_X)$ is first-countable. However, if we drop the latter assumption, the condition that the pre-image of open sets be open is strictly stronger than the assumption that convergent sequences be mapped to convergent sequences. This means that while continuity is non-conventional in the context of first countable topological spaces, it becomes conventional in the more general context of topological spaces not assumed to be first-countable.

4.2. Transport-Conventionality. The status of "non-t-conventional" is contingent on the theory in question: a concept can be non-t-conventional in theory T but t-conventional in theory T'. Note that both P- and t-conventionality play a role in the debate on simultaneity, corresponding to two different relations. P-conventionality is associated with the relation between the world and a particular relation on Minkowski spacetime—regardless of what we call it. On the other hand, t-conventionality is associated with the relation between a particular relation on Minkowski spacetime and the name "simultaneity". This description of t-conventionality immediately raises the question of whether this is merely a trivial semantic issue. As a matter of fact, Putnam dubs this type of conventionality "Trivial Semantic Conventionality" (TSM) and makes it clear that he finds it uninteresting. Though the intuition that it cannot matter what sound goes with what meaning is compelling, I do not share Putnam's view and as a matter of fact, I do not think Putnam himself is consistent in his dismissal. However, the purpose is not to engage in Putnam-exegesis and I will focus on my own reasons for sometimes taking seriously the question of matching meanings with sounds or strings of symbols. As before, the reasons have to do with scientific innovation: whether we use an old name for a new term makes a difference for the intuitions we have and ultimately for the kinds of generalisations or alterations we can imagine. And again, such differences are epistemologically potent as we shall soon see.

CHAPTER 2

Conventionalism

1. Introduction

Ben-menahem (2006) identifies two research programmes historically carried out under the banner of "conventionalism": an account of necessary truth and an analysis of the underdetermination of theories by evidence. Dürr and Read (2023) mount a defence for the continued relevance of the latter in the specific context of spacetime geometry and interpret conventionalism in this context as a selective anti-realism. The idea is that underdetermination of spacetime geometry by empirical evidence only amounts to a problem for the empirically-minded realist if geometric claims have truth values. By construing the geometry as conventional, the realist can save face since the claims that differ between different geometries reflect different modelling choices rather than genuinely different accounts of the target system. Thus, on this account conventionalism is wholly unmotivated in the absence of genuinely underdetermined options. If, for instance, prima facie cases of underdetermined options turn out to be merely notational variants, conventionalism does not get off the ground. On the other hand, maybe underdetermination can be broken by the invocation of "super/extra-empirical theory virtues, such as parsimony or unificatory power" (Dürr and Read 2023, p. 6). However, fixing the referent of some geometric term (e.g. "metric") seems cheaper than the operation required for saving the realist: throwing darts might select between two candidate geometries but will not alleviate the worry that the choice is wholly unrelated to truth. As Dürr and Read (2023, p. 6) write: "The crux for such a move is to render plausible the epistemic relevance of such virtues: ideally, one should demonstrate that they serve as reliable indicators of truth".

This observation is well-known in the conventionalism literature. For example, Putnam (1974) attempts a "refutation" of conventionality by the invocation of a particular extra-empirical virtue: coherence as simplicity, and argues that "simplicity is the mark of truth". But Ben-menahem (1990) responds by denying that "coherence" is epistemically relevant and concludes that Putnam's "refutation" merely amounts to an argument for the rationality of choosing the simpler of two theories when both are in accordance with experimental evidence. I agree with Ben-menahem (1990)'s assessment of the epistemological idleness of "coherence". In this chapter, however I will suggest certain other factors that might do a better job of breaking underdetermination in an epistemically potent way[1]. In particular, I will argue that viewing theories as structures that theorists adjust and change over time sometimes can break underdetermination in an epistemologically relevant way. Dürr and Read (2023) acknowledge one way that taking this perspective on theories might break underdetermination by following Laudan and Leplin (1991), who argue that it might be that a successor theory "reduces to, or contains in some other suitable sense" one of the original theories but not the other (Dürr and Read 2023, p. 7).

My proposal is sympathetic to Laudan and Leplin (1991) but adds to their account in a mayor way by providing a much-needed argument for the epistemological potency of taking theories to be living objects that change over time. Specifically, I argue that the project of interpreting current theories necessarily involves looking ahead towards successor theories. This means that differences between theories having to do with their ability to drive innovation cannot be showed in the pile of "non-epistemic" virtues but must be taken seriously. And this in turn implies that theories working differently for the purposes of constructing new theories ought to be differentiated

[1]Ben-menahem (1990) argues that Putnam's suggestion to use coherence to fix reference is an argument for the rationality of theory choice rather than an argument against conventionalism. As a matter of fact, Poincaré, arguably the father of modern conventionalism, writes in his seminal "Science and Hypothesis":

> It is clear that any fact can be generalized in an infinite number of ways, and it is a question of choice. The choice can only be guided by considerations of simplicity. (Poincaré 1929, p. 146)

Apparently, Poincaré himself would not object to Putnam's suggestion that coherence act as a guide in matters of theory choice.

as theories. This sometimes flies in the face of received wisdom, as we will see in a number of examples where differences in formalism result in differences in heuristic function.

Section 2 discusses Putnam (1974)'s "refutation" and distinguishes two different situations in which one might mistakenly believe one has been confronted with a conventional choice. This leads to the insight that conventionalism can be productively framed as a question of equivalence of theories. Section 3 discusses what makes conventionality either trivial or serious and presents the main contribution of the chapter: a novel equivalence relation on the space of theories called *constructive equivalence*. Section 4.1 then discusses constructive equivalence specifically in the context of different formulations of classical mechanics and relativity theory. Section 5 outlines the relationship between constructive equivalence and the background beliefs of scientists, and section 7 concludes.

2. Trivial and Serious Conventionality

Putnam (1974) sets out to refute the conventionalism found in Grünbaum's writings on space and time[2]. The strategy is to argue that the claims of his interlocutor reduce to triviality upon closer examination. In this section, I will describe Putnam's "refutation" of conventionalism.

2.1. Putnam's Refutation of Conventionalism. Complicating matters somewhat is the fact that Putnam (1974) misrepresents the view expressed in Grünbaum (1973) though not, I will argue, in a way that undermines Putnam's "refutation". First, I explain Putnam's argument and then I discuss the implications of his misrepresentation of Grünbaum.

Putnam (1974) reads Grünbaum as deriving the conventionality of the metric from a kind of multiple realizability:

> "The conclusion that Grünbaum draws from the situation just described is the following: There are certain axioms that any concept

[2]Putnam also claims that Quine's writings on radical translation exemplify what he calls "the conventionalist ploy" (Putnam 1974, pp. 28-31). See Quine (1960).

> of distance, that is to say, any metric, has to satisfy. For example, for any point x in the space, the distance from x to x is zero; for any points x and y in the space, the distance from x to y equals the distance from y to x; for any three points in the space x, y, z, the distance from x to y plus the distance from y to z is greater than or equal to the distance from x to z; distance is always a non-negative number; the distance from x to y is zero if and only if x is identical with y. But any continuous space that can be metricized at all, i.e., over which it is possible to define a concept of distance satisfying these and similar axioms, can be metricized in infinitely many different ways. (Putnam 1974, p. 27)

Putnam reasons as follows: What we mean by "distance" is exhausted by the metric axioms plus the requirement that the metric is compatible with the topology[3]. But these requirements do not suffice to fix a unique metric. If we desire any particular metric on our manifold, which of course we do when we do physics, we must pick it out as a matter of convention.

Putnam's proposed refutation of this view (which he fallaciously attributes to Grünbaum) is based on the following observation: just because the manifold admits of more than one metric compatible with the topology does not mean that one of these metrics does not stand out as the most expedient one for the purposes of theory-construction.

> We try to formulate total science in such a way as to maximize internal and external coherence. By internal coherence, I mean such matters as simplicity, and agreement with intuition. By external coherence, I mean agreement with experimental checks. But Grunbaum certainly has not proved that there are two such formulations of total science leading to two different metrics for physical space-time. (Putnam 1974, p. 33)

[3]Grünbaum *does* believe the topology to be empirically determined (Grünbaum 1973, p. 336).

In other words, we formulate our physical laws on the basis of the metric and we should not expect these laws to be equally simple across all different choices. If a particular metric makes the laws especially simple, Putnam proposes that we simply say that *that one* is what we refer to[4] when we say "metric" (Putnam 1974, p. 34).

Taking one step back, Putnam asks why we should accept Grünbaum's claim that *these* requirements exhaust the meaning of "metric" and suggests we add considerations pertaining to simplicity and intuition. Schematically, we can say that where axioms and compatibility with a topology fail to uniquely specify a metric, then perhaps axioms, compatiblity *and* expedience might succeed. The fact that we could have meant something different when we said "metric" is simply a reflection of the semantic freedom always inherent in the matching up between sounds or symbols on the one hand, and parts of reality on the other. The conventionality therefore is of the trivial sort.

There is one problem, however. Grünbaum (1973) specifically denies that the existence of alternative metrics is what makes the choice of metric conventional:

> For convention-ladenness arises from the lack of an intrinsic basis and not from the existence of an alternative metric! (Grünbaum 1973, p. 560)

To understand the implications of this misrepresentation for the strength of Putnam's argument, we have to understand what Grünbaum has in mind when he speaks of an "intrinsic basis". First, Grünbaum (1973, pp. 505-506) defines when an element of a manifold is *internal* to a given interval[5] of the manifold,

> (1) Given the elements of a manifold, we shall speak of an entity as being "internal" to an interval of the manifold (or as being an "inside" entity with respect to the interval), iff the existence of the interval depends on the existence of the entity. Thus every element belonging to an interval [a, b] is internal to [a, b] in this sense,

[4]Of course, the assumption here is that all the options are *externally coherent*, i.e. in accordance with experimental checks.

[5]Grünbaum seems to be picturing the manifold $\mathbb{R}$, but it is straightforward to generalise to general subsets of an arbitrary manifold.

> whether [a, b] is an interval of some P-manifold or of the arithmetic manifold of the real numbers. (Grünbaum 1973, pp. 505-506)

With regard to a property being *external* to an interval of a manifold, Grünbaum explains,

> (2) Now, in a given manifold, a monadic property P is said to be "external" to an interval possessing it, iff the obtaining of P depends on entities which are not internal to the interval.

Grünbaum goes on to define an *intrinsic property* of an interval as a property, i) whose constitution does not depend on any particulars, and ii) which is not external to the interval. This definition, Grünbaum insists, makes the properties relating to any particular metric extrinsic to the intervals they are about (Grünbaum 1973, p. 506). Grünbaum highlights that the "intrinsic metric amorphousness" (ima)[6] of the manifolds of GR stems directly from the cardinality being continous[7]. However, the purpose here is not to engage in Grünbaum-exegesis and whether he succeeds in presenting an argment from his definition of "intrinsic" to the claim that continous manifolds are necessarily ima, ultimately does not matter here. Our question, rather, is whether Putnam's misrepresentation of Grünbaum's view undermines the strength of the refutation. I will argue that it does not. The difference between Putnam's target and what Grünbaum actually says is that the latter demands a different starting point. Rather than starting with the metric axioms and a topology as Putnam imagines, Grünbaum's starting point is what the manifold has the resources to express "intrinsically". Grünbaum's argument is not that since the topology can be metricised in more than one way, only convention can establish a unique referent for the word "metric", but rather that what the manifold does not have the resources to express "intrinsically" must be established by convention. That means that when Putnam responds that axioms, compatibility with a topology and coherence together *can* establish a unique metric, he is talking past purposes. But we can simply amend Putnam's "refutation" to state that a unique metric can be established from what

[6]Grünbaum's denomination for those manifolds lacking an "intrinsic" metric.
[7]Presumably, the counting-metric would be intrinsic to an interval of finite cardinality.

the manifold has the resources to express "intrinsically" together with coherence instead. So rather than the refutation resting on the arbitrariness of restricting oneself to axioms and compatibility, it rests on the arbitrariness of restricting oneself to what Grünbaum refers to as "intrinsic" to the manifold. If Putnam's approach successfully establishes the triviality of the conventionality in the former context, it should do so successfully in the latter. The question therefore becomes: does Putnam successfully establish that the conventionality in play is of the trivial kind? To answer this, we must investigate what makes conventionality either "trivial" or "serious". But first, I discuss the appropriate level of analysis.

2.2. Theories and Concepts. So far, we have been discussing conventionality at the level of individual concepts, e.g. "metric", but it seems this narrow focus is inappropriate for Putnam's overall approach. This can be most easily seen in the way Putnam wants to avoid conventionality with the help of coherence. With "coherence" Putnam has in mind both external coherence, i.e. empirical adequacy, and internal coherence, i.e. simplicity. However, there is something strange in claiming that a single concept is simple, since "simple" typically refers to a property of a system. A theory can be simple but an individual concept cannot. Thus, "coherence" at the level of concepts cannot work as a refuter of conventionality.

Suppose we did use, say, "schmistance" instead of "distance" and continued to build our physical theory of space and time on this alternative concept of length. Let us try to use Putnam's scheme to avoid serious conventionality. As we just saw, we cannot apply the test of coherence to a single concept, and thus we have to focus instead on the theory resulting from our adoption of "schmistance", which we can call "Schmeneral Relativity" (S-GR). And let us take the theory resulting from the adoption of "distance" to be ordinary General Relativity (GR). Since S-GR is created by adopting an alternative concept of distance, we must expect that the vocabulary of S-GR will have to differ from that of ordinary GR throughout to make the two theories empirically equivalent. In this way, Putnam's maxim that we let coherence fix reference trades in the problem of conventionality of the reference of a single concept for the problem of theory choice, and the question of equivalence on the level of individual concepts for the question of the equivalence of theories[8].

On this reading, Grünbaum claims that the manifold underdetermines the metric and Putnam responds that availing ourselves of "coherence" breaks the underdetermination[9]. I will follow this interpretation broadly, but we will need to add some detail. First, since the conclusion of Putnam (1974) is that the claims made by conventionalists reduce to triviality, we must analyse what makes conventionality either serious or trivial. This leads to the main idea of the paper: that the question of conventionalism can be productively framed as the question of which equivalence relation is appropriate on the space of scientific theories. This in turn implies the existence of two ways in which one might mistakenly believe oneself to be faced with conventionality. I point out, for each of these, a strategy to avoid being fooled.

2.3. Two Ways of Avoiding being Fooled: Elimination and Discrimination. Serious underdetermination requires the underdetermined options to reside in the goldilocks zone. Options need to be close enough to threaten underdetermination by at least being observationally equivalent, but not so close as to be merely notational variants. This implies that there are two ways of mistakenly believing that one is facing underdetermination: first, one might have overlooked the fact that, upon closer inspection, the available options are not *really* underdetermined. This would involve the identification of one or more hitherto underappreciated factors that could help one discriminate. I will call this strategy for avoiding to be fooled "Discrimination" (Discr) since the confusion is remedied by discriminating between options that were prima facie on a par. The insight behind Discr is that one can avoid mistaking the case where we are actually able to discriminate between two theories for genuine undertermination by adopting a stronger equivalence relation based on features of the theories in question that we can identify and therefore use as the basis for discriminating between them.

[8]Note that the reference of individual concepts such as "metric" will be fixed ipso facto when the theoretical context is fixed.

[9]This is in the spirit of Ben-menahem (2006) who takes conventionality to be a question of underdetermination (Ben-menahem 2006, pp. 7-12).

Second, one might have overlooked that the options realising the supposed underdetermination are not actually different: rather, they are simply different notational variants. I call this strategy for avoiding being fooled “Elimination” (Elim) since the confusion is remedied by eliminating the problematic choice altogether. The insight behind Elim is that one can avoid mistaking the existence of multiple notational variants of the same theory for genuine underdetermination by adopting a weak equivalence relation inducing a partition of the space of theories into large equivalence classes.

For example, the fact that “il y’a une tasse sur la table”, “there is a cup on the table” and “der er en kop på bordet” apply equally well to my desk at the moment does not mean that the state of my desk is somehow underdetermined by the evidence I gleaned by looking. It simply means that the same state of affairs can be expressed in three different ways. What makes the underdetermination trivial in this instance is that the three options are not actually substantially different. Hence, the underdetermination is trivial.

Another example of trivial underdetermination is in the toy example where we are simply considering swapping the referents of “cat” and “mouse”. Here, one would probably want to say that the two resulting “theories” are merely notational variants. A straightforward application of Elim yields the desired conclusion: the cat-theory and the mouse-theory are identified by the appropriate equivalence relation. Conventionalism itself can be understood in this way: the prima facie differences between two formalisms are explained away as conventional[10]. Of course, when we are choosing between notational variants, any conventionality involved is of the trivial variant.

But in the realistic case where one is confronted with two entire theories, e.g. GR and S-GR, it is less clear how to determine whether we are confronted with a bona fide choice between different theories or merely notational variants of the same theory. Thus, it is difficult to say whether Elim is appropriate. In particular, it will no longer be the case that the reference is known: GR comes with reference to

[10]See e.g. Dürr and Read (2023) who take conventionalism to be a “selective anti-realism”.

"distance" but not "schmistance" and vice versa, so here it is not simply a question of label switching.

I will argue in section 3.2 that the recent literature on theory equivalence can be interpreted as an attempt at a more sophisticated version of Elim specifically aimed at not getting fooled by different-looking mathematical formalisms that are in fact in some sense equivalent. The underdetermination is then broken by pointing out that would-be pairs of theories realising this underdetermination really are just different formulations of the same theory. Naturally, the choice between two notational variants of the same theory involves only trivial conventionality. In the next section, we discuss this distinction between trivial and serious conventionality.

3. Conventionalism and Theory Equivalence

3.1. The Trivial and the Serious. Let us follow Putnam (1974) in using the phrase "Trivial Semantic Conventionality" (TSC) to refer to the inherent conventional element in the proces of establishing connections between parts of language and parts of reality. Nothing intrinsic to the string of symbols "mouse" nor the associated sound bit in the English language makes it refer to the household rodent. This means that we enjoy a certain semantic freedom: we would not be commiting any linguistic nor logical mistakes if we systematically switched the referents of "cat" and "mouse". In the cat-and-mouse example, the categories are fixed and the semantic freedom pertains only to which label goes with which category. But in reality our linguistic freedoms far surpass the permutation of labels among already-defined categories, as we saw in the example with "distance" and "schmistance": "schmistance" does not feature in GR nor does "distance" feature in S-GR, but this fact alone should not make us conclude that the two theories are genuinely different. Reality does not come pre-carved, and as a result, there is a real question of when two semantic conventions differ in only trivial ways.

So, when is underdetermination "serious"? To answer this question, we will have to start conceiving of underdetermination not as a predicate but as a binary relation. A choice is not underdetermined simplicitér but rather underdetermined *with* *respect to* a certain class of factors. To see why no choice is underdetermined simplicitèr simply realise that a choice always can be made by throwing darts or drawing lots. Whether the underdetermination is "trivial" or "serious" has everything to do with the class of factors with reference to which a choice is possible. Serious underdetermination comes about when options differ in substantial ways but those substantial ways are inaccessible to us so we cannot use them as reasons for choosing one option over the other. We might think, for instance, that geometrised Newtonian theory (GNT) differs from ordinary newtonian force mechanics in a substantial way because the former ascribes non-zero curvature to space whereas the latter does not. Further, the two theories are observationally equivalent and so the underdetermination is substantial. By contrast, trivial underdetermination occurs when the difference between options is merely aestethic, as, for instance, found in textbooks in two different languages.

We just saw how Putnam's insistence that coherence can help fix the reference of terms leads us to trade in the conventionality of denotation for theory choice. In the next section, I situate Putnam's argument in the modern literature on theory equivalence. We start by surveying four attempts at using the powers of logic and category theory to operationalise Elim by providing an appropriate equivalence relation based on formal structure.

3.2. Some Accounts of Theory Equivalence. Recently, a number of formal criteria of theory equivalence have been proposed in the literature. But, before laying them out, a comment is in order. It is notoriously difficult to get a handle on scientific theories in the wild, and therefore philosophers of science often engage in a bait-and-switch manoeuvre in which the actual object of investigation is what "theory" means in books on categorical logic and model theory, while the results of the analysis are taken to apply to scientific theories in the wild. Even on the most generous reading, it must be acknowledged that real scientific theories are not formulated in purely formal languages, and that some sort of non-trivial translation-scheme must therefore be invoked prior to the application of any formal criterion of theory equivalence, at least if the results are to have any ramifications for theories actually employed by scientists. So, the present discussion must proceed on the assumption that an appropriate choice of translation-scheme has been settled.

Naively, one might suggest logical equivalence[11] as the natural, formal criterion for theory equivalence. But, two theories can only be logically equivalent if they are formulated in the same signature,[12] and this is simply an implausibly strong requirement to make. Barrett and Halvorson (2016a) give the example of the theory of groups, which can either be formulated using a binary operation $\cdot$ and a constant symbol e, or with a binary operation $\cdot$ and a unary function -1 encoding inversion with respect to the binary operation[13]. We would not want to claim that these two theories are actually different, but they cannot be logically equivalent since they are written in different signatures. *Definitional Equivalence* (DE) represents an attempt at remedying this shortcoming.

The idea behind DE is that two theories T_1 and T_2 formulated in signatures Σ_1 and Σ_2 are equivalent if they can be simultaneously expanded to theories T_1^+ and T_2^+ (formulated in the same signatures as T_1 and T_2, respectively) in such a way that, 1) T_i^+ does not "say anything more than T_i" for $i = 1, 2$ and, 2) T_1^+ and T_2^+ are logically equivalent as $\Sigma_1 \cup \Sigma_2$-theories[14] (Barrett and Halvorson 2016b, p. 7). Of course, "$x \cdots$ does not say anything more than $\cdots y$" needs to be made precise, but it essentially means that all new function-symbols, predicate-symbols and constant-symbols introduced in the move from the original signatures to the union $\Sigma_1 \cup \Sigma_2$ were already definable in the original signatures, and that everything that is valid in T_i remains so in T_i^+ (Barrett and Halvorson 2016b, p. 6). The important thing to note here is that even though the signatures Σ_1 and Σ_2 are allowed to be different, they

[11] Theories T_1 and T_2 formulated in the signature Σ are *logically equivalent* if $Mod(T_1) = Mod(T_2)$, i.e. if the two theories have the same class of models. Equivalently, if for all Σ-formulas ϕ $T_1 \models \phi$ if and only if $T_2 \models \phi$. See Barrett and Halvorson (2016a, p. 468) for more.

[12] The signature Σ for a first order language is the set of predicate symbols, function symbols and constants(Barrett and Halvorson 2016a, p.468).

[13] For example, on the first formalisation, one can use the neutral element to define inversion: $\forall x \exists x$ such that $x \cdot y = y \cdot x = e$. On the second formalisation, one can use inversion to define the neutral element: $\exists x \forall y$ such that $-1(y) \cdot y = y \cdot -1(y) = x$.

[14] As both Σ_1 and Σ_2 are subsets of $\Sigma_1 \cup \Sigma_2$, we can view both T_1^+ and T_2^+ as formulated in the signature $\Sigma_1 \cup \Sigma_2$.

have to include the same sort-symbols[15] (Hudetz 2019, p. 49). The latter limitation is somewhat undesirable as the same theory often has equivalent formulations using either one sort or multiple ones[16].

Generalized Definitional Equivalence (GDE) (or Morita Equivalence) remedies this shortcoming by allowing the definition of new sorts on top of new functions, predicates and constants. The concrete way in which this is achieved is somewhat technical and the interested reader is referred to Barrett and Halvorson (2016b, p. 9) or Hudetz (2019, p. 49) for the details.

For most theories in the wild, it is difficult to see how one could translate them into the kind of languages necessary for the application of DE and GDE. This motivates the use of category theory. Given first-order theories T_1 and T_2, one can often form their associated categories of models $Cat(T_1)$ and $Cat(T_2)$, and it is standard practice to discuss the class of models of a theory. If T is a first order theory, the morphisms of $Cat(T)$ are typically taken to be elementary embeddings, i.e. maps $h : N \to M$ between the domains associated with the models M and N preserving satisfaction (Harnik 2011, p. 83). Depending on the theories, the relevant criterion of equivalence for categories will either be Categorical Equivalence (CE) or Categorical Duality (Dual), and we thus say that theories T_1 and T_2 are Categorically Equivalent, or Dual, if the associated categories $Cat(T_1)$ and $Cat(T_2)$ are equivalent or dual, respectively. We say that categories $\mathbb{C}$ and $\mathbb{D}$ are *equivalent* if there exist functors $F : \mathbb{C} \to \mathbb{D}$ and $G : \mathbb{D} \to \mathbb{C}$ such that $FG \cong \mathbf{1}_{\mathcal{D}}$ and $GF \cong \mathbf{1}_{\mathcal{C}}$. Two functors are isomorphic if they are related by a natural isomorphism (Awodey 2006, p. 134). The categories are dual just in case the opposite category[17] of one is equivalent to the other.

Sometimes duality is the appropriate notion of equivalence between categories. An example of this is the case of RING, the category of rings and ring-homomorphisms,

---

[15]A *sort symbol* is part of the formal language. For example, in epistemic logic one routinely operates with the sorts "agent" and "object" and category theory finds a natural formalization using the sorts "object" and "morphism". See Manzano and Aranda (2022).

[16]One example being category theory that allows the natural two-sorted formulation with objects and morphisms and a mono-sorted formulation with only morphisms (Hudetz 2017).

[17]The opposite category is obtained by "turning around" all morphisms.

and the category TOP of topological spaces and continuous maps. Any continuous map $f : (X, \mathcal{T}_X) \to (Y, \mathcal{T}_Y)$ corresponds to a map $g : C(Y) \to C(X)$ between the associated rings of continuous, real-valued functions on X and[18] Y. Thus RING and TOP are dual categories[19]. The duality between a category of structured sets[20] and the category in which the objects are sets of real-valued homomorphisms on said structured spaces is quite general. An example of particular relevance here is the categories Lor of Lorentzian Manifolds and isometries, and EA of Einstein Algebras and their homomorphisms. Rosenstock et al. (2015) show these two categories to be dual, and argue that this makes the Einstein Algebra formulaton of GR theoretically equivalent to the geometric formulation[21]:

> Both [Einstein Algebras and Lorentzian manifolds, red.] encode precisely the same physical facts about the world, in somewhat different languages. It seems far more philosophically interesting to recognize that the world may admit of such different, but equally good, descriptions than to argue about which approach is primary. (Rosenstock et al. 2015, p. 17)

Finally, Hudetz (2019) points out that categorical equivalence and categorical duality are inappropriately insensitive to the internal structure of the models. To illustrate this, Hudetz (2019) gives the example of the categories FinVec of finite vetor spaces and linear maps, and Num where the objects are natural numbers and the morphisms between objects n and m are $m \times n$ matrices over[22] $\mathbb{R}$. FinVec and Num are categorically equivalent, but for modeling purposes, it seems to make a real difference whether we are working with vector spaces or natural numbers. This is because all that matters for a category is the web of objects and morphisms, and not

[18] This correspondance is realized by composition with f.
[19] See e.g. Leinster (2014, p. 23) or Hudetz (2019, p. 52).
[20] And the relevant structure preserving maps as morphisms.
[21] We return to this example in a later chapter.
[22] Composition of morphisms is implemented as matrix multiplication: if $f : m \to n$ and $g : n \to k$ are morphisms, the composition $g \circ f$ simply is the matrix-matrix product $g \cdot f$. Since f is an $n \times m$ matrix and g is an $k \times n$ matrix, the product $g \cdot f$ is a $k \times m$ matrix as required for a morphism from m to k.

what is "inside" the individual object[23]. To remedy this shortcoming, Hudetz (2019) introduces a fourth formal account meant to combine the strengths of CE and GDE. The idea is to augment CE by adding the requirement that the functors realising the equivalence be "reconstruction functors", roughly meaning that for each model $M \in Cat(T_1)$ the image $F(M) \in Cat(T_2)$ is constructable from M in the signature of T_1. Of course, "$x \cdots$ is constructable from $\cdots y$" needs to be made precise. For the specifics of this construction, see Hudetz (2019, pp. 54-56).

All of the above criteria represent attempts at arguing that prima facie cases of underdetermination reduce to cases of mere notational variants of the same theory. As such, Halvorson, Weatherall and others attempt to solve the problem of underdetermination by operationalising Elim using the resources of logic and category theory. The purpose of Elim is to avoid being fooled by the existence of different ways of representing the same mathematical structure: each of the four criteria induce an equivalence relation on the space of theories that can reveal when two formalisms are really equivalent despite surface differences. On the other hand, the point of Discr is to avoid the mistaken belief that only a conventional choice can break the underdetermination when in fact epistemically potent reasons for preferring one option are available.

Putnam's "refutation" can be viewed as a version of Discr that spells out these "extra factors" as internal and external coherence. There are good reasons to be sceptical of the idea of basing theory choice on coherence[24], but I will argue that there is something fundamentally on point in Putnam's account in his insistence that we consider factors sometimes referred to as "non-epistemic"[25]. So even though "coherence" is inadequate in this respect, maybe we can find other criteria that fare better. In the next sections, I will develop a "neo-Putnamian" answer to the question

[23]The objects of a category do not need to be sets, so we might not even have access to the membership relation ϵ.

[24]Putnam takes "external coherence" to mean "empirical adequacy" but in all interesting cases of supposed underdetermination all candidates will be empirically adequate. "Internal coherence" means "simplicity", but I am doubtful that we can find a unique measure of simplicity.

[25]Though I will argue that "non-epistemic" is a misnomer in section 3.3.

of conventionality as a version of Discr where the "extra factors" follow Putnam's spirit but deviate from his letter.

The present argument is not that this neo-Putnamian version of Discr should be chosen in all cases but rather that it often will be a viable strategy, and we will see a number of examples of prima facie cases of conventionality that can be defused by applying it. A corollary will be that we should be wary of claims to the sufficiency of formal considerations for theory equivalence. This is because these "extra" factors are such that any formal criterion will be insensitive to them: at least as long as we believe any formalism to allow for at least two different interpretations, any criterion that only deals with the formalism will in principle be unable to detect some differences in interpretation.

To develop our approach to conventionalism, we need to analyse ways in which theories can differ to get a basis for discriminating between them. This analysis starts with a look at different aims scientists might be pursuing and the strategies that might help fulfil them. This will give rise to a novel equivalence relation on the space of theories that I will dub "constructive equivalence". This will be the content of section 3.4. First, we must investigate what theories do in order to get a basis for discriminating between them.

3.3. What are Scientific Theories Good For? I would like to propose an equivalence relation for the space of theories motivated by two considerations: first, the functionalist idea that an entity can be specified by what it does. In the context of scientific theories, this means that if we list the purposes for which scientific theories are devised, we get a list of ways in which they can differ. This shows how functionalist considerations are naturally associated with Discr, in that focusing on what theories do yields a basis upon which one can discriminate between them. The second consideration guiding the present analysis is that the distinction between what is sometimes referred to as "epistemic" values and "non-epistemic" values becomes tenous for actors with broadly speaking realist intuitions dealing with theories known to be incomplete[26]. This means that the factors we will identify as the basis of differentiation between theories cannot be shoved into the pile of "pragmatic" reasons or otherwise be shrugged off as secondary to reasons having to do with what a theory actually says about reality.

So, what do scientific theories do? Consider two different aims that scientists might be following: first, "standard realism", the aim of investigating what the world is like according to known theories. After the publication of Einstein's field equations in 1915, scientists undertook the work of searching for solutions and investigating the nature of such solutions. This research programme gave birth to the idea of a black hole when Schwarzschild published his famous solution showing how a singularity of curvature was allowed by Einstein's theory (Schwarzschild 1916; Hawking, S.W. and Ellis 1973). This is a prime example of standard realism where a known theory is investigated for previously unknown consequences[27].

A second aim is "construction", the aim of constructing new theories. Construction aims at the development of new theories rather than the exploration of old ones. Where standard realism views theories as static objects, construction recognises that theories change, develop and get transformed[28].

However, the two aims are not meant to be mutually exclusive nor generally independent and likely the boundary between them is fuzzy. For example, one could argue that GR is a research programme rather than a theory and that Schwarzschild's contribution quoted above should therefore fall under "construction". This line of thought can be expanded by the following considerations: since the realist knows that none of our current best theories are representationally complete, their task is not one of finding the "true" one. Rather, guided by their desire to explain empirical success by truth, the realist is in the business of trying to identify which parts of our current theories that are responsible for those theories' success. The line between

[26]As the question of conventionality is moot for instrumentalists, realists of different stripes are the relevant group, and I will simply assume that we are talking about this group of actors in all that follows.

[27]See e.g. Psillos (1999) for a reference on the standard view.

[28]The Lakatosian focus on research projects seems to fit this picture. In particular, the idea that theoretical projects are individuated by heuristics rather than the specific instantiation at some given time.

good and bad strikes through the heart of every theory. The good, or efficacious elements, will in turn be the ones the realist will seek to preserve over a coming theory change. As such, their investigation of current theories is simultaneously laying the grounds for innovation. This connection between the investigation of current theories and the development of new ones becomes even stronger once we realise that the best way of separating good from bad is to theorise[29]. The scientist will isolate the elements they believe are responsible for the success of the theory and use these as basis for further theorising. If the scientist's attempts are succesful and produce theoretical advances, then that will constitute support for their hypothesis regarding which elements were responsible for the successes of the older theory. Once a viable successor theory is developed, the scientist will then want to use the newer theory to explain the success and limitations of the older theory. In physics, this often happens as the older theory is re-located as an appropriate limit of[30] or is succesfully reduced to a successor theory[31]. This highlights how the investigation of current theories is intimately linked to the development of new ones, and therefore, how the features of theories mediating differences in heuristic function are epistemically potent: we only really understand current theories when they are no longer current, and so the project of investigating what "a theory says about the world" is necessarily a project of developing new theories. In the following, we will see examples of different formalisms working differently for the purposes of construction, with the implication that differences in formalism can be epistemically relevant – sometimes despite the formalisms being in some sense mathematically equivalent.

Three strategies associated with construction are worth discussing here. The scientist will seek a formulation that will allow them to develop and express their ideas as easily and effectively as possible. We can call this strategy "clear expression". Second, when faced with the task of unifying different theories or theory fragments, the scientist might try to reformulate theories to make the formalisms more alike. We can

[29]This paints a picture of science as continuously open. For another view emphasising the openness of science, see Norton (2014).
[30]See Feintzeig (2018, 2020) for a discussion of the classical limit of quantum mechanics. See Malament (1986) for a discussion of the classical limit of relativity theory.
[31]See e.g. Dizadji-Bahmani et al. (2010).

call this strategy "uniformity". A close cousin of uniformity is the strategy of selecting a formalism that allows for generalisation in order to facilitate construction. We can call this strategy "generalisability". It is worth noting that what counts as clear expression is intrinsically contingent on the guiding heuristics of the individual scientist whereas uniformity and generalisability are formalism-centric. In the next section we will see how different formulations of classical mechanics illustrate construction and the associated strategies discussed here.

Recall that our purpose is to develop a "neo-Putnamian" version of Discr based on a functionalist analysis of the roles played by scientific theories. In particular, we have identified "standard realism" and "construction" as two functional roles played by theories: the investigation of what the world is like, and the construction of new theories, respectively. Further, I have argued that these two roles are intimately linked and discussed strategies associated with the fulfilment of these roles. The next step is to take these functional roles seriously. For not only have we already seen how the aims of standard realism and construction are epistemically linked, a much more pedestrian observation applies: if a central role of theories is the development of new theories, then two theories that fill this role in different ways are not equivalent as theories. This specifically means that we need to investigate the strategies via which the aim of construction might be pursued and how this analysis results in a novel equivalence relation on the space of scientific theories.

3.4. A New Equivalence Relation. Recall that we have taken conventionalism to be a question of underdetermination and that the fundamental idea of Putnam's refutation is to let "coherence" help break this underdetermination. In the language developed here, this means that Putnam is following a strategy of discrimination. However, "coherence" is undesirable as a basis for discrimination since properties such as "simplicity" depend crucially on the concrete measure chosen and comparitive judgements (theory A is simpler than theory B) cannot be expexted to be stable under changes of measure. In what follows, I will develop an approach that recognises the idea of identifying factors that can help discriminate in cases of seemingly underdetermined options, but which substitutes "coherence" for properties desirable for theory construction.

Viewing scientific theories through a functionalist lens, we have identified a number of roles filled by theories together with strategies through which theories can fill these roles better. The neo-Putnamian refutation of conventionalism seeks to discriminate theories on the basis of fit between formalism, the scientist doing the theorising and their aims, pursued via the three strategies:

- Clear expression
- Generalisability
- Uniformity.

This list is not meant to be exhaustive[32] but to highlight three items of special interest. Note also that the claim is not that this approach to conventionalism will break underdetermination in every single case. In some instances, Elim will be the appropriate strategy and perhaps there are some cases of genuine underdetermination. Rather, the claim is that many of the examples held up by proponents of conventionalism are not, on closer analysis, substantial and that this can be realised using the approach outlined here.

On this account, theory choice is explicitly contingent on the aims, beliefs and guiding heuristics of the individual scientist. This means that two different scientists might not make the same judgements. Does this make underdetermination subjective? No, it does not. Rather, it makes theory choice context relative. This motivates the first of two novel equivalence relations

Definition 1 (R-Constructive Equivalence). Theories T_1 and T_2 are *constructively equivalent relative to research programme R*, if and only if, with respect to R, T_1 and T_2 function equally well with respect to clear expression, generalisability and uniformity. When this is the case, we write $T_1 \sim_{RC} T_2$.

[32]For example, Glymour (1977) and later Laudan and Leplin (1991) are "discriminators" in their approaches to questions of underdetermination, but the basis for discrimination is not on my list. Rather, their strategy is to argue that even though empirically equivalent theories will be confirmed by the same evidence, the same piece of evidence might confirm on theory more than another despite them being empirically equivalent.

R-constructive equivalence is, as the name suggests, equivalence relative to a research programme. This level mirrors the choices actually made by scientists engaged in research. On the other hand, for the purposes of individuating theories on the basis of the kind of epistemologically hard-hitting reasons discussed presently, we need a more high-level notion of constructive equivalence. That theory T_1 is preferable to Mette compared to T_2 is an objective fact even though the reason this is the case is contingent on what Mette believes about the world. This means that the equivalence relation this suggests on the space of scientific theories should distinguish between T_1 and T_2 even though for *some* scientists there may be no reasons for preferring one over the other. Even though the left-hand scissors and the right-hand scissors are equivalent for the ambidextrous, left-hand scissors are different as tools from right-hand scissors: it makes a difference for everyone that it makes a difference for someone. Let us close this section with the central definition,

Definition 2 (Constructive Equivalence Simplicitér). Theories T_1 and T_2 are *constructively equivalent* if and only if, for all research programmes, T_1 and T_2 function equally well with respect to clear expression, generalisability and uniformity. Alternatively, if and only if, for all relevant research programmes R, we have $T_1 \sim_{RC} T_2$. When this is the case, we write $T_1 \sim_C T_2$.

Because of the quantification over all research programmes[33], this equivalence relation cuts finely and possibly there are no interesting cases of equivalent theories. However, I believe this is at it should be. First, science is a collective effort which is reflected in the quantification over *all* the relevant research programmes. This also means that the $\sim_C$ relation might change over time. For instance, when a new research programme is born, theories hitherto equivalent may become inequivalent because certain differences that did not matter before suddenly do. Conversely, when a research programme is discontinued, theories hitherto inequivalent may become equivalent because certain differences that used to matter suddenly do not. This also

[33]Note that if the theories T_1 and T_2 are wholly irrelevant to a particular research programme, T_1 and T_2 wil ipso facto function equally well with respect to the three criteria.

strikes me as the right call: not only is science an inherently social endeavour, it is a human endeavour and therefore it is only right that differences in heuristic function are differences for human scientists. Last, as per the discussion in section 3.3, the differences discerned by $\sim_C$ are absolutely not epistemically idle and as such cannot be dismissed as merely "pragmatic concerns". Rather, differences in heuristic function are intimately linked to different judgements on what makes our currently best theories succesful. Ignore at your own peril.

I will now illustrate with an example.

4. A Case Study: Analytic Mechanics

It is well known that classical mechanics has three formulations typically taken to be equivalent by physicists. This is, however, contested in the philosophical literature on classical mechanics. North (2009) argues that Lagranian mechanics and Hamiltonian mechanics are different based in part on simplicity considerations and that Hamiltonian mechanics is the "natural" formulation of classical mechanics. Curiel (2014) argues that North is right to distinguish, but for the wrong reasons and argues that a careful examination of the mathematical structure of Langrage and Hamilton reveals that classical mechanics is really Lagrangian. Barrett (2019) employs the tools of category theory for the question of the equivalence of Lagrangian Mechanics and Hamiltonian Mechanics and shows that depending on how one chooses the morphisms, the category of models of Lagrangian Mechanics and the category of models of Hamiltonian Mechanics either are or are not categorically equivalent. I argue that there is yet another sense in which Lagranian mechanics and Hamiltonian mechanics are in-equivalent: they are constructively inequivalent.

4.1. Analytic Mechanics. First, in Newton's force formulation the dynamics are guided by his famous 2nd law stating that $\vec{F} = m\vec{a}$ where $\vec{F} = \nabla U$ is the resultant force on the object, where the potential U is a smooth scalar field, m is the mass of the object and $\vec{a}$ is the acceleration of the object.

Second, Lagrange showed that one can obtain precisely the same dynamics based on the so-called "Hamilton's principle of least action". The guiding idea here is that nature is lazy and prefers the easiest route. Technically, one defines a functional on the space of possible trajectories the body could take and solves an extremisation problem. This functional is what is referred to as the *action*. Specifically, the functional takes a curve γ to the integral $\int_\gamma \mathcal{L} d\mu$, where $\mathcal{L}$ is called the *Lagrangian* and is a function on the tangent bundle, the space of possible states the object could be in. Normally, $\mathcal{L} = K - U$, where K is kinetic energy and U is potential energy. Let x be the n-dimensional vector of position-coordinates, and let $\dot{x}$ be the n-dimensional vector of velocity-coordinates. Then one can show that the curve γ solves the minimization problem precisely if the so-called "Euler-Lagrange" equations hold[34]:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}}\right) - \frac{\partial \mathcal{L}}{\partial x} = 0 \tag{1}$$

Third, Hamilton noticed that it was sometimes difficult to actually solve the Euler-Lagrange equations and suggested applying the Legendre transformation to transform the n 2nd order Euler-Lagrange equations into the $2n$ 1st order so-called Hamilton's equations. In Hamiltonian mechanics, it is customary to use the variable q instead of x, so the Hamilton equations are:

$$\dot{p} = -\frac{\partial H}{\partial q} \tag{2}$$

$$\dot{q} = \frac{\partial H}{\partial p} \tag{3}$$

where $H(p, q, t) = p\dot{q} - \mathcal{L}(q, \dot{q}, t)$ (Arnold 1974). In the special case where $p = mv$, with $q = x$ and $\mathcal{L} = K - U$, we get $H(p, q, t) = mv \cdot v - K + U = mv^2 - \frac{1}{2}mv^2 + U = K + U$. But this means that "the Hamiltonian" H equals the total energy of the system. In order to discuss the role of energy conservation in Hamiltonian mechanics, it is necessary to introduce the formalism in more mathematical detail.

[34]We can think of γ as a curve on n-dimensional space parametrized by time. Both x and $\dot{x}$ can then be thought of as real-valued functions of time and thus equation 1 can be evaluated along $\gamma(t) = x(t)$.

In a typical modern formulation, Hamiltonian mechanics takes place on a *symplectic manifold*, i.e. a pair (M, ω) where M is an even-dimensional C^∞-manifold, ω is a non-degenerate and closed two-form on M called a *symplectic form.* The fundamental dynamical principle is then that the dynamics is governed by a vector field X, which preserves the symplectic structure in the sense that $\mathcal{L}_X w = 0$ uniformly. Thinking of the integral curves of X as parametrised by time, this is naturally interpreted as time-translation invariance of the dynamics[35]. Importantly, this is enough to derive Hamilton's equations in local form and it is a theorem of the system that energy is conserved[36]. Since Hamilton's principle of least action can be satisfied for non-conservative systems, no equivalent theorem holds in Lagrangian mechanics[37].

I will argue that this formal difference is reflected on the level of heuristics. Concretely, by deriving the dynamics from a principle of time-translation invariance of the symplectic structure, the Hamiltonian approach is local in nature and the dynamic is along indifference-curves for the Hamiltonian. This places the conservation of energy in a conceptually central place that makes this approach quite ill-suited for the theorist who suspect this conservation might be broken in a future theory. On the other hand, we will see soon how the Hamiltonian formalism lends itself well to a (mathematical) generalisation that in a sense points towards Quantum Mechanics. This means that Lagrangian Mechanics and Hamiltonian Mechanics are inequivalent viz. construction, and as discussed above, this makes them constructively inequivalent as theories.

One can of course work with a time-parametrised set of Hamiltonians but this gives up what is thought of by many as the central dynamical principle of Hamiltonian mechanics, namely local time-translation invariance and energy conservation[38]. On the other hand, it is perfectly possible to satisfy Hamilton's action-extremisation

[35]See Roberts (2021, p.66-67) for an in-depth discussion of time-translation invariance in Hamiltonian mechanics.
[36]See e.g. Arnold (1974, p. 207) or Abraham and Marsden (1978, p. 188) proposition 3.3.3.
[37]See Curiel (2014) for a discussion of the conservation of energy in Lagrangian and Hamiltonian mechanics.
[38]See Roberts (2021) for a philosophical analysis. See Arnold (1974) and Abraham and Marsden (1978) for technical discussion.

principle with a time-dependent Lagrangian and thus with a time-dependent description of energy. In sum, the conservation of energy follows directly from the governing dynamical principle in Hamiltonian Mechanics but not from the ditto in Lagrangian Mechanics.

4.2. Rule-Following in Science. The question of how two or more formulations of a model can differ has been discussed by Vorms (2011, 2012). What Vorms (2011) dubs the "format" of a representation is supposed to capture the difference in the way different representations convey information to their users (as in the case of the three formulations of classical mechanics). Specifically, she defines the *format* of a device R representing a scene S as a triple (I, CC, CK) where:

- I is the kind and quantity of information about S a particular agent A in a particular context C can draw from R;
- CC is the relative length of the inferential process P—or the number of inferential steps, if they can be counted—necessary for A in C to draw I from R (the cognitive cost);
- CK is the kind of cognitive operations involved in P (Vorms 2011, p. 289).

It is clear from the definition of "I" that this definition makes the format of a representation relative to the agent as well as the context. She writes:

> I will now show that assuming such a standard user is misleading, and quite problematic for a study of the use of models in theorizing, in particular if one wants to analyze their role in theory development as well as in scientific learning." (Vorms 2012, p. 265)

Vorms intends the format of a representation to be relative to the particular scientist rather than to some "standard user". Further, it is clear that Vorms intends the format of a representation to play a role in the developement of new theories:

> Moreover, Hamiltonian equations reveal the deep relations between Classical Mechanics and other fields of physics — such as statistical mechanics, quantum mechanics and relativistic quantum mechanics. (Vorms 2011, p. 292)

Consider the role of rule-following in scientific innovation. Vorms (2012) points out that though we need to assume that users of models follow certain rules, this should not make us overlook the dynamical nature of rule-following in science. Vorms returns to the example of analytical mechanics: if a scientist does not know how to solve differential equations, obviously none of the approaches are going to lead anywhere. But, on the other hand, we should not think that the process of reasoning with models for the purpose of scientific discovery is strictly algorithmic (Vorms 2012, p. 268). This is at the heart of innovation: striking a balance between the adherence to and the bending of rules. Adhere to the rules too much and you will be stuck in the status quo, bend too many rules and the lack of constraints will render theorising impossible. Vorms (2012) stresses that the format of a representation is dynamical and changes with time. She then points to the example of Hamilton's insight that using the Legendre transformation could make it easier to solve Lagrange's extremisation problem as an example of the dynamical nature of formats.

The case of Hamiltonian Mechanics illustrates the dynamic interplay between formalism and innovation: on one level, Hamilton simply provided a way to solve the Euler-Lagrange equations, thus staying within the confines of the Lagragian conceptualisation that as we have seen was global in nature. But, Hamilton's approach engendered an entirely different approach to classical mechanics that was local in nature and based on the conservation of energy. While putting the focus on the pragmatics of solving the Euler-Lagrange equations with the aid of the Legendre transformation does emphasise how formats can be useful, it fails to elucidate how the developement of new formats influences the developement of new theories. Only when Hamilton's approach is conceived of less as a trick to help solve the Euler-Lagrange equations and more as an independent framework based on the conservation of energy is the format rich enough to drive innovation.

4.3. From Classical to Quantum. Once we are faced with two formats, the point is that, while they might be equivalent within a current theory, they might suggest different ways forward. This point is recognised by Ben-menahem (2006), who suggests that this difference in innovative functionality makes conventionality at most temporary:

> empirically equivalent interpretations of a physical theory may well evolve into nonequivalent theories. The freedom to make a conventional choice may thus be a transitional phase. (Ben-menahem 2006, p. 36)

Feynman (1965) too supports the idea that different formats of classical mechanics function differently for the purposes of innovation. In a discussion of the three formulations, he concludes:

> Second, psychologically they [the three formulations of classical mechanics] are different because they are completely unequivalent when you are trying to guess new laws. As long as physics is incomplete, and we are trying to understand the other laws, then the different possible formulations may give clues about what might happen in other circumstances. (Feynman 1965, p. 53)

While the difference between Lagrange and Hamilton is typically taken to be one of convenience within classical mechanics, there is a real way in which Hamilton's approach works differently from Langrange's for the purposes of innovation. Interestingly, even Rosenstock et al. (2015) entertain the idea that different, albeit mathematically equivalent, formalisms might be differentiated by heuristic function:

> For our part, we see no reason to choose between these approaches, at least in the absence of new physics that shows how one bears a closer relationship to future theories.

Last, Barrett (2019, p. 1187) chimes in:

> Having different viewpoints on a theory may help catalyse progress to newer and better theories. But in order to reap these benefits we first have to recognize when two viewpoints are providing us with views of the same theory, and when they are instead providing us with views of different theories.

Not only does Barrett believe having access to different formalisms might drive innovation but at the same time he argues that there is a certain order to things: first you figure out which equivalence relation is appropriate on the space of theories and only then do you attempt innovation. This view enables a neat categorisation of values into those epistemologically potent ones we use to sort our theories into equivalence classes and the epistemologically impotent ones we use to make a choice *within* each class for the purposes of innovation. As I have argued above, this view is untenable as it is based on the idea that we have complete access to our current theories when they are still current, while in reality, the project of investigating current theories is intrinsically linked to the project of constructing new ones. I will proceed to give a concrete example of how the formalisms of Hamilton and Lagrange function differently for the purposes of innovation that will simultaneously illustrate why Barrett's imagined order of operations fail to do justice to the dynamic of theory development.

At first, neither Lagrange's nor Hamilton's approaches seem to apply in the context of quantum mechanics, since they rely on a total specification of the system at all times — something for which Heisenberg's uncertainty principle does not allow. However, the algebra of the Hamiltonian approach lends itself to generalisation in a way that makes it suitable for the quantum context. To see this, define the *Poisson bracket* of two smooth functions f, g as $\{f, g\} = \frac{\partial f}{\partial q}\frac{\partial g}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial g}{\partial q}$. It turns out that equations 2, 3 can be re-written in terms of the Poisson bracket as:

$$\frac{dq}{dt} = \{q, H\} \tag{4}$$

$$\frac{dp}{dt} = \{p, H\} \tag{5}$$

While the Poisson bracket is a helpful anti-symmetric operator for smooth functions, the Moyal, or "commutator", bracket is a helpful anti-symmetric operator for quantum mechanics. This results in a strong similarity between equations 4, 5 and the

Heisenberg equations governing the dynamics in Quantum Mechanics:

$$\frac{d\hat{q}}{dt} = i\hbar[\hat{q}, \hat{H}] \tag{6}$$

$$\frac{d\hat{p}}{dt} = i\hbar[\hat{p}, \hat{H}] \tag{7}$$

Historically, Dirac went from equations 4, 5 to equations 6, 7. Further, there are a variety of mathematically exact senses in which the Poisson bracket reappears as the limit of the Moyal bracket when $\hbar$ tends to zero[39].

Firstly, this illustrates the interdependence of standard realism and construction: having the old theory appear in the limit of the successor theory teaches us something about why the older theory was successful despite being incomplete. However, this also illustrates how different choices of formalism will function differently for the purposes of construction via the strategy that we have called generalisation: the structure of Heisenberg's equations is clearly suggested by Hamilton's equations in a way that it simply is not by either Newton's second law or the Euler-Lagrange equations[40].

4.4. A Couple of Other Examples. Another example is Einstein's famous resistance to adopting Minkowski spacetime as a model of special relativity. There is little doubt that his reversal on this point was instrumental in the development of the general theory: the moment your model is a Lorentzian Manifold $(\mathbb{R}^4, \eta)$, it is implicitly suggested that you choose a different manifold and a different metric. In contrast, the four-vector formulation of the special theory as an affine space is closed around itself and does little to point towards the general theory[41]. Here we again see generalisation in action.

[39]See Dirac (1947) for a historic reference on "quantisation"—the operation of turning a classical description of a system into a quantum mechanical one. See Landsman (2005) for a more modern discussion of different approaches to quantisation and see Roberts (2021, p. 201) for references on various impossibility results.

[40]The so-called "path-integral" approach to Quantum Mechanics uses the Lagrangian formalism but is fraught with problems. In any case, the argument here is not that only Hamilton's formalism can lead to Quantum Mechanics but rather that Hamiltonian and Lagrangian mechanics have different heuristic functions. In particular, the path-integral appeals to a small community of physicists with Feynman most notable among them.

[41]See Appendix B for a discussion of affine spaces.

A third example is the formulation of GR. While GR is typically formulated using Lorentzian Manifolds, Geroch (1972) showed that one can equally well express the theory in the language of so-called Einstein Algebras. This algebraic re-formulation illustrates all three of the strategies described above: for the scientist believing that perfectly localised spacetime points will have to be sacrificed in a successor theory of quantum gravity, it will arguably be easier to theorise using a formalism that does not explicitly quantify over spacetime points (clear expression). Further, since quantum mechanics is typically formulated using Hilbert spaces, it might be easier to unify GR and QM if GR too is formulated algebraically (uniformity). Last, the Einstein Algebras are commutative algebras, which means that the generalisation to non-commutative algebras is straightforward (generalisation). These non-commutative algebras can then be interpreted in geometric terms in line with their commutative brethren yielding what is called "non-commutative geometry". There are at least some indications that non-commutative geometries are useful in the development of theories of quantum gravity[42].

These are examples of how different formalisms work differently for the purpose of construction through the different strategies discussed in the previous section. In these examples, we saw differences in heuristic function driven both by the formalisms and by the beliefs held by individual scientists. The lesson we should draw from the formalism-centric strategies generalisation and uniformity is one of tolerance: we are invited to tolerate and even encourage a wide diversity of models for the sake of enabling innovation[43]. On the other hand, since scientists working on the development of new theories do not enjoy the benefit of hindsight, it will often be prudent for the scientist to look for the model that best allow them to develop their thinking relative to their theoretical commitments and expectations. This motivates a closer look at the strategy we have called "clear expression".

[42]See Connes (1994) for a reference on non-commutative geometry. See Parfionov and Zapatrin (1995) and Butterfield and Isham (1999) for discussions of the use of non-commutative geometry in the development of theories of quantum gravity.

[43]See Lal and Teh (2017) for a discussion of using category theory to generate new models.

5. A Closer Look at Clear Expression

Let us start with Newton's second law of motion: $\vec{F} = m\vec{a}$. Since $\vec{p} = m\vec{v}$ and $\vec{a} = \frac{d\vec{v}}{dt}$, we have $\vec{F} = m\vec{a} = m\frac{d\vec{v}}{dt}$. We have by linearity of the differential operator that $m\frac{d\vec{v}}{dt} = \frac{d(m\vec{v})}{dt} = \frac{d\vec{p}}{dt}$. So, we can equally well formulate Newton's second law as $\vec{F} = \frac{d\vec{p}}{dt}$ — the two are equivalent in Newtonian Mechanics. In relativity theory, however, the situation is different. Interpreting $\vec{p}$ as relativistic three-momentum, the equation $\vec{F} = \frac{d\vec{p}}{dt}$ still holds, while $\vec{F} = m\vec{a}$ is false[44] (Pletyukhov 2018). Though the two formulations of Newton's second law are equivalent within Newtonian Mechanics, they are not equivalent in relativity theory. For a physicist in search of a successor for Newtonian theory, it could very conceivably make a difference which one they think of as primary. In particular, the relationship between force and momentum expressed by the second version of Newton's second law is retained over theory change whereas the relationship between force and acceleration expressed by the first version is not.

The realist can look at Newtonian Mechanics and ask themself what in this demonstrably incomplete theory nevertheless accounts for its remarkable empirical success. Depending on the realist's ground beliefs and commitments, they can decide that the relationship between force and acceleration is central and thus should be retained over a coming theory change. This might lead them to speculate that the law $\vec{F} = m\vec{a}$ should remain true. On the other hand, they might decide instead that the relationship between force and momentum is the central one and that hence it should live through theory change. This might then lead them to the hypothesis that the law $\vec{F} = \frac{d\vec{p}}{dt}$ should remain true. From our vantage point, it seems that taking the latter route would be advantageous if the purpose is to come up with the theory of relativity, but from the point of view of clear expression, which of the two one should opt for depends wholly on one's beliefs and guiding heuristics. Let us next consider the conservation of energy.

Imagine a physicist who had among their background beliefs that conservation of energy should probably not be a part of the ultimate theory of everything[45]. For

[44] Though Malament (2012, p. 143) points out a sense in which the second law reappears in GR.
[45] Niels Bohr famously entertained this possibility concerning β-decay (Guerra et al. 2012).

this physicist, there will likely be a bad fit between their beliefs and the Hamiltonian approach where conservation of energy is an extremely salient part of the guiding heuristic. For this physicist, it would likely be better to adopt Lagrange's framework. It is important to note here that within classical mechanics, the conservation of energy is a theorem in both the Lagrangian formulation and in Newton's force formulation: the difference is how centrally the conservation of energy features in the guiding heuristics. Feynman (1965) frames this difference in terms of the axiom-theorem distinction and observes how one's choice of axioms can make a difference for the developement of scientific theories:

> If you have a structure that is only partly accurate, and something is going to fail, then if you write it with just the right axioms maybe only one axiom fails and the rest remain, you need only change one little thing. But if you write it with another set of axioms they may all collapse, because they all lean on that one thing that fails. We cannot tell ahead of time, without some intuition, which is the best way to write it so that we can find out the new situation. We must always keep all the alternative ways of looking at a thing in our heads; so physicists do Babylonian mathematics, and pay but little attention to the precise reasoning from fixed axioms. (Feynman 1965, p. 54)

Let us go through a toy example to unpack the idea that different axiomatisations of the same theory can fare differently for the purpose of developing new theories.

Consider two propositional theories, $I_1 = \overline{\{r, p \wedge q \leftrightarrow r\}}$ and $I_2 = \overline{\{p, q, p \wedge q \leftrightarrow r\}}$. One can check that the two theories are equivalent in the sense of having identical deductive closures. Now imagine a "theory change" over which we learn that $p \wedge q \leftrightarrow r$ is false. Updating in the obvious way yields versions I_1^* and I_2^* as $I_1^* = \overline{\{r\}}$ and $I_2^* = \overline{\{p, q\}}$ so I_1^* and I_2^* are not necessarily equivalent. One could imagine one theoretician preferring I_1 because they think that r is fundamental in underwriting the success of the theory while the connecting axiom $p \wedge q \leftrightarrow r$ merely adds a convenient equivalence. Another could prefer I_2 because they think that p and q are what account for the success and r is the convenient equivalent. Further, before the theory change this difference in attitude makes no difference for what propositions are believed to be true since the two theories are equivalent. However, once a new piece of evidence is introduced, the two scientists are led in different directions.

Connecting back to the running example of analytical mechanics, r could be the conservation of energy and I_1 a theory in which this conservation is an axiom, whereas the equivalent I_2 merely has the conservation of energy as a theorem.

Though it is somewhat artificial to think of scientific theories in terms of axioms and theorems, there is a clear analogy between the axiom-theorem distinction and the more salient-less salient distinction. Psychologically, it is immediately clear that the axioms will be true of the system whereas the truth of any given theorem can be more or less surprising. Similarly, the most salient features of a theory are obvious for the user whereas the less salient features can take a background role.

As mentioned above, the physicists in question will find themselves in a situation quite different than ours: whereas we look back in time with full knowledge of the developments of relativity theory and quantum mechanics, they look forwards through a future covered in opaque clouds. But their inability to foresee what will come in the future does not make the choice of model arbitrary: rather, they will have good (albeit subjective) reasons for thinking e.g. that the central relation expressed by Newton's second law is between force and momentum and not force and acceleration, or that the conservation of energy is central and should be retained and thus be searching for future theories using Hamilton's framework where the conservation of energy is the starting point. Their background beliefs will influence not only their choice of model but also the way they use the model to develop new theories.

The sciences typically do not progress by wholesale rejection of old concepts but rather by slight modifications of existing definitions and relations of and between concepts. As these definitions and relations are typically codified in equations, we should expect to see some equations surviving theory change with little or no structural alterations. As the example with Newton's second law showed, it can also be that while one formulation of an equation is retained over theory change another formulation is not. Further, correctly "spotting" the version that will remain true arguably puts one at an advantage for the purpose of theory development. This is just a version of the insight from selective realism discussed in section 3.3: guessing what accounts for the success of current theories should tell one what is true about them and what must therefore be retained over theory change.

There is no doubt that actual scientists display these sort of preferences all the time. Further, though such preferences are to some extent idiosyncratic, they are not arbitrary as they are grounded in the background beliefs of the individual scientist. These background beliefs in turn are the result of experience and academic training. Further, they play the very important role of narrowing down the space of possible successor theories. Feynman, displaying the scientists' usual dismay over "metaphysics", writes:

> People may come along and argue philosophically that they like one better than another; but we have learned from much experience that all philosophical intuitions about what nature is going to do fail. One just has to work out all the possibilities, and try all the alternatives. (Feynman 1965, p. 53)

But this is, of course, a non-starter. No one can manage "all the possibilities" so there is an absolute need to narrow down the operating space. Not having guiding heuristics is not an option so insisting on wanting "to try all the possibilities" simply amounts to not being aware of one's guiding heuristics. To portray scientists as blank slates is highly misleading. Rather, we should acknowledge the role played by the background beliefs of scientists and acknowledge that science is an inherently human activity instead of trying to explain it away. The philosophy of science should pay heed to the practice of doing science.

What we have seen is that different formalisms fit quite differently with different sets of background beliefs and as such can function differently for the purposes of theory construction. Such differences are real differences and should be taken seriously regardless of which mathematical relationships that might hold.

6. Realism and the French Objection

French (2011) considers a number of strategies for breaking the types of underdetermination that might threaten "the realist stance" and he dubs one of these strategies "appeal to heuristic fruitfulness" (Ibid.). By a "heuristically fruitful" formulation of a theory we are to understand a formulation:

> [...] leading to, or, weakly, as indicating (again in some sense) an empirically successful theory (French 2011, p. 5),

but French ultimately rejects the idea that appeal to heuristic fruitfulness can exonerate the realist in the face of putative underdetermination of theories by evidence[46]. What I have called "construction" uses the potential of a theory to lead to other theories in the future to break underdetermination and so I must either reconcile French's argument with my own claims or show where he errs. I will do the latter.

French (2011) starts his rebuke by describing what he takes the "appeal to heuristic fruitfulness" to be about:

> The idea is that we should prefer that formulation which is more heuristically fruitful, in some sense, where that sense can be broadly characterized, strongly, as leading to, or, weakly, as indicating (again in some sense) an empirically successful theory (see Pooley op. cit.) (French 2011, p. 5)

Note that he takes putative underdetermination to obtain between formulations rather than theories. This fundamental assumption is reflected in French's first challenge to the proponent of appeal to heuristic fruitfulness, that since the crux of the strategy is to choose the formulation leading to an empirically successful theory, it is necessary that a formulation and a theory can enter into:

[46]Which is to say, French rejects that heuristic fruitfulness counts as evidence in the relevant sense.

> the sorts of inter-relations that come to be established following certain heuristic moves; more particularly, there is the question whether the well-known kinds of moves that one can discern as leading from one theory to its successor, also hold between a formulation and a future theory (where it is not yet clear whether 'successor' is the appropriate term here). (Ibid.)

I am doubtful that the argument can work: if these formulations are merely notational variants of the same theory, we might reasonably fear that that the relevant relations cannot obtain between present formulations and future theories, but in this case it is doubtful that the realist stance was ever threatened in the first place. If, on the other hand, these formulations differ substantially in reference, the underdetermination might actually threaten the realist. But in this case the "formulations" look a lot like theories, and so it is non-obvious that there should be problems with them entering into relations with future theories in the way French imagines.

However, French's main argument does not hinge on the assumption that putative underdetermination obtains between mere formulations. He proceeds by way of an example using Hamiltonian and Lagrangian mechanics to show that appeal to heuristic fruitfulness must fail. Making an argument similar to mine from section 4.3, that there is a sense in which Hamiltonian mechanics led to quantum mechanics through the deformation of the Poisson bracket to the Moyal bracket, French argues as follows: either we attempt to evaluate the heuristic fruitfulness of the two formulations before the advent of quantum mechanics, in which case any reason we might proffer for preferring one over the other can only act as a "promissory note", or else we might try this strategy after the advent of quantum mechanics. But now French argues, "theoretical developments" have made the choice for us and no underdetermination remains to be broken by heuristic fruitfulness. In the former case, the strategy is ineffective, and in the latter, it is superfluous. There is not much hope for heuristics to play a meaningful role in the breaking of underdetermination then. Let us call this argument "the French fork".

Before responding to the fork, let us consider what French has in mind when he talks of "breaking underdetermination". He concludes:

> in the case of the Lagrangian and Hamiltonian formulations, one can justifiably claim that each demonstrated a degree of fruitfulness, and the relevant elements an associated degree of plasticity, so in this case one can't even make a retrospective determination. But the point is that even if one could, even if it were clear which formulation turned out to be more fruitful than the other, such considerations are really no help in breaking the underdetermination at all: either they are mere promissory notes, or there is no underdetermination to break! (Ibid.)

French argues that both Hamiltonian and Lagrangian mechanics demonstrate heuristic fruitfulness and plasticity. As a result, heuristic considerations cannot even retrospectively break the underdetermination.

While I agree that both frameworks have proved heuristically fruitful, the conclusion only follows if they were fruitful in the same way and for the same purposes. This is simply not the case. As I have argued in section 3.4, judgements of heuristic function are often relative to a research programme. This means that since each framework is preferable to some research community, the two should be counted as different theories. It is important to remember that we are not facing a choice between some incomplete options and one complete one but rather between flawed theories. The realist is convinced the success of each option stems from the fact that they each get something right, and the realist project therefore is to search for the kernel of truth in theories that are incomplete. Realism should not rest on the assumption that at each point there is one absolutely superior option.

More importantly, the reasons that lead each scientist to adopt a theory or theoretical framework are not idle speculation but will include the available experimental data. We have already seen an example of how experimental data influence the choice of formalism in section 4.4, where we saw that Geroch (1972) proposed an algebraic formalism for GR. For our purposes here, the interesting observation is that Geroch himself motivated the algebraic framework due to its alledged ability to better reflect the "smearing out of events" suggested by Quantum Mechanics[47] (Geroch 1972). The supposed smearing inherent in the particle-wave duality and the Heisenberg uncertainty relation ultimately stems from experimental data obtained by Planck, Einstein and others. As a matter of fact, any particular feature of Quantum Mechanics may not make it into a functioning theory of Quantum Gravity but the experimental record will have to be made sense of. Geroch is very much taking the experimental record into account when he suggests an algebraic framework for GR. Neither should one think that experiments "talk for themselves" so it is no wonder that different scientists might take different cues from the same body of experimental data. French imagines "breaking underdetermination" involving picking out one theory whose superiority everyone can agree on. I think this picture is flawed.

I will now respond to the two prongs of the French fork one at a time, starting with the idea that heuristic fruitfulness merely amounts to a promissory note before the relevant theoretical developments have been made. I am sympathetic to this view and in section 3.4 I explicitly emphasised how two theories may be constructively equivalent in the absence of a research programme for which some difference between the two matters. In other words, I accept completely that there may be times at which differences in heuristic function will do little to facilitate a choice between theories. It may even be that analytic mechanics around 1890 is such a case. Whether or not this is the case would depend on what research programmes were active at the time.

That said, I think French gives promissory notes an unduly bad reputation. What kind of scientific evidence is so conclusive as to not contain an element of future promise? Rather than simply shrugging away heuristic fruitfulness as irrelevant or weak, we should evaluate why a particular scientific community opts for a particular theoretical framework. If such choices are based on sound interpretations of the

[47]One might fear that this argument is ineffective due to the second prong of the French Fork: after all, Geroch writing in 1972 is long after the advent of Quantum Mechanics. However, Geroch's ultimate purpose is the development of a theory of Quantum Gravity, which means that his article was written long before the relevant "theoretical developments".

available experimental data together with appraisal of how these data are in tension with existing theories, I think it is very inappropriate to dismiss them as idle promises.

This leads me to the second prong of the French fork: the idea that there simply is no underdetermination to break once the relevant theoretical advances have been made. Once the anomalies pile up enough to drive theoretical advances, the underdetermination is broken by "theoretical developments". The idea is that the advent of quantum mechanics will break the underdetermination, obviating the need for heuristic fruitfulness. As I just discussed, breaking underdetermination often means different scientists coming to different conclusions after having examined the available experimental evidence, the available theoretical frameworks and their own background beliefs. What leads to these theoretical developments are different research programmes attempting to innovate in different ways and the fact that these aims make certain theories more appropriate than others. In the language of section 3.4, theories that used to be constructively equivalent ceased to be so in the face of new research programmes. So the reason why there is no underdetermination to break later on, if that is indeed the case, is precisely that heuristic considerations have done their work. As soon as we realise that breaking underdetermination often amounts to different researchers making different choices motivated by their particular aims together with the experimental record, one thing becomes clear: that underdetermination can be broken long before such drastic theoretical innovations as quantum mechanics enter the fray.

7. Conclusion

Section 2 presented Putnam (1974)'s refutation and introduced the strategies Elimination (Elim) and Discrimination (Discr). Whereas the purpose of elimination was to avoid mistaking the presence of notational variants for serious underdetermination, the purpose of Discrimination was to avoid overlooking factors that could help one choose one theory over the alternatives. This led to the idea that conventionalism can be productively framed as a question of choosing an equivalence relation on the space of theories. Section 3 discussed what makes conventionality either trivial and serious and situated Elim in the modern literature on theory equivalence a la Halvorsson, Weatherall, Rosenstock and Barrett. I argued that Putnam (1974)'s refutation should be understood as an instance of Discr and showed how analysing the functions of theories could aid the project of formulating a "neo-Putnamian" refutation. I then introduced the aims of "standard realism" and "construction" and argued that the project of investigating known theories is epistemologically linked to the project of constructing new theories. This insight led to a discussion of "clear expression", "uniformity" and "generalisability" as three strategies for selecting a formalism expedient for the construction of new theories. Last, section 3 introduced a novel equivalence relation based on the ability of a formalism to drive innovation through clear expression, uniformity and generalisability. Section 4 showed examples of how different formulations of classical mechanics and relativity theory work differently with respect to uniformity and generalisability. Finally, section 5 showed examples of how different formalisms work differently with respect to clear expression.

CHAPTER 3

Einstein Algebras

1. Introduction

Einstein Algebras have been suggested as a way to make General Relativity (GR) more conducive to interpretations that seek to either eliminate or downgrade the ontological status of spacetime points[1]. Opponents object that since Einstein Algebras and Lorentzian Manifolds are equivalent in the sense of being categorically dual, nothing can be gained by changing from one to the other[2]. I argue that there might still be good reasons for prefering one framework over the other. In particular, I propose that mathematically equivalent models of spacetime physics can differ in an important way: different models might suggest different paths forwards towards the development of new theories and models. Although this idea seems to have been forgotten in the recent philosophical literature[3], it is not new in the context of Einstein Algebras, which were originally proposed as a tool for the development of a theory of quantum gravity (Geroch 1972). I will make this precise by pointing out three ways in which these formalisms generally can fare differently in the construction of new theories.

This analysis helps shed some light on a recent discussion between Bain (2013) and Lam and Wüthrich (2015) regarding Einstein Algebras. Despite them disagreeing over whether category theory can help underwrite an object-less ontology, they agree that the lack of explicit quantification over spacetime points in an Einstein Algebra does not imply that reference to spacetime points is truly eliminated, since the algebra maintains the resources to construct point-correlates. I argue that there

[1]Geroch (1972) introduces Einstein Algebras and shows how they allow one to construct the structures needed to do General Relativity. See Earman (1989) for an overview of classic interpretations of spacetime.
[2]See Rosenstock et al. (2015) for a proof and discussion of this fact. For an earlier argument that the two frameworks are equivalent see Rynasiewicz (1992).
[3]For an older reference to this idea see Laudan and Leplin (1991).

nonetheless can be a positive reason for choosing an Einstein Algebra over an equivalent Lorentzian Manifold if the aim is to eliminate reference to spacetime points. This reason has to do with the model's capacity to assist in the work of developing new theories. This conclusion has implications for the burgeoning literature on theory equivalence in the tradition of Glymour (1970), Rosenstock et al. (2015) and Barrett and Halvorson (2016b), since any formal criterion of theory equivalence will be insensitive to differences in a model's ability to drive innovation.

In section 2, I review GR in the standard formulation using tensor fields on a manifold, and then introduce the Einstein Algebra formulation due to Geroch (1972). I review the different motivations presented by Geroch (1972) and Earman (1989) for adopting Einstein Algebras, and discuss a sense in which the two framewords are equivalent, due to Rosenstock et al. (2015). Section 3 introduces the debate about the interpretive significance of these formulations between Bain (2013) and Lam and Wüthrich (2015). In particular, I discuss the interpretive significance of the fact that an Einstein Algebra has the resources to construct point-correlates and argue that Einstein Algebras might still offer a good tool for eliminating reference to spacetime points.

Section 4 contains my response to Bain (2013) and Lam and Wüthrich (2015) and my analysis of the significance of spacetime theory based on Einstein Algebras. One function served by theories is the construction of new theories, and I argue that different formalisms can perform differently relative to this aim even when the formalisms are in some sense mathematically equivalent. Then I list three strategies through which this heuristic function can manifest and situate comments made by Lal and Teh (2017) in this framework. Lastly, I draw a lesson for theory equivalence. Section 5 concludes.

2. Relativistic Spacetimes and Einstein Algebras

A common model of GR is an ordered pair[4] (M, g) where M is a four-dimensional C^∞-Manifold and g is a $(0,2)$-rank[5], symmetric, non-degenerate tensor field with Lorentzian signature, known as a *Lorentzian metric*[6]. I will refer to such a model as a *Lorentzian Manifold.* As Lorentzian manifolds are studied within the purview of differential geometry, this structure is commonly viewed as geometric in nature. However, it is also possible to build models of GR that are manifestly algebraic in nature. This latter structure is referred to as an *Einstein Algebra*, and in this section I would like to define and introduce some connections between these two structures. I will begin by identifying some of the natural algebraic structures appearing in Lorentzian Manifolds.

A Lorentzian Manifold contains a rich algebraic structure among its tensor fields. At each point $p \in M$ sits a $4d$ vector space T_pM referred to as the *tangent space.* These spaces are standardly constructed by first considering the ring $C_p^\infty(M)$ of smooth, locally defined, real-valued functions[7] at p. The spaces $C_p^\infty(M)$ automatically have the structure of an algebra under pointwise addition, pointwise multiplication and multiplication by real scalars. This means we can define a *point derivation* ξ_p on $C_p^\infty(M)$ as an $\mathbb{R}$-linear, real-valued map satisfying the (pointwise) Leibniz rule, $\xi_p(fg) = \xi_p(f)g(p) + \xi_p(g)f(p)$. The space of all derivations on $C_p^\infty(M)$ naturally forms a $4d$ vector space (Malament 2012, proposition 1.2.3), which is the tangent space T_pM. The vectors of T_pM are typically referred to as *contravariant* while the elements of the dual spaces $(T_pM)^*$ are referred to as *covariant.* By forming the disjoint union of all the tangent spaces we form the *tangent bundle* $TM := \cup_p T_pM$. A (contravariant) *vector field* ξ can now be defined as a smooth section of the tangent

[4]See Hawking, S.W. and Ellis (1973), Earman (1989). There are indeed many formulations of GR, but comparing them all goes beyond the scope of this article. See e.g. Landsman (2021) on the $3+1$ "PDE-formulation".
[5]I.e. for each $p \in M$, g takes two vectors and produces a real number.
[6]see e.g. Malament (2012, section 2.1)
[7]A function is locally defined around a point if it is only defined on an open subset containing the point. These functions are called *germs.* Technically, a germ is an equivalence class of C^∞-functions, but this detail won't play a role in what follows. See e.g. Tu (2011).

bundle, i.e. a smooth map $\xi : M \to TM$ such that for all $p \in M$ the value ξ_p falls in T_pM.

More complicated tensor fields are constructed by first considering, for each point $p \in M$, multilinear, real-valued maps on ordered tuples[8] of vectors and co-vectors. Such multi-linear maps are referred to as *tensors*. For example, a tensor at p taking one vector and two co-vectors as input would have rank $(2, 1)$ and the space of all rank-$(2, 1)$ tensors at p could be denoted $\mathcal{T}_1^2(T_p(M))$. Taking the disjoint union $\cup_p \mathcal{T}_1^2(T_p(M))$ then allows us to form the *tensor bundle* $\mathcal{T}_1^2TM$ and a *tensor field* Λ of rank $(2, 1)$ would be a smooth section of $\mathcal{T}_2^1TM$, i.e. a smooth map from M to $\mathcal{T}_1^2TM$ such that $\Lambda_p \in \mathcal{T}_1^2(T_p(M))$ for all p. Note that this means that, while tensors at a point take vectors and co-vectors at that point as input, tensor fields take other fields as input. In particular, the metric tensor g is a rank-$(0, 2)$ tensor field that is *symmetric* and *non-degenerate*, the former meaning that $g(\xi, \eta) = g(\eta, \xi)$ for all vector fields ξ and η and the latter that if for some vector field ξ we have $g(\xi, \eta) = 0$ for all vector fields η, then $\xi = 0$ (i.e. ξ_p is the zero vector for all p.). This means that the metric tensor equips each point in the manifold with a generalised inner product on the associated tangent spaces[9]. Last, the metric tensor is required to be of *Lorentz signature*, which means that for any point $p \in M$ there is a basis $\{\xi_1, ..., \xi_4\}$ for the tangent space T_pM such that:

$$g(\xi_i, \xi_j) = \begin{cases} 1, \; i = j = 1 \\ -1, \; i = j \text{ but } i \text{ and } j \text{ are different from } 1 \\ 0, \; i \neq j \end{cases} \tag{8}$$

One standardly[10] takes a given choice of metric and manifold (M, g) to represent spacetime, with the elements of the manifold representing individual events in

[8]For the expert: in the so-called “abstrat index notation” by Penrose, the arguments of tensors are not ordered because the abstract indices does the work of bookkeeping which arguments goes where.
[9]Ordinarily, an inner product $\langle \cdot, \cdot \rangle$ on a vector space H satisfies the conditions $\langle x, x \rangle \geq 0 \, \forall x \in H$ and $\langle x, x \rangle = 0 \, iff \, x = 0$. None of these conditions will be satisfied for “inner products” considered in GR due to the metric tensor having Lorentz signature. See e.g. Rudin (1987).
[10]See Malament (2012, pp. 119-120).

physical spacetime. Any further geometric or matter-energy structure is then represented by further tensor fields[11].

Notably, when defining the tensor fields in accordance with the above schema, we do so by specifying the value of the tensor fields at each point of the manifold. This makes the tensor fields depend on the points of the manifold for their very definition. Moreover, the points of the manifold are introduced without reference to the tensor fields. In this sense, the formalism suggests that the spacetime events represented by the points of the manifold are ontologically primary and the fields merely properties borne by these events.

In contrast, the Einstein Algebra approach to GR reverses the order of these definitions. This led relationists following Earman (1989) to suggest that Einstein Algebras might be pertinent for the formulation of spacetime theories that avoid taking spacetime points to be fundamental[12]. Here is how these structures are defined.

Geroch (1972) begins the presentation of Einstein Algebras with a construction of tangent spaces and tensor fields on a manifold that differs slightly from the standard approach presented above, where we cannot even define the various tensor fields without already having access to the points of the manifold. Recall that we defined the vectors of the tangent spaces T_pM as point derivations on the space of locally defined smooth functions and used these tangent spaces to form the tangent bundle of which the vector *fields* were smooth sections. There is an alternative construction that defines the vector fields directly: vector fields can be defined as derivations on an abstract algebra $\mathcal{R}$ isomorphic to $C^\infty(M)$, where a *derivation* is an $\mathbb{R}$-linear map $\xi : \mathcal{R} \to \mathcal{R}$ satisfying the Leibniz rule $\xi(fg) = \xi(f)g + \xi(g)f$ for all $f, g \in \mathcal{R}$. It is

[11]Other geometric components, such as the connection ∇_a, the Riemann curvature tensor R^a_{bcd}, the Ricci tensor R_{ab} and the Riemann curvature scalar field R, are indispensable for doing GR. The reason we can get away with only picking a manifold and a metric is that these other pieces of kit are determined by the metric: the metric determines a unique connection and the connection determines the Riemann tensor(Malament 2012).

[12]See Butterfield (1989) for a discussion of Einstein Algebras as a tool to underwrite a relationist interpretation of GR. See Rynasiewicz (1992) for an early argument that Einstein Algebras are equivalent to Lorentzian Manifolds. Note that Earman calls them “Leibniz Algebras” in homage to the historic debate between Leibniz and Newton.

a simple exercise in differential geometry to show that that the two definitions are equivalent[13].

The advantage of Geroch's approach is that it does not rely on the specification of the elements of $C^{\infty}(M)$ as smooth, real-valued functions on a manifold. Since the only place the manifold enters explicitly is in the construction of the algebra $C^{\infty}(M)$, we might as well begin with an abstract algebra without making reference to spacetime points. As we have just seen, the (vector-)space of all contravariant vector fields is simply the space of derivation on the algebra. We can denote this space $\mathcal{D}$. Then we get the space of co-vector fields as the dual space[14] $\mathcal{D}^*$. Since more complicated tensor fields are simply multilinear, real-valued maps taking vector fields and co-vector fields as input, we can piggy-back on our algebraic construction of vector fields and co-vector fields to obtain the other tensor fields we need[15] (curvature, torsion, the Einstein tensor, the stress-energy tensor etc.). For the metric tensor, non-degeneracy means that g induces an isomorphism between $\mathcal{D}(\mathcal{R})$ and $\mathcal{D}(\mathcal{R})^*$ sending a vector field ξ to the dual vector field $g(\xi, \cdot)$.

So far we have constructed only fields: if we want an algebraic analogue of vectors at a point, we need to first construct 'points'. The guiding observation here is that the algebra $C^{\infty}(M)$ in a sense is the dual of the manifold M. Hence, our construction can rely on the ubiquitous identification of a space with its double dual[16]. This approach is borne out in that the elements of the dual space $C^{\infty}(M)^*$ stand in one-to-one correspondence with the points of M. Thus, in the general case where we are given only some abstract algebra A over $\mathbb{R}$, we simply form the dual space $|A|$ of real-valued homomorphisms on A, called *the set of points*. We then interpret the

[13]If ξ is a derivation on $C^{\infty}(M)$, we can define a section s of the tangent bundle by letting $s(p)f := (\xi f)(p) \in \mathbb{R}$. On the other hand, if s is a section of the tangent bundle we can define a derivation ξ on $C^{\infty}(M)$ by letting $(\xi f)(p) := s(p)f \in \mathbb{R}$.

[14]If V is a vector space, we define the *dual space* V^* to be the vector space of real-valued, linear maps on $\mathbb{R}$.

[15]See Heller (1992).

[16]The relation between V and V^* can be realized via the "evaluation" map $\Lambda : V \to V^*$ defined by $(\Lambda x)f = f(x)$. This map is *canonical* in the sense that it can be represented without recourse to a basis for V.

elements of $|A|$ as points of some manifold and define a vector ξ_f at a point $f \in |A|$ as a pair (ξ, f) where ξ is a vector field[17].

Using similar thinking, Geroch (1972) shows how one can obtain the covariant derivative and the Lie-derivative from an appropriate algebra. The reader who is familiar with the theory of Lorentzian Manifolds might at this point wonder how it is possible to carry out a certain class of operations in purely algebraic terms—without reference to spacetime points. Namely, these are operations that are normally thought of as involving comparisons of different points of the manifold. To illustrate, I will offer the example of *parallel transport* of vectors.

In a Lorentzian Manifold, the metric tensor determines uniquely a so-called *covariant derivative*[18] ∇ said to be *compatible* with[19] g. One feature of the covariant derivative is that it allows a notion of constancy[20] of tensor fields along curves on the manifold: if γ is a curve[21] and ξ is a vector field defined on the image of γ[22], we say that the vector field ξ is *constant* along γ just in case $\gamma'(s)\nabla\xi = 0$ for all $s \in I$, where $\gamma'(s)$ is the derivative of γ with respect to s. If ξ_0 is a vector at $\gamma(s_0)$, one can prove that there exists a unique vector field (tensor field in general) ξ on the image on γ "extending" ξ_0, i.e. such that $\xi_{\gamma(s_0)} = \xi_0$. The value of this vector field at another point on the curve we say arises from *parallel transport* of ξ_0 along[23] γ. The upshot is that we have a standard of equivalence between tensors sitting at different points of the manifold (and hence covariant derivatives are sometimes referred to as *connections* because they "connect" the spaces of tensors at different points.)

[17]Note that we can think of ξf as a map taking an element x from the algebra $\mathcal{R}$ to the real number $f(\xi(x))$ analogous to how tangent vectors take smooth functions to reals on a Lorentzian Manifold.
[18]See e.g. Malament (2012) or Tu (2011) for a definition.
[19]This derivative is called the *Levi-Civita* connection and is the unique, torsion-free covariant derivative such that any vector field constant wrt. g is constant wrt. ∇. One can show that this is equivalent with the requirement that $\nabla g = 0$. See Malament (2012, p. 76-77).
[20]The metric tensor and the covariant derivative each come with a natural notion of constancy and as noted in the previous footnote, compatibility between the two amounts exactly to the requirement that these two notions be equivalent.
[21]I.e., $\gamma : I \to M$ is smooth and $I \subset \mathbb{R}$ is open.
[22]The definition extends analogously for more complicated tensor fields.
[23]See Malament (2012, p. 57).

In an Einstein Algebra, we can get the covariant derivative compatible with the metric by explicit definition[24] of its action on an arbitrary field. With ∇ in hand, one simply uses the definition of parallel transport from above, thus obtaining a standard of equivalence between fields sitting at different "points"—without ever quantifying over points.

Thus, we can do without the manifold at all and just start with an algebra isomorphic to $C^\infty(M)$. The above constructions show how we can retain all the structure we need to do GR in this setting. This leads Geroch to the definition of an *Einstein Algebra*.

Definition 3 (Einstein Algebra). An *Einstein Algebra* is a pair $(\mathcal{R}, g)$ where $\mathcal{R}$ is a commutative ring containing a subring $\mathcal{F} \subset \mathcal{R}$ isomorphic to the ring of real numbers, and g is a Lorentz signature metric, i.e. an isomorphism between the space of contravariant vector fields and its dual satisfying[25] Equation 8.

A couple of comments are in order: the point of the sub-ring $\mathcal{F}$ is to retain the structure of the constant functions in the set $C^\infty(M)$, and standardly, g is taken to represent the metric field, whereas the matter-energy and other tensor fields are represented by certain algebraic constructions on $\mathcal{R}$. Note also that if the aim is to provide an independent way of formulating GR, it would defeat the purpose to simply let $\mathcal{R} = C^\infty(M)$ for some manifold M.

What is the relation between these two formalisms? Clearly each spacetime (M, g) gives rise to an Einstein Algebra: simply form the ring of smooth functions $\mathcal{R} = C^\infty(M)$. So we might speculate that the two formulations are in some sense equivalent. Recently, the language of category theory has been employed to investigate such structural equivalences, and it will be helpful to do so here[26].

In our first way of building up GR, we viewed it as a theory of Lorentzian Manifolds. This perspective can be captured by the category *Lor*, in which the

[24] See e.g. Geroch (1972).

[25] Where $g(\xi, \eta) := g(\xi)\eta$. Since $g(\xi)$ is a co-vector field this expression evaluates to a scalar function as it should.

[26] I will not rehash the basic definitions of category theory here, but the reader can consult e.g. Awodey (2006).

objects are Lorentzian Manifolds and the morphisms isometries[27]. On the other hand, if one views GR as a theory of Einstein Algebras, the relevant category is EA in which the objects are Einstein Algebras and the morphisms are homomorphisms of Einstein Algebras[28]. The suspicion that the two formulations of GR are in some sense equivalent is borne out by Rosenstock et al. (2015), who show that the categories Lor and EA are dual[29]. This means that there are contravariant functors $F : Lor \to EA$ and $G : EA \to Lor$ such that the compositions $F \circ G$ and $G \circ F$ are *naturally isomorphic*[30] to the identity functors 1_{EA} and 1_{Lor}, respectively. The duality of the categories Lor and EA implies that we can translate back and forth between geometric models and Einstein Algebras without loss of information. In particular, we have already seen that, given an Einstein Algebra, it is possible to reconstruct the points of the associated Lorentzian Manifold. This will be relevant in section 3.

3. Spacetime Structuralism

At first it seems that there is a choice to be made by the physicist: either to go with the standard formulation using tensor fields on a manifold with the explicit reference to spacetime points, or to use the Einstein Algebra formulation where this reference is seemingly eliminated. On the other hand, a standard response is that insofar as the two frameworks are mathematically equivalent, the elimination of spacetime points inherent in the move to Einstein Algebras is an elimination in name only. This is precisely the critique levelled by Lam and Wüthrich (2015) in a response to Bain (2013). In this section, I would like to review this debate and then point out a sense in which Lam and Wüthrich's critique can be avoided. Further, I will identify

[27]An isometry ϕ between Lorentzian manifolds (M, g) and (M', g') is a diffeomorphism such that $\phi^* g' = g$. See Malament (2012, p. 85).

[28]An Einstein Algebra homomorphism ϕ between the Einstein Algebras $(\mathcal{R}, g)$ and $(\mathcal{R}', g')$ is an algebra-homomorphism such that $\phi_* g = g'$. For details, see Rosenstock et al. (2015), or for the full version, Rosenstock (2019).

[29]Note that when making a category out of the models of a theory one faces a choice of morphisms. As such, the categories Lor and EA are only one possibility for modelling GR in the geometric formulation and the algebraic formulation, respectively.

[30]Let C and D be two categories and $f : C \to D$ and $g : C \to D$ be two functors. Then f and g are *naturally isomorphic* if there exists a family of isomorphisms $\eta_c : FC \to GC$ in D, indexed by the objects $c \in C$, such that for all morphisms $f : c \to c'$ in C, $Gf \circ \eta_c = \eta_{c'} \circ Ff$ (Awodey 2006, pp. 134, 136).

a positive reason for adopting Einstein Algebras that they do not consider, regarding the role of Einstein Algebras in theory construction. This will be developed further in the remainder of this paper.

Bain (2013) sets out to defend the Radical Ontic Structural Realism (ROSR) of Ladyman (1998), French (2010) and others, against the well-known problem of how to interpret relations without the existenc of any objects that could serve the role as relata. Since relations are standardly defined extentionally in Set Theory[31], opponents claim that the main tenet of ROSR is simply incoherent. Call this objection the Incoherence Objection. Bain's strategy for defending ROSR against the incoherence objection is to look for frameworks other than Set Theory that might be more hospitable to the structuralist project and to argue that Category Theory could be one such framework.

To give an example of a situation where one formalism is more appropriate for the structuralist project than another, Bain offers up Einstein Algebras. As we have seen above, a straightforward interpretation of GR in the geometric formulation lets the manifold points represent individual events in physical spacetime. Since the structuralist would like to do away with individual events, the move to the Einstein Algebra formalism might look prima facie promising as the manifold points seem to be eliminated:

> in general relativity (GR, hereafter), I will now argue, moving from the tensor formalism to the Einstein Algebra formalism supports an ontology of spatiotemporal structure in which the articulating role that spacetime points play in the former is eliminated. (Bain 2013, p. 1625)

However, as we have just seen, there is a sense in which the Einstein Algebra formalism is mathematically equivalent to the geometric formalism and in particular that there is a way of retrieving the spacetime points from an Einstein Algebra. Bain (2013) already recognises this:

[31]This means i.e. that a binary relation R on a set A is a subset of the cartesian product $A \times A$.

> Insofar as the relata associated with tensor models are distinct from those associated with EA models, in adopting the EA formalism, we eliminate explicit reference to manifold points. On the other hand, one might question whether this is an elimination of manifold points in name only. Given the 1-1 correspondence between tensor models and EA models, to every manifold point in the former, there corresponds a maximal ideal in the latter (and vice-versa). Thus any reference to a manifold point in a tensor model of GR will be translatable in a 1-1 fashion into a reference to a maximal ideal in an EA model. (Bain 2013, p. 1625)

Bain (2013) argues that for solutions to the field equations with asymptotic boundary conditions, the move to Einstein Algebras amounts to a non-trivial elimination of reference to spacetime points. He then makes two points: first, that treating solutions with asymptotic boundary conditions in the Einstein Algebra formalism is more parsimonious because all solutions—including those with asymptotic boundary conditions—can be embedded into one category of so-called sheaves of Einstein Algebras. This, Bain (2013) argues, is in contrast to the geometric formalism where solutions with and without asymptotic boundary conditions belong to different categories. Second, that the relevant algebraic structure for modelling such solutions in general does not have the resources to construct correlates of spacetime points.

Lam and Wüthrich (2015) then argue that not even for solutions with asymptotic boundary conditions is implicit reference to spacetime points retained[32]. While the resolution of this disagreement is outside the scope of this article, I argue that both Bain (2013) and Lam and Wüthrich (2015) are too quick to be impressed by what we could call the Elimination Objection: the objection that adopting the Einstein Algebra approach fails because the spacetime points re-emerge as maximal ideals. Thus, the spacetime points are eliminated "in name only". But in order to evaluate

[32]They argue that the Einstein sheaves do have the resources for constructing "local points" (Lam and Wüthrich 2015, p. 628).

the elimination objection, we must first return to the maximal ideals from the Lam and Wüthrich (2015) quote above.

The construction both have in mind when they state the equivalence of the two frameworks is not the one we have already seen that was based on homomorphisms on the algebra. Rather, it starts with the following observation: if we construct the algebra $C^{\infty}(M)$ of smooth, real-vaued functions over a manifold M, we can recover the point $p \in M$ as the set $I_p = \{f \in C^{\infty}(M) \mid f(p) = 0\}$. If instead of starting with a manifold and creating the algebra of smooth real-valued functions, we are instead given an abstract Einstein Algebra A, we of course cannot make sense of the sets I_p because we have not got the set of points to which the "p" could refer. But it turns out there is another way of characterising the sets I_p that is not dependent on the points. The sets I_p are *maximal ideals* of the algebra $C^{\infty}(M)$. An ideal I of an algebra A is a sub-algebra that is closed under multiplication by an arbitrary element from A, and an ideal is *maximal* if it is not properly contained in any other ideal. Conversely, it turns out that all the maximal ideals of $C^{\infty}(M)$ are of the form $I_p = \{f \in C^{\infty}(M) \mid f(p) = 0\}$ for some $p \in M$ so the correspondence is one-to-one. In the general case where we are simply given an abstract Einstein Algebra A, it remains true that the maximal ideals are in one-to-one correspondence with the points of a particular manifold—namely the one that can be constructed using the fact that the categories *Lor* and *EA* are dual. So there is indeed a sense in which we can construct spacetime points from an Einstein Algebra using maximal ideals.

To evaluate the strength of the Elimination Objection, it is necessary to take a closer look at the two different ways of constructing manifold points from an Einstein Algebra we have seen so far. Consider first the construction based on maximal ideals. If we minimally assume that the proponent of an Einstein Algebra based approach to GR will want to commit ontologically to referents for the elements of the Einstein Algebra, it seems unavoidable that also sets of elements from the algebra will form part of the ontological commitments of the theory. If I believe in the pen, the cup and the piece of paper on my desk, there is a natural sense in which I also believe in, say, the union of the pen and the cup[33]. At least if the referents of the elements of the Einstein Algebra are localised in spacetime, it seems that the proponent of Einstein Algebras will be committed to referents for the maximal ideals that are also localised in spacetime.

However, as we saw in the previous section, there is another way of constructing manifold points from an Einstein Algebra relying on the homomorphisms of the algebra, and it is at least in line with physical practice to expect structure preserving maps on a mathematical space to represent properties of the elements of said space. Take as an example kinetic energy in the Hamiltonian approach to classical mechanics. Hamiltonian Mechanics takes place on the cotangent bundle T^*M over the phase space M equipped with generalised position coordinates q_i and generalised momentum coordinates p_i. The kinetic energy E_K is then a function taking (q_i, p_i) to the real number $\sum_i \dot{q}_i p_i$. Likewise, the elements of $|A|$ are real-valued maps on the Einstein Algebra. So, as the natural interpretation of kinetic energy is as a property of each point in the system under description, so one natural interpretation of the set of "points" in an Einstein Algebra is that a point refers to a property of the algebra representing the relational system, rather than to entities of their own. So, it looks like the Elimination Objection only really works if points *must* be reconstructed as maximal ideals. But, as we have just seen, there is another way and so the objection fails[34].

There is a further positive reason to adopt Einstein Algebras, which is the main reason for which I would like to argue. Right at their inception, Geroch (1972) saw Einstein Algebras as a potentially helpful tool in the quest to construct a theory of quantum gravity. Geroch suggested that:

[33]One could object to the implication from a commitment to the existence of individuals to a commitment to the existence of sets of individuals on metaphysical grounds. My main point do not hinge on this and I will not discuss the point further.

[34]One might worry that there is no real disagreement here: Lam and Wüthrich claim that Einstein Algebras maintain reference to points; I claim that it is possible to interpret them not do so. As the question is whether a move to the algebraic formalism can be used to escape reference to points, the possibility of a points-free interpretation is what separates our positions.

> it is perhaps reasonable to expect that, in a quantum theory of gravitation, the mathematical formalism will, at some point, suggest a "smearing out of events". (Geroch 1972, p. 271)

As the points on the manifold are typically taken to represent pointlike events in physical spacetime, one could speculate that eliminating the points would prove conducive to the project of quantising gravity. However, we have already seen how the Einstein Algebras have the resources to construct point correlates, so this raises the question: what could Geroch (1972) mean by the formalism "suggesting" a smearing out of events? We will return to this question in the next section. For now, we note that if the aim of Einstein Algebras is to facilitate the developement of a theory of quantum gravity by "suggesting a smearing out of events", the success of the project should be judged relative to this aim[35].

A brief note on terminology is needed first. Earman (1989) repurposed Einstein Algebras as a tool to formulate spacetime theories in a way intended to be hospitable to the relationist project in spacetime. In the context of spacetime physics, the term "relationism" carries historical baggage as the object of contention in the famous discussion between Newton and Leibniz. In modern times, the interest in relationism has been revitalised in the discussion of the so-called "hole argument"[36]. For our purposes, it suffices to note that relationists see relations on spacetime as fundamental and relegate localised points to a watered-down ontological status. On the other hand, "structuralism" is typically associated with the program of structural realism,[37] which is a brand of selective realism trying to avoid the threat from theory change while staying realist in spirit. In the context of the philosophy of physics, structural realism has been pushed heavily in Ladyman et al. (2007), and from the discussion between Bain (2013) and Lam and Wüthrich (2015), it is clear that the structural realist intends to do away with the spacetime points. Going forward, I will use the term

[35]Bain considers this aim as well (Bain 2013, p. 1631).

[36]For classical references see Earman (1989) and Butterfield (1989).

[37]See e.g. Worrall (1989) for a historical reference or Frigg and Votsis (2011) for an overview of the field up until 2011.

"structuralism" with respect to spacetime in this minimal sense: a view that intends to do away with individual, localised spacetime points.

4. Einstein Algebras Revisited

The realist might worry that no amount of evidence could help make the choice between a spacetime theory formulated using Lorentzian Manifolds and one using Einstein Algebras, and that therefore the choice would have to be fixed by convention. In chapter 2 of this thesis, I argue that conventionalism in the sense of underdetermination reduces to a question of which equivalence relation to employ on the space of scientific theories. Moreover, I identify two strategies for breaking underdetermination. The first, "Elimination" (Elim) involves adopting an equivalence relation so coarse that the prima facie underdetermined theories turn out to be merely notational variants of the same theory. We have already seen this strategy employed in the context of Lorentzian Manifolds and Einstein Algebras by Rosenstock et al. (2015), who show that the categories Lor and EA are categorically dual and argue that they are therefore theoretically equivalent. This, of course, absolves the realist, since the multitude of options reflects not a multitude of ways the world can be but rather a multitude of ways one can represent the same state of affairs. No one needs to blush just because Newtonian mechanics is sometimes taught in Mandarin.

The problem with Elim in the context of these two spacetime theories is that they are simply not equivalent as theories. This becomes clear once one attends to the aims of scientists motivating the formulation of theories in the first place. In particular, I will argue that the two spacetime theories discussed here fare differently viz. the aim of constructing new theories and that they therefore should not be counted as the same theory[38]. This necessitates the second strategy from chapter 2, which I call "Discrimination" (Disc). Whereas Elim involves the adoption of a coarse equivalence relation on the space of spacetime theories, Disc proceeds via the choice of a more finely grained equivalence relation together with an argument that T_1 and T_2 are distinguished in a way that is epistemically accessible to us. I will

[38]See chapter 2 for more on this.

argue that there are epistemically relevant reasons for preferring Einstein Algebras over Lorentzian Manifolds. To see this, we need to first turn an algebra into a theory of space and time by supplying it with an interpretation.

Recall from section 2 that an ordinary model of GR is an ordered pair (M, g) of a four-dimensional smooth manifold and a metric tensor. On a straightforward realist interpretation, the elements of M are interpreted to stand for physical points in spacetime, and the metric tensor to imbue these points with metric properties. At its most fundamental, this picture represents spacetime points as the bearers of properties supplied by the various tensor fields on the manifold. To give a physical interpretation to an Einstein Algebra, we turn this relation on its head and posit the field provided by the metric together with other fields that can be constructed by multilinear algebra on the algebra as the fundamental entities, and say that among the properties of these fields is a particular property we could call "localisation", in some sense mirroring the spacetime points in the standard formalism.

To see why this algebraic theory of spacetime is different in an epistemically relevant way, consider the consequent reduction in cognitive dissonance for the phycisist trying to come up with a theory of quantum gravity. It is easy to imagine our physicist believing that GR demands ontological commitment to completely localized spacetime points while at the same time believing that QM calls for the "smearing out"(recall Geroch's phrase) of the very same points. For this physicist, adopting Einstein Algebras in lieu of Lorentzian Manifolds would likely afford them an advantage: rather than trying to unify one theory (Quantum Mechanics), which seemingly demands the "smearing out" of spacetime points, with another theory (GR) in which spacetime points form the ontological bedrock, the conflict is reduced to one over a single property of the fields involved. The physicist will have no problem imagining that one property will change in the move from GR to a theory of Quantum Gravity,

In general, I have argued in chapter 2 of this thesis that the effects of formalism on innovation are mediated by (at least) three strategies:

- Clear expression

- Uniformity
- Generalisability,

with the spacetime theory based on Einstein Algebras displaying all three. "Clear Expression" refers to the strategy of choosing a formalism that complements one's beliefs and guiding heuristics facilitating the developement of new theories and models. In the case at hand, choosing Einstein Algebras over Lorentzian Manifolds could very well be expedient for a scientist guided by the idea that localised spacetime points should give way in the move to a theory of quantum gravity. For even though Einstein Algebras have the resurces to construct point-correlates, they will serve the heuristic function differently than Lorentzian Manifolds in which the quantification over points is explicit: by not explicitly quantifying over points, the scientist guided by the belief that spacetime points should be abandoned in the quest for a theory of quantum gravity is able to mould the formalism in their pursuits without having the points staring them in the face. The point is not that the algebra-based theory of space and time *is* already somehow a theory of quantum gravity but that the algebra-based theory is more amenable as a cognitive tool in the *pursuit* of a theory of quantum gravity for the scientist with structuralist intuitions. The search for new theories typically involves adjustments to old theories, and different formalisms will allow different adjustments.

What I call "Uniformity" refers to the selection of a formalism specifically to render two theories or theory fragments formally homogenous, in the hope of aiding unification. For example, GR is usually formulated in terms of differential geometry, while Quantum Theory is typically set in a Hilbert space setting. In this context, the strategy of uniformity would likely involve either the reformulation of Quantum Mechanics in geometric terms, or of General Relativity in terms of operator algebras. Here we are focusing on the latter approach. The hope is that having the formalisms look alike will facilitate theoretical unification.

Last, Generalisability is the strategy of formulating a theory in a framework that suggest a particular generalisation. The reformulation of Special Relativity from

Einstein's original formulation into the contemporary Minkowski formulation is one such example: taking a model of spacetime to be a manifold and a metric implicitly suggests that we might choose a different manifold or a different metric, paving the way towards GR. In the case of Einstein Algebras and Lorentzian Manifolds, the former implictly suggests the adoption of e.g. a non-commutative algebra, whereas the latter does not. And in the context of developing a theory of Quantum Gravity, the resulting so-called "non-commutative geometries" form the cornerstone of a prominent research programme[39] which is not formally equivalent to GR.

In the context of the discussion between Bain (2013) and Lam and Wüthrich (2015), the point that different models might suggest different generalisations despite being in some sense mathematically equivalent was picked up by Lal and Teh (2017). They argue that while Bain (2013) fails to show that category theory can help defend an object-less ontology, two strategies for using category theory to operationalize what I call Generalisability can be extracted from Bain's arguments. I will discuss the two in turn.

Generalisation by categorical duality (GenDual) is a strategy for generalising the models of a physical theory. First, one forms the category of models for theory T. Then, one looks around for another category T', dual to T, and uses the duality to establish equivalence between T and T'. Last, T' is embedded into a larger category $T*$ by dropping some requirement on T' and the larger category $T*$ is interpreted as a generalisation of T. A classical example is non-commutative geometry, where T is the category of locally compact Hausdorff spaces and continuous proper maps, T' is the category of commutative C^*-algebras and proper *-homomorphisms and $T*$ the category of all C^*-algebras. The complement $T^* - T'$ is then interpreted as consisting of "non-commutative" topological spaces. In this context, the commutative C^*-algebras and the locally compact Hausdorff spaces does in a strong sense contain the same information, but the commutative C^*-algebras have the virtue of "suggesting" that

[39]See Connes (1994) for a mathematical reference. See Parfionov and Zapatrin (1995) and Butterfield and Isham (1999) for discussions of non-commutative geometry in the development of theories of quantum gravity.

one drops the commutativity and thus one obtains something that can be viewed and investigated as an extension of the theory of locally compact Hausdorff spaces.

Since the category of locally compact Hausdorff spaces is dual to the category of commutative C^*-algebras, there must be some way of carrying out this generalisation directly on the topological category. But whereas one can easily imagine an algebra that is not commutative, it is not easy to imagine a topological space without points. This means that physicists are extremely unlikely to get the idea for this generalisation if they have access only to the geometric models.

Generalisation by Categorification (GenCat) is a strategy for generalising physical concepts. First, one picks a category that has relevance to physics and singles out a relevant property. If the category was *Hilb*, of finite dimensional Hilbert spaces and bounded linear operators, the property could be "entanglement". The next step is formulating the property in category-theoretic terms. Here, the relevant property of *Hilb* is that it is monoidal but non-cartesian, i.e. there is a notion of product between objects (namely, the tensor product $\otimes$), but not every element of the product of two objects can be realised as a pair of elements. Specifically, we can form the product $H_1 \otimes H_2$ in *Hilb* but if $\psi \in H_1 \otimes H_2$ there is no guarantee that we can find $\psi_1 \in H_1$ and $\psi_2 \in H_2$ such that $\psi = \psi_1 \otimes \psi_2$ (Lal and Teh 2017).

Having realised that "entangled" can be translated to "monoidal and non-cartesian", one can look around after other monoidal and non-cartesian categories. The hope is that possible other categories with the relevant property can be interpreted to have physical meaning. Indeed, Lal and Teh (2017) point to an example where the monoidal, non-cartesian category *Rel* has been used to describe non-deterministic classical processes.

I am sympathetic to the approach of Lal and Teh (2017), and our projects are really best thought of as complementary. Their focus is on how one can use GenCat and GenDual as mechanisms for what I have called "Generalisation"[40]. My main claim is that this is one of a class of practices that distinguishes physical theories even

[40]For an example of someone following "Unification", see Pitowsky (1984).

when they are formally equivalent. Before concluding, there is one more lesson we can draw regarding the recent literature on equivalence of theories.

We have seen two formulations of GR: one using tensor fields on a manifold, and one using Einstein Algebras. According to Rosenstock et al. (2015) the two are equivalent, but according to the present analysis, Einstein Algebras fare differently from Lorentzian Manifolds in at least three different ways viz. the development of a theory of quantum gravity. But this means that if we acknowledge that the development of theories is a central aim of Physics, then we are forced to say that there is a significant way models can differ, and to which categorical duality is completely insensitive.

Of course, the proponent of categorical duality as a criterion for equivalence of theories[41] can claim that the development of new theories is not a central aim for Physics and that this difference is therefore insignificant. However, this reply is implausible if one does not view theories as frozen objects. It is more reasonable to acknowledge that a formal criterion cannot cover all aspects of theory equivalence, and to take this as an invitation to investigate precisely what one learns from being told that two theories have categories of models that are dual.

As a first attempt, we can notice how having some kind of equivalence between two pieces of formalism can allow one to translate an interpretation of one piece into an interpretation of the other piece. If I have three pens on my desk, *yellow*, *blue* and *green*, I can represent them using the set $\{a, b, c\}$ as a representational vehicle. Concretely, I might let a represent *yellow*, b represent *blue* and c represent *green*. We can define an interpretation function I to bear this out so that $I(a) = \textit{yellow}$, $I(b) = \textit{blue}$ and $I(c) = \textit{green}$.

The sets $\{a, b, c\}$ and $\{1, 2, 3\}$ are obviously isomorphic, so we can choose an isomorphism of sets $\phi : \{1, 2, 3\} \to \{a, b, c\}$. This isomorphism can be used to "move" the interpretation since the composite $I \circ \phi$ is an interpretation function on the set $\{1, 2, 3\}$. The situation is showed in Figure 1. Consequently, the fact that the two sets are isomorphic means that if one of them stands in a representational relation to

[41]For example, Rosenstock, Barrett and Weatherall.

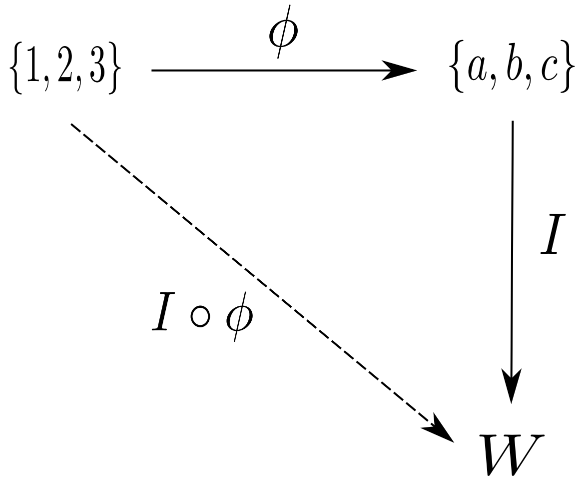

FIGURE 1. W stands for "world" as in external reality. If a map from a set to a piece of the world makes you queasy, you can think of W as a set of names naming objects in the real world. We do not need to solve the puzzle of reference presently.

some phenomenon, there is a way to make the other structure do precisely the same representational work by "lifting" the interpretation function via an isomorphism.

Two equivalent mathematical structures can do the same work. However, they need not. Let us imagine the process of going from the three colour theory to a theory of colours based on a continuum of different frequencies. It seems that with respect to Generalisation, the sets $\{a, b, c\}$ and $\{1, 2, 3\}$ are going to perform differently. The numbers $1, 2$ and 3 would likely be conceived of as natural numbers in the three colour theory, but the formalism suggests an obvious generalisation: embed $\mathbb{N}$ in $\mathbb{R}$ and fill in the continuum. On the other hand, while the set $\{a, b, c\}$ does suggest generalisation to the rest of the alphabet, we are only going to get to 25 (or 28 if we opt for one of the mighty Scandinavian alphabets). The idea of a continuum of colours is implicitly suggested by one formalism but not the other, despite the two being isomorphic as sets.

In the case of EA and GR, the fact that the two frameworks are in some sense equivalent means that we can transfer the standard interpretation of GR in the geometric formulation to a model in the algebraic formulation. Concretely, we can let the maximal ideals of the Einstein Algebra do the work of the points of the Lorentzian Manifold and represent localised events in physical spacetime. But that does not mean that is the only way of interpretating Einstein Algebras. I have suggested a different interpretation in which the fields constitute the ontological bedrock and the spacetime points become a property of these fields, and it is clear that this property-object reversal is something to which no functor is sensitive.

This shows that there can be a meaningful difference between the geometry-based spacetime theory and the algebra-based spacetime theory[42] despite them being categorically dual. In particular, work by Parfionov and Zapatrin (1995); Butterfield and Isham (1999) suggest that an algebra-based and non-commutative approach to quantum gravity holds promise. The work of any scientist is inevitably guided by background commitments working as heuristics, and this difference might very well make one framework preferable over the other.

5. Conclusion

I have argued that Einstein Algebras work differently to Lorentzian Manifolds for the purposes of constructing of new theories. Section 2 presented the theories of Lorentzian Manifolds and Einstein Algebras and their relationship, noting that the two are categorically dual. Section 3 discussed the interpretative significance of this duality; in particular, whether the move to Einstein Algebras constitutes an elimination of spacetime points. I discussed two different ways of producing "spacetime points" in the algebraic formalism and argued that in a plausible interpretation, the "points" in an Einstein Algebra can be viewed as referring to properties rather than to localised entities. Section 4 presented the main contribution of the chapter: three different ways in which Einstein Algebras function differently from Lorentzian Manifolds for the purposes of constructing new theories. This breaks the underdetermination by allowing the scientist to make the choice of formalism that best suits her aims: rather than convention, theory choice is fixed by epistemically potent reasons having to do with a formalism's capacity to aid in the process of scientific innovation. Another

[42]One might want to say "between the geometric and algebraic formulations of GR", but the point is precisely that these are two different theories rather than being two different formulations of the same theory.

result is that formal criteria of theory equivalence are incomplete: the differences discerned in the heuristic function are simply too subtle for functors and isomorphisms to detect. Last, I discussed a sense in which mathematically equivalent models *can* do the same interpretive work though they do not *need* to, as shown by the example of Einstein Algebras and Lorentzian Manifolds.

CHAPTER 4

Conventionality of Simultaneity and the Newtonian Limit

1. Introduction

Conventionalists regarding simultaneity claim that there is more than one candidate for what "simultaneity" means in the context of Minkowski spacetime, and that the choice between them is a matter of convention[1]. Opponents of the conventionality thesis claim there are good reasons to reject the non-standard candidates for relativistic simultaneity. How do we evaluate these competing interpretations of simultaneity?

In this paper, I propose that progress can be made through two theses. First, that some well-known responses to the conventionality of simultaneity can be helpfully interpreted as functionalist responses in a sense recently set out by Butterfield and Gomes (2023). Specifically, I will argue that viewing classic responses from Reichenbach (1958), Malament (1977) and Sarkar and Stachel (1999) through a functionalist lens makes explicit the sense in which each proposes to provide a *reduction* of the concept of simultaneity to some less problematic language. Second, I will argue that the only viable functionalist constraint on the concept of relativistic simultaneity is its behaviour in the non-relativistic limit. I go on to show that this constraint is extremely weak, allowing many different simultaneity relations, which I argue implies that either relativistic simultaneity is redundant on pain of being co-extensional with another well-known relation on Minkowski spacetime, or that the concept is simply devoid of meaning. In either case, my conclusion is that there is no place for simultaneity in special relativity.

In section 2, I review Einstein (1905)'s discussion of relativistic simultaneity and the approach to functionalist reduction of concepts from Lewis (1970) and Lewis

[1]See Jammer (2006) for a comprehensive overview of the literature as of 2006. For more recent contributions, see e.g. Valente (2016), Valente (2018), Bacelar Valente (2018), Hinchliff (2000), Ben-Yami (2019), Ben-Yami (2015), Thyssen (2019), Rovelli (2019), Weingard (1972), Savitttl (2000) and Balashov and Janssen (2003). For an overview of the literature post-2006, see Janis (2018).

(1972). I further discuss functionalism in a sense recently laid out by Butterfield and Gomes (2023), present a schema for functionalist reduction of concepts and discuss the relationship between functional reduction of concepts and conventionality. Section 3 presents the responses to the conventionality of simultaneity by Reichenbach (1958), Malament (1977) and Sarkar and Stachel (1999) and shows how to interpret these as functionalist responses. Section 4 presents my second thesis: that the only viable functionalist constraint on the meaning of relativistic simultaneity stems from the Newtonian context. Here I show—through pictures, as well as rigorous proof—that pretty much any relation one can write down on Minkowski spacetime converges on Newtonian simultaneity. Section 5 argues that the concept of simultaneity is either redundant or meaningless in the context of Minkowski spacetime. Section 6 concludes.

2. Simultaneity and Functionalism

2.1. Einstein's Definition of Simultaneity. The second paragraph of Einstein's landmark 1905 article set out a definition of what was widely assumed to be utterly unproblematic: the relation whereby two events in spacetime are counted as *simultaneous*. The definition comes in two stages: first, we say what it means for two spatially separated clocks to be *synchronised*, and then use this to define simultaneity. Einstein (1905) characterises the former as follows:

> Let a ray of light start at the "A time" t_A from A towards B, let it at the "B time" t_B be reflected at B in the direction of A, and arrive again at A at the "A time" t'_A. In accordance with definition the two clocks synchronize if $t_B - t_A = t'_A - t_B$. (Einstein 1905, p. 3)

Since this will be central in the discussion to come, I will go through the exercise of explicating Einstein's famous definition[2].

A light ray is sent at A-time t_A from A towards B (Figure 2a), and it is reflected back at B-time t_B towards A (Figure 2b), arriving back to A at time t'_A (Figure 2c). But if the A-clock and the B-clock are synchronous, Einstein reasons, we ought to

[2]For a discussion of Einstein's criterion of simultaneity and the literature surrounding it, see (Jammer 2006, chapter 7).

(A) At t_A the light beam is sent towards B.

(B) At t_B the light beam is reflected at B.

(C) At t'_A the light beam is received back at A.

FIGURE 2. A light beam travelling back and forth.

be able to calculate the time it takes for the ray to travel from A to B by taking the difference $t_B - t_A$, and in the same vein, the time it takes the ray to travel from B to A by calculating the difference $t'_A - t_B$. And *if* we further believe that the speed of light is the same in all directions, and thus that the light ray used the same amount of time to make the trip from A to B as it did to make the trip from B to A, we can say that the A-clock and the B-clock are synchronised if the local times bear out this fact, that is, if $t_B - t_A = t'_A - t_B$.

Given Einstein's definition of clock synchronicity, we now have a natural definition of simultaneity for spatially separated events. Imagine two observers, Annette and Barbara, located at A and B, respectively, both holding a coin. And assume the existence of local clocks, synchronised in accordance with Einstein's definition. If Annette releases her coin at A-time t_A, and Barbara releases her coin at B-time t_B, we will say that the two releases happened simultaneously precisely if $t_A = t_B$. In plain English: events associated with synchronised clocks are simultaneous if and only if the clocks display the same times.

There is another equivalent way of presenting Einstein's criterion of simultaneity that possibly is more intuitive. Let us place a third observer, Cecilie, at the spatial midpoint C between A and B, and let us say that Annette releases her coin simultaneously with Barbara so that $t_A = t_B$ (the A-clock and the B-clock are synchronised). When will the light rays from Annette and Barbara reach Cecilie's retina? A straightforward argument shows that they will reach her at the same time[3] $t_C = t_A + \frac{1}{2}(t'_A - t_A)$. Thus we have the following formulation of Einstein's criterion: spatially separated events A and B are simultaneous precisely if an observer situated at the spatial midpoint between A and B sees A and B as simultaneous, in that the light signals would arrive there at the same event.

Einstein had a weighty reason to be elaborate. The concept of simultaneity had been supposed to be absolute for centuries, and he was about to propose a different criterion whereby judgements of simultaneity would suddenly become relative to the observer. If such a radical proposal had not been prefaced with ample motivation, one could easily see how it would be viewed as controversial or even problematic. Understanding this, Einstein gave a reduction of the "problematic" concept of simultaneity in operational terms, which he expected the community of physicists to view as entirely unproblematic. Hence, he formulated his criterion in terms of familiar measurement devices such as clocks and rods, and the transmission of light rays as coordinated by these devices.

I would like to point out that Einstein's general philosophical strategy here, to reduce a problematic context to a description in terms of other well-known concepts, is precisely in the spirit of a well-known research programme in philosophy called *functionalism*, espoused most notably by **?**, Lewis (1970) and Putnam (1960). In their terms, I claim, Einstein is providing what is known as a *functionalist specification* of simultaneity. So, I will begin by briefly reviewing a classic presentation of that programme[4].

---

[3]Let us say that the light ray from A would have arrived at B at time t'_B and that the light ray from B would have hit A at time t'_A. By synchronicity we get that $t'_B - t_A = t'_A - t_B$, but we already know that that $t_A = t_B$ so $t'_A = t'_B$. But this means that on the A-clock, the light ray leaves A at time t_A and would have arrived at B at time $t'_B = t'_A$. But that means that the light ray going from A to B reaches Cecilie's retina at A-time $t_A + \frac{1}{2}(t'_A - t_A)$. Likewise, the light ray going from B to A leaves B at time $t_B = t_A$ and would have arrived at A at time t'_A. But that means that the light ray going from B to A also reaches Cecilie's retina at A-time $t_A + \frac{1}{2}(t'_A - t_A)$. So Cecilie is hit by the two light rays at the same time and, equivalently, she sees the two coins starting to fall simultaneously.

[4]This is inspired by the analysis in **?**.

2.2. Introducing functionalism. Lewis (1972) illustrates the basic idea of a functionalist specification:

> X, Y and Z conspired to murder Mr. Body. Seventeen years ago, in the gold fields of Uganda, X was Body's partner... Last week, Y and Z conferred in a bar in Reading... Tuesday night at 11:17, Y went to the attic and set a time bomb... Seventeen minutes later, X met Z in the billiard room and gave him the lead pipe... Just when the bomb went off in the attic, X fired three shots into the study through the French windows. (Lewis 1972)

Lewis asks the reader to imagine that in our world, this specification is true of precisely one triplet of names (Plum, Peacock, Mustard) and concludes that the natural conclusion is that X, Y and Z refer to Plum, Peacock and Mustard, respectively. This is what it means to say that Lewis' story provides a functional specification of Plum, Peacock and Mustard in the actual world: by describing the functional role of the placeholder (X, Y, Z), a unique referent is determined.

So far so good. But now imagine another world w that is free of Plums, Peacocks and Mustards. If the terms used in Lewis' story are well-understood by all relevant parties, and we agree that the story is appropriate, then we can transport our concepts to w by simply interpreting our functional specification in w (provided the inhabitants of the world w understand the terms "Uganda", "Reading", "Tuesday" and so on). Of course it may be that houses in w look different than in the actual world, or that billiards is played without cues, but if a certain triplet of names, say (Anne, Bill and Frankie), makes the story true in the context of w, Lewis argues that one should infer that Anne realises X, that Bill realizes Y and that Frankie realises Z. That is, the two triplets (Plum, Peacock, Mustard) and (Anne, Bille, Frankie) are each *realisers* of the terms (X, Y, Z) in their respective contexts. If our functionalist detective is competent, the functionalist specification allows us to understand what was previously unfamiliar (X, Y, Z) in terms of the familiar (the terms employed in the detective story).

Butterfield and Gomes (2023) point out that the detective story gives an example of a functionalist specification that allows one to transport a concept from one context to another. I will now add three ways in which one might fail to have a functionalist specification. First, it might be that there were simply no elements satisfying the functional role in a particular world. For example, it could be that no triplet of people in w would play the role of (X, Y, Z). Let us call this problem, "realiser-non-existence".

Secondly, it could be that the story fails to pick out one unique triplet in a particular world. For example, the detective story might be true of two triplets of people in w. In this case, we would have a choice about which realiser to use, e.g. of whether to use: "Anne" or "Catherina" for X, and that choice would have to be settled by some other means. Call this problem "realiser-non-uniqueness".

Thirdly, it could be that we have more than one functional specification to choose from in the first place. For example, maybe we could produce another story also picking out the triplet (Plum, Peacock, Mustard) in the actual world, but which produces a different set of realisers than the original story once re-interpreted in w. Then the choice of story itself would be underdetermined, even though each story might have a unique realiser in w. Call this problem "specification-non-uniqueness".

We shall encounter both realiser-non-uniqueness and specification-non-uniqueness in our subsequent discussion of relativistic simultaneity: Reichenbach will exemplify the former, while Sarkar and Stachel will exemplify the latter. But first we need a more careful discussion of the languages in which the functional specifications are formulated.

2.3. Languages of Functional Specification. This section will make the functionalist ideas of the last section more precise and give a more general account of how to employ functional descriptions to transport concepts between theoretical contexts. In particular, the language of a detective story is too coarse for the purposes of applying functionalism to physics. For the latter, it will be helpful to review a classic example from the literature: the functional specification of "pain"[5]. I will follow an example given by Butterfield and Gomes (2023), who are themselves inspired by Lewis, since their emphasis on the use of functional specifications for the purposes of reduction will play an important role in what follows. According to Butterfield and Gomes (2023), the fundamental idea is to divide the concepts of a theory into those that are well-understood outside the theory—the "unproblematic" terms—and those introduced by the theory—the "problematic" terms[6]. This idea can be viewed as a kind of intertheoretic reduction, from the "problematic concepts" to the "unproblematic concepts", with the added benefit that bridge laws will be theorems of the reducing theory rather than extraneous empirical postulates.

Given the two theories *folk* = Folk Psychology and *neuro* = Neurophysiology we seek to reduce the folk-term "pain" to neuro. The folk-terms are characterised by being introduced by Folk Psychology and include names of mental states such as "pain", "desire" and "intention". Likewise, the neuro-terms are characterized by being introduced by Neurophysiology and include physiological concepts like "neuron" and "pre-frontal cortex". Both theoretical vocabularies also contain common language: terms that were understood prior to the introduction of either theory. For the sake of argument, we can say that the common language includes descriptions of behavioural dispositions, and that the folk-term "pain" can be specified in terms of displaying aversive behaviour, a desire to make the pain stop and an intention to avoid the cause of the pain.

What does it mean to say that pain has been specified by these descriptions? According to Lewis, it means that precisely one folk-term fits the description, in that it is a realiser of the specification. "Pain" is then argued to be the unique folk-term realising the specification, "An agent A is experiencing pain if and only if i) A is displaying aversive behaviour, ii) desires to make the pain stop, and iii) intends to avoid the cause of pain." Further, since descriptions such as "aversive behaviour" and

[5]In the Philosophy of Mind, the idea of identifying mental states with brain states via functional specifications originates independently in **?** and Lewis (1966).

[6]The designations "problematic" and "unproblematic" are meant mainly as aide-mémoires and should not be taken to mean that there is something suspect about the so-called "problematic" terms.

"desiring to make the pain stop" are equally available to the neurophysiologist, we can ask whether there are neuro-terms realising that same specification. If we are so lucky that a *unique* neuro-term realises the specification (typically, the philosopher's candidate of choice is "C-fiber firing"), then we can reason as follows:

- The folk-term "pain" is functionally specified by the three behavioural dispositions above;
- The neuro-term "C-fiber firing" is also functionally specified by the same three behavioural dispositions;
- Conclusion: "Pain" and "C-fiber firing" are realisers of the same functional role.

On the well-known "Nagelian" approach to reduction[7], we would have to rely on bridge laws to match a folk-concept to a neuro-concept since they are different languages. But, if we are lucky enough to be able to prove that a certain neuro-term occupies the same functional role as the folk-term "pain", then this theorem provides the bridge law without any further assumptions.

Following Lewis (1972), Butterfield and Gomes (2023) emphasise that a successful functionalist reduction must be unique. Therefore, they require not only proof that "C-fiber firing" realises the pain role in neurophysiology, but also that it is the unique neuro-term realising the specification. In the language of Section 2.2, the observation is that in cases of realiser-non-uniqueness, the functional specification simply fails to refer. We will return to the question of the significance of non-existence and non-uniqueness in the next section.

Traditionally, inter-theoretic reduction takes entire theories as the unit of analysis[8]: we might attempt to reduce Folk Psychology to Neurophysiology, or Thermodynamics to Statistical Mechanics. However, it is also possible to focus more narrowly on individual concepts and still reduce a problematic concept to a less problematic one. One might argue, there is more epistemic security in our folk psychological "pain"

[7]This view is developed in Nagel (1961). See Dizadji-Bahmani et al. (2010) for a recent developement and review of this approach.
[8]Cf. Nagel (1961).

concept once we learn that it is reducible to "C-fiber firing". Likewise, we probably gain greater confidence in the "temperature" of an ideal gas after learning that it is reducible to "mean kinetic energy". This suggests the following general scheme for the successful application of functionalism:

(1) Start with an unproblematic concept in a theoretical context T and specify this concept functionally. The aim is to export this concept to a different theoretical context T' in which the meaning of the concept is undefined or problematic.
(2) Export the functional specification from T to T', using only vocabulary shared between them. Determine the extension of the functional specification in T'. Compute the extension of each realiser of the specification in T'.
(3) Conclude that these extensions characterise the meaning of the concept in T'.

This scheme puts constraints on what a concept originating in one theoretical context might mean in a different theoretical context. "Temperature" originates historically in thermodynamics (TD) and so prima facie is meaningless in the context of statistical mechanics (SM). The scheme gives us a way of ascribing meaning to "temperature" in SM: specify "temperature" functionally in TD and re-interpret the specification in the context of SM. The well-known result is that, in some cases, "Temperature" is realised by "mean kinetic energy"—a concept that is perfectly meaningful in the context of SM.

In the next section, I will argue that Einstein's criterion of simultaneity can be viewed as an application of functionalism in this sense. But first, I will point out two senses in which functionalism can result in conventionality.

2.4. Functionalism and Conventionality. Consider again the schema for functional reduction of concepts from section 2.3: under what circumstances will it make sense to say that the transported concept is conventional? To say that a concept in a theory is conventional means that the meaning of the concept is not uniquely specified by the theory and thus is only fixed by convention. So conventionality occurs for the functionalist when the schema fails to produce a unique meaning for the concept in the particular theoretical context in question[9].

First, it could be that once the functional specification from T (the theoretical context in which the concept is unproblematic) is interpreted in T', more than one concept is picked out. In the example of pain, this would mean that more than one neuro-concept realised pain. In this case, the realiser is not unique and so the resultant concept is conventional. This is what we called realiser-non-uniqueness in Section 2.2, and so we might call it *realiser-conventionality*. Second, it could be that we could specify the concept in more than one way. In the example with pain, this would mean that one could find more than one appropriate specification for the folk-concept of pain. This in itself does not ruin uniqueness for it could be that each specification produced the same realiser in T'. But if there are at least two appropriate specifications that pick out different concepts in T', then the choice between them is conventional. This is what we called specification-non-uniqueness in Section 2.2, and so we might call it *specification-conventionality*.

In the next section, I will argue that these two notions of conventionality form the crux of the simultaneity debate. To see this, we need only take the novel step of viewing the literature through a functionalist lens.

3. Competing Functionalist Approaches to Simultaneity

In section 2.1 we saw how Einstein (1905) offers a specification of a problematic concept—"relativistic simultaneity"—in terms of a less problematic theoretical context T consisting of rulers, clocks, light rays and mirrors. Specifically, according to Einstein, two events e, e' are simultaneous exactly if an observer standing at the spatial midpoint between them receives light rays from e and e' at the same time.

[9]Here I am following the Putnam tradition in which the interesting sense of conventionality is viewed as arising out of underdetermination. This view is made explicit in Ben-menahem (2006) (see pp. 7-12). Note that the approach is fully compatible with the interpretation favoured by Poincaré, for whom the equivalence of different geometries is absolutely central to his argument for the conventionality of geometry. Of course, one is not forced to draw a Poincaré-style conclusion from underdetermination, but it is certainly permitted to do so.

However, one can question whether this specification is appropriate for simultaneity. Why, for instance, does the observer have to stand at the midpoint and not, say, one third of the way towards one end? While it seems reasonable to ground the concepts of a revolutionary new theory in concrete physical operations, it was soon pointed out that Einstein's was not the only possible operational criterion. In this section, we review three classical responses to the simultaneity debate and show how they can be understood through a functionalist lens. First, we turn to the response from Reichenbach (1958); in section 3.2, we discuss the response from Malament (1977); and in section 3.3, we attend to a development of Malament's response.

3.1. Reichenbach: Topological Simultaneity. Reichenbach (1958) argued that an entire family of operationalist criteria are available to give meaning to simultaneity, each one associated with a parameter ϵ taking values between 0 and 1. Since this specification does not uniquely specify the meaning of simultaneity, he concluded that simultaneity is conventional. From the discussion in section 2.4, we can already see that simultaneity becomes realiser-conventional on Reichenbach's account.

Let γ be a timelike line in Minkowski spacetime representing the worldline of an inertial observer, and consider some inertial coordinate system[10] $(\vec{x}, t)$, in which the observer begins at an event $A = (\vec{x}_A, t_A)$ on γ. Some events are accessible to the observer at A via a signal, such as an event $B = (\vec{x}_B, t_B)$ where a light signal from A can be bounced off and returned to the observer at $A' = (\vec{x}'_A, t'_A)$, as in Figure 3. One might reasonably conclude that for this observer, event A is *earlier than* event B. However, in other moments, event B is not accessible to the observer in this way, such as events S and B that cannot be connected by any signal moving at or less than the speed of light. Reichenbach calls these latter events *topologically simultaneous.* It turns out that the time coordinates for the events S on the observer's worldline that are topologically simultaneous with B have a time coordinate given by,

$$t_S = t_A + \epsilon(t_{A'} - t_A) \tag{9}$$

[10]Here we choose a coordinate system while staying uncomitted to any notion of simultaneity. Two events in spacetime can have the same time coordinate without it following that they should be counted as simultaneous.

for some $\epsilon \in (0, 1)$ (Reichenbach 1958, p. 127).

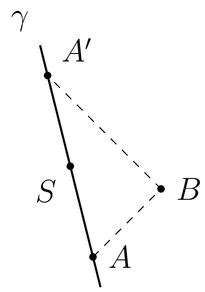

FIGURE 3. Every point S on the line γ between A and A' is topologically simultaneous with B.

Note that if we set $\epsilon = \frac{1}{2}$, Equation 9 identifies t_S as the midpoint between t_A and $t_{A'}$ (see Figure 5). This precisely corresponds to the original definition from Einstein (1905), that B is simultaneous with S on γ if and only if light rays would hit the midpoint at the same time. I will refer to this as *the standard simultaneity relation* for an observer on a worldline. A *non-standard* simultaneity relation can then be defined in terms of values of ϵ between 0 and 1 that are not equal to $\frac{1}{2}$. Visually, different choices of ϵ correspond to differently tilted foliations into surfaces of simultaneous events as in Figure 4. The simultaneity-relation associated with each parameter ϵ will be denoted Sim_ϵ (in particular, the standard simultaneity relation can also be denoted $Sim_{1/2}$).

Let us now see how the Reichenbachian approach fits into the schema from Section 2.3. First, Reichenbach specifies the concept of relativistic simultaneity in terms of an operation involving the bouncing light rays off of distant mirrors. But while Reichenbach agrees with Einstein that a context consisting of concrete operations involving lightrays and mirrors is unproblematic, they disagree as to which operational specification is appropriate for simultaneity. As a consequence, when Reichenbach's functional specification is interpreted in the context of Minkowski spacetime, the entire set of relations $\{Sim_\epsilon \mid 0 < \epsilon < 1\}$ is picked out. Consequently, relativistic simultaneity for Reichenbach is realiser-conventional.

3.2. Malament's Theorems. The functional specifications of simultaneity due to Einstein and Reichenbach are not the only possible ones: in 1977, David Malament

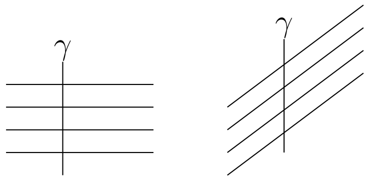

FIGURE 4. Two foliations corresponding to different values of ϵ for an observer moving along γ.

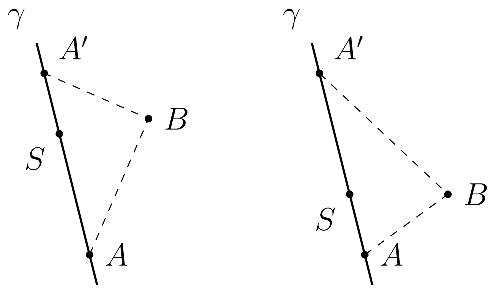

FIGURE 5. The point S indicates the point on γ simultaneous with B. In the left picture, the speed of light is highest in the direction BA'. In the right picture, the speed of light is highest in the AB direction.

proved a theorem that can be interpreted to show that the standard simultaneity relation is the unique realiser of a third functional specification. In this section, I will review Malament's original theorem, together with a recent variant of it, and illustrate the sense in which these results can be understood as functional specifications of simultaneity that are more restrictive than Reichenbach's.

Whereas the functional specification of Reichenbach was motivated by a desire to ground a problematic concept in concrete physical operations, Malament was inspired by the exposition in Robb (1914) of special relativity as a theory of Minkowski spacetimes. Robb (1914) constructs Minkowski spacetime from the earlier-than-relation together with 19 axioms (as such, one could say that Robb offers a functional reduction of Minkowski spacetime based on a number of axioms that he takes to be unproblematic). He writes:

> The special object here aimed at has been to show that spatial relations may be analyzed in terms of the time relations of before and after, [...]. The present work is the outcome of an endeavour to get rid of certain

obscurities in connection with some of the fundamental parts of Physical Science. (Robb 1914, p. v)

This perspective led Malament to approach the question of the conventionality of distant simultaneity as a question of the symmetries of Minkowski spacetime. Malament showed that under the assumption that simultaneity is a time-reversal invariant equivalence relation, and that every point-event is simultaneous with some point-event on the worldline of the observer, the standard simultaneity relation is the only non-trivial candidate. Note how this can be interpreted as Malament offering a third functional specification for simultaneity based on the symmetries of Minkowski spacetime.

Prima facie, one might think that an appropriate simultaneity relation on Minkowski spacetime ought to be invariant under the full Poincaré group, its group of symmetries. But the only equivalence relations on Minkowski spacetime invariant under the full Poincaré group are the two trivial ones, i.e. the relation under which each point event is only simultaneous to itself, and the relation under which every two point events are simultaneous, as I will discuss in detail in section 3.3 (Giulini 2001, Theorem 4). Malament thus motivated his assumption by restricting attention to those symmetries that preserve both the structure of Minkowski spacetime *and* a timelike line representing some observer for whom the simultaneity relation is being defined. The object consisting of Minkowski spacetime with a timelike line has strictly more structure than bare Minkowski spacetime, and hence fewer symmetries than Minkowski spacetime itself, including only the time translations and the time reversal symmetries that preserve the line. As a result, we should expect a bigger class of potential simultaneity relations. Malament's result shows a sense in which imposing a minimal extra requirement (in the theorem this is requirement ii) suffices to enlarge this class by exactly one relation: the standard simultaneity relation.

Theorem 1 (Malament 1977). Let L be a timelike line in Minkowski spacetime, and let S_L be a non-trivial two-place relation on the spacetime that satisfies the following three conditions: i) S_L is an equivalence relation, ii) all events in spacetime are simultaneous with some event on L and iii) S_L is preserved under temporal reflections with respect to hyperplanes orthogonal to L. Then S_L is the standard simultaneity relation due to Einstein (1905).

Using functionalist vocabulary, we can say that Malament (1977) takes the notions of "equivalence relation" connecting to L and "time reversal invariance" to be unproblematic. The former can be motivated by noting that only if simultaneity is an equivalence relation will the simultaneity slices form a partition of spacetime, and the latter by the intuition that judgements of simultaneity should not depend on the direction of time. Insofar as this is correct, these properties can then be taken to yield a functionalist specification of the concept of "relativistic simultaneity". The upshot of Theorem 1 is then that this specification has precisely one realiser in the context of Minkowsi spacetime—namely, the standard simultaneity relation due to Einstein (1905). This is a paradigm example of a successful functionalist reduction according to the schema in section 2.3. If appropriate for simultaneity, then it established that relativistic simultaneity is not conventional but uniquely realised.

In an answer to Malament (1977), Sarkar and Stachel (1999) point out that Malament's uniqueness-result does not go through without the assumption of time reversal symmetry. They argue that this assumption is "unphysical" based on the discovery of time reversal symmetry violation in 1964 (see Roberts (2015)). Interestingly, there is a ready answer to Sarkar and Stachel's critique, which was already noted by Stein (1991) and Spirtes (1981), and which Malament himself presented in an unpublished set of notes (Malament 2009, Proposition 3.4.2, p. 62)[11]. Instead of defining simultaneity relative to a single world line L, one can define simultaneity relative to a congruence $\mathcal{L}$ of inertial co-moving observers, sometimes called a "frame". The structure consisting of Minkowski spacetime with time orientation (to break time reversal symmetry), together with a congruence of timelike lines $\mathcal{L}$, has all the spacetime translations among its symmetries. Apart from the two trivial relations, the only

[11]Indeed, Sarkar and Stachel (1999) are themselves aware of this.

equivalence relation invariant under all the symmetries of this new structure is again the standard simultaneity relation:

Theorem 2 (Spirtes, Stein, Malament 2009). Let $\mathcal{L}$ be a frame and $S_{\mathcal{L}}$ a non-trivial two-place equivalence relation, such that for some $L \in \mathcal{L}$ all events in spacetime are simultaneous with some event on L. Further, suppose $S_{\mathcal{L}}$ is invariant under the actions of translations and reflections in some $L \in \mathcal{L}$ of the two-plane containing L and some other (arbitrary) point $p \in A$. Then S_L is the standard simultaneity relation due to Einstein (1905).

The actual mathematical work being done by re-defining simultaneity relative to a frame makes use of the fact that a frame is invariant under spatio-temporal translations and rotations. Hence, the upshot of Theorem 2 is that time-reversal invariance can be substituted for invariance under translations and reflections, thus providing an alternative functional specification of relativistic simultaneity, which also yields the standard simultaneity relation uniquely in the context of Minkowski spacetime. It is worth pointing out that all it would take to oppose either of Malament's specifications is to reject one of the assumptions, say, that simultaneity must be an equivalence relation or necessarily invariant under any particular symmetry. We have here a case of two specifications for the same concept differing only in terms of which particular symmetries simultaneity is said to be invariant under. It does not yet amount to specification-conventionality, since the specifications in theorems 2 and 3 are both uniquely realised by the standard simultaneity relation. However, in the next section we will see yet another variation from Sarkar and Stachel (1999) with a different realiser, which does give rise to specification-conventionality.

Malament (1977) was extremely influential for the subsequent debate, with authors such as Wesley et al. (1992) taking the debate to have been settled in favour of non-conventionality:

> Contrary to most expectations, he was able to prove that the central claim about simultaneity of the causal theorists of time was false. (Wesley et al. 1992, p. 222)

There were also critics like Grünbaum, who (in private correspondence to Salmon) laments that by taking the relation of simultaneity to be an equivalence relation, one rules out any relation that fails to partition spacetime, such as Wesley et al. (1992); Grünbaum (2010), (the relation of topological simultaneity described in Section 3.1)[12]. Another critic is Brown (2005), whose commitment to the "dynamics-first" approach makes him discount Malament's geometric approach to the question[13] (Brown 2005, pp. 98-102). Brown argues that since Malament only countenances a timelike line and the conformal structure of Minkowski spacetime, there simply is not enough structure to have something deserving of the label "time". Brown ultimately endorses conventionalism with respect to simultaneity, but in a later chapter he discusses the non-relativistic limit and after having derived the Galilean transformation, finds it pertinent to assure the reader that the resulting classical simultaneity relation is the appropriate one:

> But there is a last twist in the story. The Galilean transformations (6.63) are formally right (or would be if they were exact) but what guarantees that the way time is spread through space corresponds to the synchrony convention in Newtonian mechanics? It is here that the Eddington–Winnie theorem, mentioned in section 6.3.2, shows its worth. For it establishes that the Poincaré–Einstein synchrony convention in S is equivalent to that of slow clock transport, and that is all we need. (Brown 2005, p. 111)

Or, in other words: the classical limit of relativistic simultaneity is the well-known relation of Newtonian simultaneity. As will soon become clear, I wholeheartedly agree

[12]There have been many other critiques of Malament. See e.g. Jannis (1983), Redhead (1993) and Debs, Talal and Redhead (1996). They point out that if we denote the standard simultaneity with respect to a timelike line L (the world line of an inertial observer) by Sim_L and if we let L, L' be two distinct timelike lines, we get two distinct simultaneity relations Sim_L and $Sim_{L'}$. The point of Debs, Redham and Janis is that if we construe Sim_L as the relation of simultaneity for an inertially moving observer whose worldline is L', we get a non-standard simuiltaneity relation, which they claim is not ruled out by Malament's theorem. However, one can dismiss this objection insofar as we are seeking a concept of simultaneity-for-an-observer, and take the question of matching relations indexed by timelike lines $\{Sim_L\}$ with inertial observers as unproblematic.

[13]Note that it is not my intention to survey the different ways in which simultaneity has been motivated in special relativity but only to evaluate the conventionality claim.

with Brown that considerations of the classical limit are highly relevant to the question of the status of simultaneity. I take this insight further, considering not only the limit of the standard simultaneity relation but of a whole host of alternative candidates.

Even among those taking a "geometry-first" approach there is a substantial question of exactly which symmetries of Minkowski spacetime "simultaneity" ought to be invariant under. This leads us to the critique of Sarkar and Stachel (1999).

3.3. Sarkar and Stachel: Backwards Lightcones. Why should we think that simultaneity is invariant under translations and rotation and not, for example, invariant under boosts? All three are symmetries of Minkowski spacetime, but as mentioned above, they cannot all be symmetries of a non-trivial relation. Which one do we choose? Sarkar and Stachel (1999, p. 218) make a similar point, which is made precise by the Giulini (2001, Theorem 4) No-Go theorem.

Theorem 3 (Non-Existence). Let L be a timelike line in Minkowski spacetime, and let S_L be a two-place relation on the spacetime that satisfies the following three conditions: i) S_L is an equivalence relation, ii) all events in spacetime are simultaneous with some event on the worldline and iii) S_L is invariant under both the actions of translations by a fixed spacetime vector as well as boosts. Then S_L is trivial.

The upshot is that a particularly greedy functional specification insisting on both boost-invariance and invariance under translations and reflections suffers from what we called realiser-non-existence in Section 2.2: there is simply nothing in Minkowski spacetime answering to the functional specification. Consequently, this functional specification cannot help us make sense of simultaneity in the context of Minkowski spacetime, except to advise that it does not exist. I will return to this possibility in section 5.

We have already seen that time reversal invariance and invariance under translations and rotations both yield the standard simultaneity relation. But, what would happen if we substituted in boost invariance instead? It turns out we get the so-called "backwards lightcone criterion". Einstein (1905, p.2) considered this criterion:

> We might, of course, content ourselves with time values determined by an observer stationed together with the watch at the origin of the co-ordinates, and co-ordinating the corresponding positions of the hands with light signals, given out by every event to be timed, and reaching him through empty space. Einstein (1905)

In modern parlance, the criterion states that two events are to be counted as simultaneous for an inertial observer on a worldline L if they lie on the surface of a backwards lightcone defined by that observer as in Figure 6.

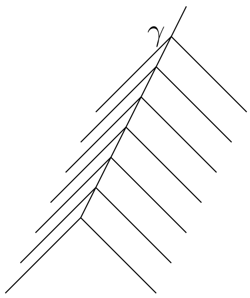

FIGURE 6. The backwards lightcone simultaneity criterion for an inertial observer following the worldline γ.

Suppose we adopt a functional specification that is very similar to the one in Theorem 2 except where invariance under translations and reflections is substituted for invariance under boosts, and where the condition that simultaneity must be an equivalence relation is dropped. Then we get the following result suggested by Sarkar and Stachel (1999)[14]:

THEOREM 4 (Backwards Lightcone Criterion). Let L be a timelike line in Minkowski spacetime, and let S_L be a non-trivial two-place relation on the spacetime that satisfies the following three conditions: i) S_L is an equivalence relation, ii) all events in spacetime are simultaneous with some event on L and iii) S_L is invariant under boosts. Then S_L is the backwards lightcone criterion.

Interestingly, Einstein immediately rejects the criterion for not being invariant under spatial translation, but without explaining why translation invariance is

[14]They do not provide a rigorous proof.

somehow more important than invariance under boosts[15]. In the language of functionalism, Sarkar and Stachel (1999) countenance two different functional specifications, each having a different, unique realiser: one based on invariance under spatio-temporal translation and one on invariance under boosts. But this means that we have to make a choice between two conventions and, provided that each convention is equally plausible, this choice exhibits specification-non-uniqueness and thus becomes specification-conventional. This means that the literature at best establishes that relativistic simultaneity is specification-conventional.

To provide a path forward, the next two sections introduce and expound my second thesis: that the only viable functionalist constraint on the concept of relativistic simultaneity is its behaviour in the non-relativistic limit.

4. Conventionality in the Classical Limit

4.1. The Classical Simultaneity Thesis. Giulini (2001), discussing the theorems by Malament and by Sarkar and Stachel, makes the following observation:

> Here the issue of uniqueness comes in because one strategy adopted to refute this thesis [the conventionality thesis] is to first identify non-conventionality with uniqueness and then to prove the latter. Clearly, this identification can be challenged upon the basis that every proof of uniqueness rests upon some hypotheses which the simultaneity relation is supposed to satisfy and which may themselves be regarded as convention

This is precisely the situation we find ourselves in with three different functional specifications resulting in three different sets of simultaneity-relations on Minkowski spacetime. Belot (2010) reaffirms the difficulty of the situation:

[15]One could object that while translations and reflections preserve the world line, boosts do not, and therefore the two specifications are not equally plausible. However, note that boosts do preserve Minkowski spacetime itself, so the breaking of symmetry only happens on the level of Minkowski spacetime plus world line. Further, it is unclear why we should expect this extra bit of structure to be preserved. This is not uncontroversial with either authors such as Hinchliff (2000) or Sarkar and Stachel (1999) explicitly considering the backwards lightcone criterion.

> Now, while every student ought to be exposed to this observation[Malament's Theorem], it does little to settle the question of the nature of time in special relativity[...] This is quite typical: symmetry arguments are of little polemical value in situations where fundamental questions are at stake, since those are the cases in which there will be little agreement as to whether a given structure provides an acceptable point of departure for such an argument. (Belot 2010, pp. 396-397)

Luckily, the functionalist framework can help us to avoid a free-for-all as to what provides an acceptable point of depature for relativistic simultaneity. The question we have to start with is this: what makes a functional specification, a specification of *simultaneity*? Clearly not all relations on Minkowski space-time are plausible candidates for relativistic simultaneity, lest every single concept becomes conventional. It is not, for instance, any help to consider a relation that countenances timelike separated events as simultaneous, nor does it have any interesting implications for the conventionality thesis. But how do we know that "x and y are timelike separated" is not a plausible candidate for relativistic simultaneity? What is responsible for the constraints on what relativistic simultaneity might mean? Or, using functionalist vocabulary, what specifications are specifications of simultaneity?

The word "simultaneity" and its synonyms, such as "synchronous" and "contemporaneous", were used in ordinary English long before Einstein (1905). Jammer (2006) tracks the concept back to the fifth century BCE, when Aristotle analysed Zeno's paradoxes in the *Physics*. Moreover, popular and scientific notions of simultaneity have coincided at least since Newton formalised the idea of absolute time, and the related notion of absolute simultaneity, in his Principia. These pre-relativistic notions are what allow us to make sense of the question "What does relativistic simultaneity mean?". This is because the only context in which the concept of simultaneity is completely unproblematic is in this pre-relativistic popular-scientific consensus. Consequently, this is a natural concept to transport to relativity theory when we seek the meaning of relativistic simultaneity.

Once we recognise that what makes us able to discern talk of the conventionality of relativistic simultaneity from talk of some other relation on Minkowski space-time is our knowledge of the pre-relatistic concept, the question of which relations are plausible candidates for relativistic simultaneity reduces to the question of how the Newtonian concept constrains the relativistic concept[16]. This moreover suggests an answer to the question I just posed: if the functional specifications of Reichenbach, Malament and Stachel and Sarkar are specifications of *simultaneity*, it is because they are appropriately related to the classical concept. The question then is how to bear out "appropriately related".

Fortunately, Minkowski spacetime and Newtonian spacetime stand in a well understood and mathematically rigorous relationship: Newtonian spacetime comes about as the classical limit of Minkowski spacetime. Intuitively, that limit is characterised by the idea that Minkowski spacetime becomes Newtonian in the limit where the speed of light tends to infinity. But, this can be made completely precise[17]. However, what we really need is not the limit of Minkowski spacetime but the limit of putative simultaneity relations defined on Minkowski spacetime. Fortunately, the classical limit of a spacetime gives a natural way of understanding the classical limit of a relation defined on that spacetime (see section 4.2). This allows us to verify what I will call "the classical simultaneity thesis" (CST) in the following manner:

[Strong CST Limit Thesis] A relation on Minkowski spacetime is a realiser of simultaneity if and only if it converges on classical simultaneity in the classical limit.

It may be that there are other ways for a relation to be "appropriately connected" to classical simultaneity than through the classical limit, but my subsequent argument only hinges on this condition being sufficient. So if the weaker version is more palatable, you may adopt:

[16] One might worry here that we allow ourselves to be constrained in our theorising by a demonstrably false theory—Newtonian Mechanics. But Newtonian Mechanics is not just any false theory: it was, and still is, wildly succesful in a well-circumscribed domain. Further, I am not claiming that the relativistic notion needs to *be* the classical notion, but merely that we should be able to make sense of the success of Newtonian theory from the vantage point of relativity theory.

[17] For a technical exposition of the classical limit see Malament (1986), who draws on the framework of Künzle (1972, 1976).

FIGURE 7. A manifold of smoothly connected events. We have not yet imposed any coordinates.

[Weak CST Limit Thesis] A relation on Minkowski spacetime is a realiser of simultaneity if it converges on classical simultaneity in the classical limit.

As we shall see shortly, the set of relations converging on classical simultaneity in the classical limit is both huge and extremely heterogenous.

Last, note that I limit myself to the question of extension. Of course, there is a question of how to export something like the Fregean sense of a concept, but I will only be concerned with the export of reference[18]. Section 4.2 explains the main idea mainly through pictures, although these can be made mathematically precise, arguing that every conceivable simultaneity relation will converge on Newtonian simultaneity in the classical limit. I will argue that this suggests that relativistic simultaneity not only fails to be conventional, but that it is even devoid of content.

4.2. The Classical Limit. I will begin by explaining the classical limit of relations on Minkowski spacetime, with the aim of producing the set of relativistic relations realizing simultaneity in accordance to CST.

First we need the idea of a spacetime manifold (see Figure 7), containing a smoothly connected set of events in space and time[19]. To simplify the illustrations, my diagrams will suppress all but one spatial dimension. At this point the spacetime in Figure 7 could be either Newtonian or Relativistic.

[18]For a discussion of meaning vs. referent of "simultaneity", see Friedman (1977).
[19]For our purposes, the manifold of events is going to be $\mathbb{R}^4$ but nothing hinges on this.

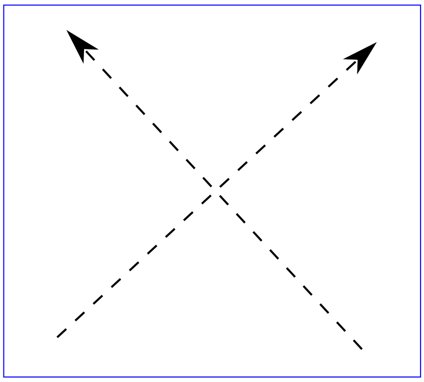

Figure 8. Two lightrays going left and right, respectively.

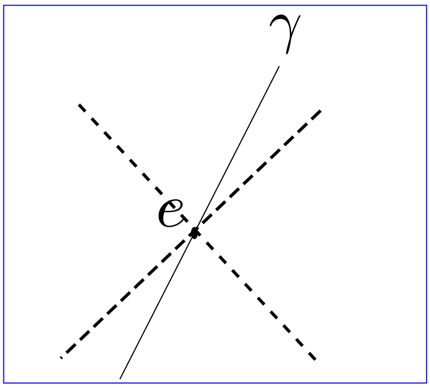

Figure 9. A worldline γ for an observer moving to the right with constant velocity and a lightcone for the point e on the world line.

Next we distinguish between Newtonian spacetime of classical physics and Minkowski spacetime of special relativity. A feature of the latter is that light rays move with speed c (approximately $3 \cdot 10^5$ kilometers pr. second) relative to any inertial reference frame. For convenience, we define units so that $c = 1$, and draw light rays at 45 degree angles in Minkowski spacetime[20] (see Figure 8).

Since nothing can move faster than the speed of light in special relativity, there is a sense in which the light rays through a point in Minkowski spacetime demarcate the region of spacetime that is causally accessible to an observer at that point. We can represent this by drawing a "lightcone" through each point (see Figure 9).

[20]Nothing hinges on this. The light rays could be drawn asymmetrically to indicate that the one-way speed of light in one direction differs from the one-way speed of light in the other direction. Of course, the average speed of light has to be c to save the phenomena.

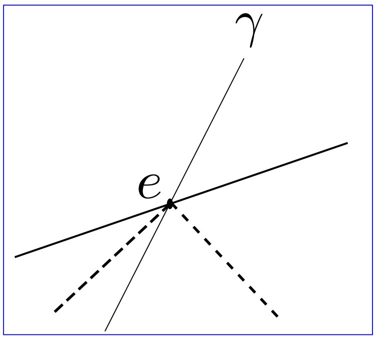

FIGURE 10. The standard simultaneity relation and the backwards lightcone criterion of simultaneity for the point e.

A lightcone around a point ("the middle vertex") divides spacetime into three areas[21]. Inside the lightcone we have two sections: the bottom area is the section of spacetime from which our observer could have been causally influenced; the top area is the section of spacetime that our observer could themself causally influence. Outside the lightcone, the right and left areas form the section of spacetime too far away from our observer for any signals to be able to travel back and forth. For this reason, it is often said that the lightcones represent the "causal structure" of relativistic spacetime. We say that points inside the lightcone are "timelike" separated from the middle vertex, whereas points outside the lightcone are "spacelike" separated from the middle vertex[22][23].

At this point we can illustrate some of the candidate simultaneity relations we have encountered above. In Figure 10, I have drawn the backwards lightcone criterion and the standard simultaneity relation. Note that both the backwards lightcone criterion and the standard simultaneity relation define simultaneous events to be non-timelike separated. This is also the case with Reichenbach's Sim_ϵ relations (see figure 11).

[21]In $2D$, the left- and right-hand sides of the lightcone are separated, but in more than 2 dimensions one can simply "go around" the middle vertex.
[22]More precisely, a vector $\xi^a \in T_pM$ is timelike if $\xi^a\xi_a > 0$, spacelike if $\xi^a\xi_a < 0$ and lightlike if $\xi^a\xi_a = 0$ where $\xi^a\xi_a = g_{ab}\xi^b\xi_a$ with g_{ab} being the metric tensor.
[23]Note that all the simultaneity slices drawn in Figures 8 and 9 are for an observer travelling along γ. Note also that although the lightcones are drawn symmetrically here the speed of light need not be isotropic.

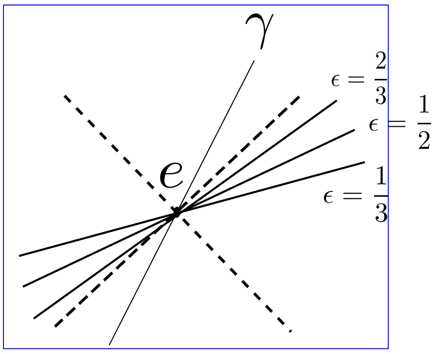

FIGURE 11. The simultaneity slices for $Sim_{\frac{3}{4}}$, $Sim_{\frac{1}{2}}$ and $Sim_{\frac{1}{3}}$ are drawn. The lightcone is included for reference.

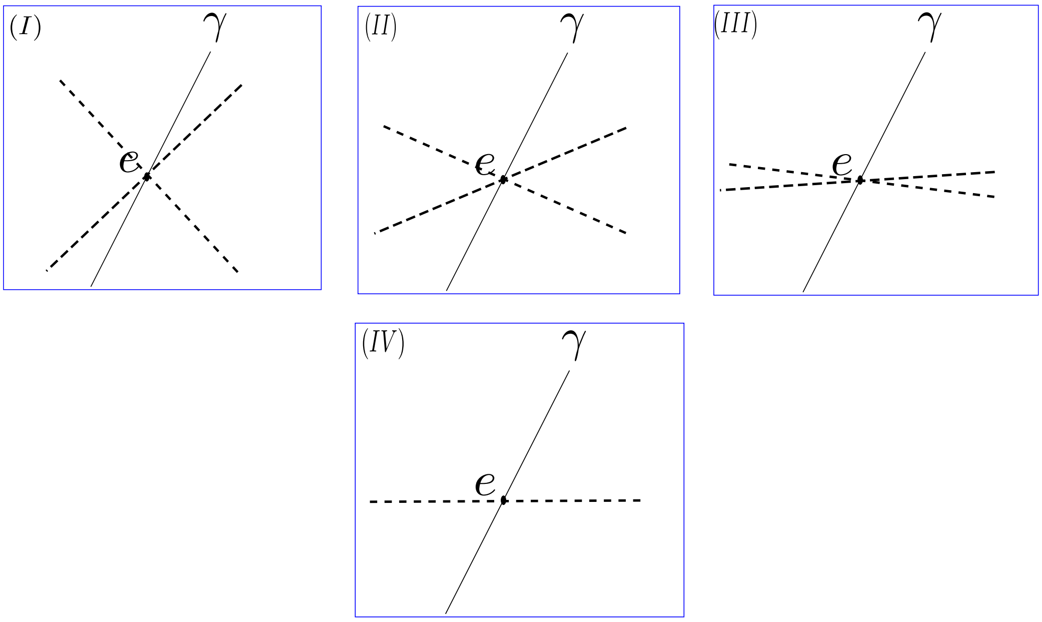

FIGURE 12. The lightcones flatten as the speed of light increases from (I) through (III). In (IV) we reach the limit $c \to \infty$ and the lightcone has disintegrated into a line.

So everyone seems to agree that the simultaneity slices are located in the non-timelike region. But, what happens to this region when we take the classical limit? Recall that a way of understanding this limiting process is by letting the speed of light tend to infinity. We can represent the changing speed of light visually by letting the light cones flatten[24] (see Figure 12).

[24]Higher speed means less time lapsing for the same distance. In our spacetimes, we have time up the y—axis and so higher speeds means flatter lines.

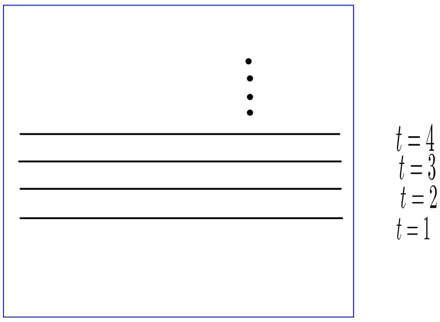

FIGURE 13. Four Newtonian simultaneity slices corresponding to the global times $t = 1$, $t = 2$, $t = 3$ and $t = 4$.

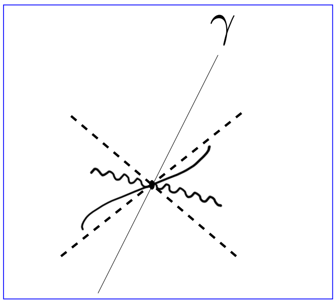

FIGURE 14. Two "odd" relations that will both tend to Newtonian simultaneity in the limit.

As every candidate simultaneity relation so far has been located in the side-areas of the light cones, they must therefore all be squished to the line in (IV). But the line in (IV) is just a Newtonian simultaneity slice[25][26] (see Figure 13).

So, all of Reichenbach's Sim_ϵ-relations (including the standard simultaneity relation) and the backwards lightcone criterion become squished onto Newtonian simultaneity slices in the classical limit. As a matter of fact, *any* relation locating the simultaneity slices in the side-areas of the light cones will tend to Newtonian simultaneity in the limit. I have drawn a few examples in Figure 14.

What is suggested by these diagrams can be stated as a theorem, which says that an enormous class of simultaneity relations incuding all those currently occurring

[25]Again, it would make no difference if the speed of light was not isotropic. The lightcones would be skewed, but as long as the one-way speed of light in any direction tends to infinity, the lightcones would collapse onto Newtonian simultaneity slices.

[26]In drawing the lightcones, I have implicitly assumed a rest frame with a vertical t-axis through e. Choosing a different rest frame (say, the rest frame of the observer moving along γ) would simply tilt all the light cones uniformly and result in a different foliation of Newtoian slices, tilted accordingly.

in the literature, all have the same classical limit. As a result, they are all equally viable simultaneity relations from the perspective of the classical simultaneity thesis. A proposition makes this precise[27], covering not only the cases of Reichenbach's Sim_ϵ-relations and the backwards lightcone criterion, but also a whole host of other possible relations.

PROPOSITION 1. Given Minkowski spacetime, and given a putative simultaneity relation that has some classical limit with respect to a Newtonian spacetime, the following holds: if the simultaneity relation consists of a foliation into spatial hypersurfaces, then it converges in the classical limit to the Newtonian hypersurfaces of simultaneity. A similar conclusion holds if the simultaneity relation consists of a foliation into backwards (or forwards or indeed entire) lightcones.

In the appendix, I will lay out precisely what it means for a putative simultaneity relation to have a classical limit and provide a proof of the proposition. For now, it suffices to note that not only all Reichenbach's Sim_ϵ-relations together with the backwards lightcone criterion will converge on Newtonian simultaneity, but *all* relations foliating spacetime into spacelike or lightlike hypersurfaces. This class is both enormous and highly heterogeneous—see e.g. Figure 14.

5. Functionalism and Conventionality Revisited

In the case of relativistic simultaneity, Reichenbach's specification exhibits realiser-conventionality. Further, the existence of alternative specifications due to Malament and to Sarkar and Stachel means that relativistic simultaneity suffers from specification-conventionality as well.

But the size of the conventionality matters. If we learned that some T-term was realised by every single T'-term, we would probably conclude that the T-term was devoid of meaning in the context of T', and not that it was conventional. The existence of precisely one realiser means that the concept in question is factual and

[27]Proof in the appendix.

non-conventional; a moderate number of realisers means that the concept is conventional and factual; but, as the number of realisers climbs, the concept gradually loses its content and becomes trivial. On my account, conventionality refers to the multiple realisability of a term in a specified theoretical context. On the other hand, a term can be non-trivial even though it is multiply realised: the loss of content happens gradually as the number of realisers increase.

It is not only the number of realisers that matters for assessing the content of a concept, but also the level of heterogeneity among the realisers must be taken into account. If a concept is multiply realised, one way of thinking about that concept is in terms of whatever all the realisers have in common. We can think of "illness" as that which all illnesses have in common and part of what makes "illness" a useful concept is that we can find a system in the class of phenomena qualifying as illnesses (maybe "illness" just *is* the system). If someone insisted that "happiness" should also count as an illness, we would rightly feel that our concept of illness was somehow broken.

We can now pull everything together: I argued in section 4 that the only thing constraining what "relativistic simultaneity" might mean is our understanding of the classical context. This constraining-relation was borne out via the Classical Simultaneity Thesis, saying that a relation on Minkowski spacetime is a realiser of simultaneity if it converges on Sim_N in the classical limit. But now we have seen that the set of relations realising simultaneity on Minkowski spacetime contains (or equals if we adopt the strong Classical Simultaneity Thesis which states that converging on Sim_N in the classical limit is both necessary and sufficient for being a realiser of simultaneity) the set of all relations locating the simultaneity slices in the side-areas of the lightcones. This set is both gargantuan and highly heterogeneous.

This gives us two options: either we conclude that "simultaneity" is devoid of content in the context of Minkowski spacetime, on pain of the realising set being too large and too heterogeneous, or we allow that maybe all the elements in the realising set have something in common after all. Since the latter option seems more neutral, it is my preferred option, and if we take this route, I think the natural option is to let "simultaneneous with p" mean "related to p under the first realiser, or related to p

under the second realiser, or...". But this just amount to letting "simultaneneous with p" mean "non-timelike separated from p", which is just Reichenbach's definition of topological simultaneity. In either case we have no use for "simultaneity" in relativity theory—either because the concept is devoid of meaning, or because it collapses into a different concept already at our disposal.

Before concluding, let us consider the objection that takes the above as a reductio of the CST. Surely, this objection would go, since the CST places such a weak constraint on the meaning of simultaneity, we must ipso facto be able to conclude that there is more to this concept than the CST would suggest. What this objection overlooks is that "simultaneity" was not a new term in 1905 when Einstein published "On the Electrodynamics of Moving Bodies". Rather, the term has a long history both as a word in ordinary language and as a technical term in Newtonian Mechanics. Hence, if we do not intend to simply reassign a meaning to the string of symbols "simultaneity", the constraints on what we might take relativistic simultaneity to be must come from the classical context. Last, note that over 70 years of analysis has not produced any agreed upon alternative to the CST. On the other hand, the CST justifies every candidate relation presented in the literature as a candidate for "simultaneity" and simply takes the logic to its natural conclusion.

6. Conclusion

Having shown that virtually all interesting relations on Minkowski spacetime one can write down will converge to a Newtonian simultaneity slice, we conclude that "simultaneity" is a concept entirely foreign to the special theory: either because it is devoid of meaning due to the realising set being too large and too heterogeneous, or because "simultaneity" collapses into a different concept already at our disposal. This invites us to radically rethink the nature of time and temporal relations in the relativistic setting and suggests that, rather than asking about the status of simultaneity in relativity, we should begin searching for physically interesting relations endemic to special relativity instead.

CHAPTER 5

Summary and Open Questions

1. Summary

In chapter 2, I offered a framework through which to understand the debate over conventionalism as a debate over different equivalence relations on the space of theories. This led to two strategies for avoiding being fooled by apparent underdetermination: "Elimination" aimed at not mistaking the peaceful co-existence of multiple notational variants of one theory for the potentially perilous one of multiple different theories between which an epistemically well-motivated choice is impossible. In terms of equivalence relations, elimination consists of the adoption of a weak relation such that many theories will come out equivalent. Naturally, equivalent theories do not threaten underdetermination.

"Discrimination", on the other hand, aimed at not overlooking the existence of epistemically potent reasons for choosing one theory over the alternatives. In terms of equivalence relations, discrimination consists in the adoption of a strong equivalence relation, such that few theories will come out equivalent, together with an argument that the reasons cited are available to the chooser to form the basis for a choice. As we saw in chapter 2, I recommend a version of discrimination based on the capacity of formalisms to aid in the process of developing new theories and models. Specifically, I offered three mechanisms through which a formalism can do this work: "clear expression", "generalisability" and "uniformity". I further argued that differences between formalisms along these lines are epistemically potent and should therefore be taken seriously.

In chapter 3, I conducted a case study using the analysis developed in chapter 2 on the case of Einstein Algebras and Lorentzian Manifolds. Here we saw an example of two formalisms that on natural interpretations performed differently for the purposes of developing a theory of quantum gravity despite being mathematically equivalent[1]. In particular, I discussed the significance of the one-to-one correspondence between the spacetime points in a manifold M and the maximal ideals of the ring $C^{\infty}(M)$. This led me to review an argument made by Lam and Wüthrich that due to this one-to-one correspondence, the move to Einstein Algebras yields an elimination of spacetime points "in name only". I offered two independent arguments against this view: one based on the existence of an alternative construction of "points" in an Einstein Algebra and one based on the capacity of Einstein Algebras to point towards a theory of quantum gravity. Since the differences in heuristic function pointed to here are too subtle for any of the purely formal criteria for theory equivalence suggested in the recent literature to be able to discern, my argument is simultaneously an argument against the completeness of these criteria. Chapter 3 ended with a discussion of what one plausible *can* learn from a purely formal equivalence result.

Where chapters 2 and 3 dealt with conventionality as underdetermination, chapter 4 offered a novel interpretation of the question of the conventionality of relativistic simultaneity as a question of the transport of concepts from one theoretical context to another. I developed a mechanism for transporting concepts between theoretical contexts based on functionalism together with an analysis of when the transported concept exhibits conventionality. The upshot was a grounding of the meaning of relativistic simultaneity in the well-known concept of classical, absolute simultaneity. Applying the mechanism, I showed how to transport the classical concept to the theoretical context of Minkowski spacetime. Via the classical limit, I argued that "simultaneity" is fundamentally foreign to Minkowski spacetime on pains of a kind of violent multiple realisability.

I finish this thesis with a number of open questions raised by the present analysis.

[1]They are categorically dual (Rosenstock et al. 2015).

2. Open Questions

2.1. Mechanisms to Facilitate Construction. In chapter 2, we saw three mechanisms through which the choice of formalism can aid the construction of new theories: clear expression, generalisability and uniformity. This list is not meant to be exhaustive, and it would be an interesting project to identify other such mechanisms. In a footnote, I pointed to Glymour and Laudan and Leplin who suggest discriminating on the basis of differential support for equivalent theories from the same piece of evidence, but undoubtably other mechanisms can be found.

2.2. Realism and Constructive Equivalence. Dürr and Read (2023) identifies conventionalism as a selective anti-realism. The idea is that the empirically minded realist can save face when confronted with multiple geometries by insisting that geometric claims are conventional rather than "about the world". The same logic can apply to the more general case of the underdetermination of theories. In a number of cases the theories involved will be constructively inequivalent and consequently my recommendation will be to insist on their difference and to use their constructive inequivalence to discriminate between them. In chapter 2, I argued that the differences discerned by $\sim_C$ are epistemically potent, on the basis of a view of science as continuously open. It would be interesting to built out the resulting view of science and in particular classify the status of theoretical values such as simplicity and coherence. If these come out as epistemic values, traditional metaphysics might be validated in some of its conclusions (albeit for very different reasons than the ones traditionally cited).

2.3. Einstein Algebras and Quantum Gravity. In chapter 3, I argued that Einstein Algebras might be expedient for the development of theories of quantum gravity. It is going to be interesting to see what we can learn about space and time from a functioning theory of quantum gravity by e.g. relocating relativistic spacetime in an appropriate limit. This might provide further evidence that an Einstein Algebra based theory of spacetime is superior to one based on Lorentzian Manifolds.

2.4. Transport of Other Concepts. In chapter 4, I developed a method for transporting concepts from one theoretical context to another. It would be interesting to apply this method to other concepts and other contexts. Perhaps one could transport "mass" or "momentum" to special relativity or quantum mechanics.

Bibliography

Abraham, R. and Marsden, J. E. (1978). *Foundations of Mechanics*, Addison-Wesley Publishing Company, Inc.

Arnold, V. (1974). *Mathematical Methods Of Classical Mechanics*, Springer-Verlag.

Awodey, S. (2006). *Category Theory*, Oxford University Press.

Bacelar Valente, M. (2018). The Gauge Interpretation of the Conventionality of Simultaneity, *Lato Sensu: Revue de la Société de philosophie des sciences* **5**(2): 1–13.

Bain, J. (2013). Category-theoretic structure and radical ontic structural realism, *Synthese* **190**(9): 1621–1635.

Balashov, Y. and Janssen, M. (2003). Presentism and relativity, *British Journal for the Philosophy of Science* **54**(2): 327–346.

Barrett, T. W. (2019). Equivalent and Inequivalent Formulations of Classical Mechanics, *British Journal for the Philosophy of Science* **70**(4): 1167–1199.

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

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

Belot, G. (2010). Symmetries in physics, *AIP Conference Proceedings*, Vol. 1271, pp. 65–89.

Ben-menahem, Y. (1990). Equivalent Descriptions, *British Journal for the Philosophy of Science* **42**(1): 1–19.

Ben-menahem, Y. (2006). *Conventionalism*, Cambridge University Press.

Ben-Yami, H. (2015). Causal Order, Temporal Order, and Becoming in Special Relativity, *Topoi* **34**(1): 277–281.

Ben-Yami, H. (2019). Absolute Distant Simultaneity in Special Relativity, *Foundations of Physics* **49**(12): 1355–1364.

Brown, H. (2005). *Physical Relativity*, Oxford University Press.

Butterfield, J. (1989). The Hole Truth, *British Journal for the Philosophy of Science* **371**(6525): 116–119.

Butterfield, J. and Gomes, H. (2023). *Functionalism as a Species of Reduction*, Springer International Publishing, pp. 123–200.

Butterfield, J. and Isham, C. (1999). Spacetime and the philosophical challenge of quantum gravity, *Physics Meets Philosophy at the Planck Scale*, number 2000, pp. 33–89.

Cao, C. and Carroll, S. M. (2018). Bulk Entanglement Gravity without a Boundary: Towards Finding Einstein's Equation in Hilbert Space, *Physical Review D* **97**(8): 1–29.

Cao, C., Carroll, S. M. and Michalakis, S. (2017). Space from Hilbert Space: Recovering Geometry from Bulk Entanglement, *Physical Review D* **95**(2): 1–37.

Carnap, R. (1922). *Der Raum: Ein Betrag zur Wissenschaftslehre*, Vol. 56 of *Kant-Studien, Ergänzungshefte.* Doctoral Thesis, translated to English as, "Space: A Contribution to the Theory of Science" in *The Collected Works of Rudolf Carnap, Volume 1: Early Writings*, A.W. Carus et al. (Eds.) 2019, Oxford: Oxford University Press.

Connes, A. (1994). *Noncommutative Geometry*, Academic Press, San Diego.

Curiel, E. (2014). Classical Mechanics Is Lagrangian; It Is Not Hamiltonian, *British Journal for the Philosophy of Science* **65**(2): 269–321.

Debs, Talal and Redhead, M. (1996). The Twin "Paradox" and the Conventionality of Simultaneity, *American Journal of Physics* .

Dirac, P. A. M. (1947). *The Principles of Quantum Mechanics*, Oxford University Press.

Dizadji-Bahmani, F., Frigg, R. and Hartmann, S. (2010). Who's Afraid of Nagelian Reduction?, *Erkenntnis* **73**(3): 393–412.

Dürr, P. and Read, J. (2023). Reconsidering conventionalism: An invitation to a sophisticated philosophy for modern (space-)times.
URL: *https://philsci-archive.pitt.edu/22172/*

Earman, J. (1989). *World Enough and Space-Time*, MIT University Press.

Einstein, A. (1905). On the Electrodynamics of Moving Bodies, *Annalen der Physik* .

Einstein, A. (1921). *Geometrie und Erfahrung*, Verlag von Julius Springer, Berlin. Extended version of a lecture held on 21 January 1921 at the Prussian Academy of Sciences in Berlin, available via the Einstein Papers project: `https://einsteinpapers.press.princeton.edu/vol7-doc/431`, pagination from the English translation 'Geometry and Experience' in Einstein, Ideas and Opinions, trans. Sonja Bargmann (New York: Crown, 1982), `https://einsteinpapers.press.princeton.edu/vol7-trans/224`.

Feintzeig, B. H. (2018). The classical limit of a state on the Weyl algebra, *Journal of Mathematical Physics* **59**(11): 1–23.

Feintzeig, B. H. (2020). The classical limit as an approximation, *Philosophy of Science* **87**(4): 612–639.

Feynman, R. (1965). *The Character of Physical Law*, BBC.

French, S. (2010). The interdependence of structure, objects and dependence, *Synthese* **57**(5): 613–627.

French, S. (2011). Metaphysical underdetermination: Why worry?, *Synthese* **180**.

Friedman, M. (1977). Simultaneity in Newtonian Mechanics and Special Relativity, *Minnesota studies in the philosophy of science* **8**: 403–432.

Frigg, R. and Votsis, I. (2011). Everything you always wanted to know about structural realism but were afraid to ask, *European Journal for Philosophy of Science* **1**(2): 227–276.

Geroch, R. (1972). Einstein Algebras, *Communications in Mathematical Physics* .

Giulini, D. (2001). Uniqueness of Simultaneity, *British Journal for the Philosophy of Science* **52**.

Glymour, C. (1970). Theoretical Realism and Theoretical Equivalence, *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp. 275–288.

Glymour, C. (1977). The Epistemology of Geometry, *Noûs* **11**(3): 227–251.

Grünbaum, A. (1973). *Philosophical Problems of Space and Time*, Vol. 29, D. Reidel Publishing Company.

Grünbaum, A. (2010). David Malament and the Conventionality of Simultaneity: A Reply, *Foundations of Physics* .

Guerra, F., Leone, M. and Robotti, N. (2012). When Energy Conservation Seems to Fail: The Prediction of the Neutrino, *Science and Education* **23**(6): 1339–1359.

Harnik, V. (2011). Model theory vs. categorical logic: two approaches to pretopos completion, *Centre de Recherches Mathématiques CRM Proceedings and Lecture Notes*, Vol. 53, pp. 79–106.

Hawking, S.W. and Ellis, G. (1973). *The large scale structure of space-time*, Cambridge University Press.

Heller, M. (1992). Einstein algebras and general relativity, *International Journal of Theoretical Physics* **31**(2): 277–288.

Hinchliff, M. (2000). A Defense of Presentism in a Relativistic Setting, *Philosophy of Science* **67**.

Hudetz, L. (2017). The semantic view of theories and higher-order languages, *Synthese* **196**(3): 1131–1149.

Hudetz, L. (2019). Definable Categorical Equivalence, *Philosophy of Science* **86**(January): 47–75.

Jammer, M. (2006). *Concepts of Simultaneity*, The John Hopkins University Press.

Janis, A. (2018). Conventionality of Simultaneity, *in* E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2018 edn, Metaphysics Research Lab, Stanford University.

Jannis, A. (1983). Simultaneity and Conventionality, D. Reidel Publishing Company, pp. 101–110.

Künzle, H. P. (1972). Galilei and Lorentz structures on space-time: comparison of the corresponding geometry and physics, *Annales de l'I.H.P.* **4**: 337–362.

Künzle, H. P. (1976). Covariant Newtonian Limit of Lorentz Space-Times, *General Relativity and Gravitation* **7**(5): 445–457.

Ladyman, J. (1998). What is structural realism?, *Studies in History and Philosophy of Science Part A* **29**(3): 409–424.

Ladyman, J., Ross, D., Spurrett, D. and Collier, J. (2007). *Everything Must Go*, Oxford University Press.

Lal, R. and Teh, N. (2017). Categorical generalization and physical structuralism, *British Journal for the Philosophy of Science* **68**(1): 213–251.

Lam, V. and Wüthrich, C. (2015). No categorial support for radical ontic structural realism, *British Journal for the Philosophy of Science* **66**(3): 605–634.

Landsman, K. (2021). *Foundations of General Relativity*, Radboud University Press.

Landsman, N. P. (2005). Between classical and quantum, *Handbook of the Philosophy of Science*, pp. 417–553.

Laudan, L. and Leplin, J. (1991). Empirical Equivalence and Underdetermination, *Jorunal of Philosophy* **88**(9): 449–472.

Leinster, T. (2014). *Basic category theory*, Cambridge University Press.

Lewis, D. (1966). An Argument for the Identity Theory, *The Journal of Philosophy* **63**(1): 17–25.

Lewis, D. (1970). How to define theoretical terms, *Journal of Philosophy* **67**(13): 205–224.

Lewis, D. (1972). Psychophysical and theoretical identifications, *Australasian Journal of Philosophy* **50**(3): 249–258.

Malament (1977). Causal Theories of Time and the Conventionality of Simultaneity, *Noûs* **11**(3): 293–300.

Malament (2009). Notes on Geometry and Spacetime.

Malament, D. B. (1986). Newtonian Gravity, Limits and the Geometry of Space, *in* R. G. Colodny (ed.), *From Quarks to Quasars: Philosophical Problems of Modern Physics.*

Malament, D. B. (2012). *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*, The University of Chicago Press.

Manzano, M. and Aranda, V. (2022). Many-Sorted Logic, *in* E. N. Zalta and U. Nodelman (eds), *The Stanford Encyclopedia of Philosophy*, Winter 2022 edn, Metaphysics

Research Lab, Stanford University.

Nagel, E. (1961). *The Structure of Science*, Harcourt, Brace and world, New York.

North, J. (2009). The "structure" of physics: A case study, *Journal of Philosophy* **106**(2): 57–88.

Norton, J. D. (2014). A material dissolution of the problem of induction, *Synthese* **191**(4): 671–690.

Parfionov, G. N. and Zapatrin, R. R. (1995). Pointless spaces in general relativity, *International Journal of Theoretical Physics* **34**(5): 717–731.

Pitowsky, I. (1984). Unified Field Theory and the Conventionality of Geometry, *Philosophy of Science* **51**(4): 685–689.

Pletyukhov, V. A. (2018). Newton's second law and the concept of relativistic mass. **URL:** *https://arxiv.org/abs/1803.07909*

Poincaré, H. (1929). *Science and hypothesis*, The Walter Scott Publishing Co. LTD.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*, Routledge.

Putnam, H. (1960). Minds and Machines, *in* S. Hook (ed.), *Dimensions of Minds*.

Putnam, H. (1974). The Refutation of Conventionalism, *Nous* **8**(1): 25–40.

Quine, W. V. (1960). *Word and Object*, MIT University Press.

Redhead, L. M. (1993). The Conventionality of Simultaneity, *Philosophical Problems of the Internal and External Worlds: Essays on the Philosophy of Adolf Grünbaum*, University of Pittsburgh Press, chapter 5.

Reichenbach, H. (1958). *The Philosophy of Space & Time*, Dover.

Robb, A. (1914). *A Theory of Space and Time*, Cambridge University Press.

Roberts, B. W. (2015). Three merry roads to T-violation, *Studies in History and Philosophy of Science Part B* **52**: 8–15.

Roberts, B. W. (2021). *Reversing the Arrow of Time*, Cambridge University Press.

Rosenstock, S. (2019). *A Categorical Consideration of Physical Formalisms*, PhD thesis, UCI.

Rosenstock, S., Barrett, T. W. and Weatherall, J. O. (2015). On Einstein algebras and relativistic spacetimes, *Studies in History and Philosophy of Science Part B* **52**: 309–316.

Rovelli, C. (2019). Neither Presentism nor Eternalism, *Foundations of Physics* **49**(12): 1325–1335.

Rudin, W. (1987). *Real and Complex Analysis*, McGraw-Hill Book Company.

Rynasiewicz, R. (1992). Rings, Holes and Substantivalism: On the Program of Leibniz Algebras, *Philosophy of Science* **59**(4): 572–589.

Sarkar, S. and Stachel, J. (1999). Did Malament prove the non-conventionality of simultaneity in the special theory of relativity?, *Philosophy of Science* **66**(2): 208–220.

Savitttl, S. F. (2000). There's No Time like the Present (In Minkowski Spacetime), *Philosophy of Science* **67**.

Schlick, M. (1920). *Space and Time in Contemporary Physics*, Oxford University Press, New York. Translated from German.

Schwarzschild, K. (1916). On the Gravitational Field of a Mass Point according to Einstein's Theory, *General Relativity and Gravitation* **35**(5): 951–959.

Spirtes, P. L. (1981). *Conventionalism and the Philosophy of Henri Poincare*, PhD thesis, University of Pittsburgh.

Stein, H. (1991). On Relativity Theory and Openness of the Future, *Philosophy of Science* **58**(2): 147–167.

Thyssen, P. (2019). Conventionality and Reality, *Foundations of Physics* **49**(12): 1336–1354.

Tu, L. (2011). *An Introduction to Manifolds*, Springer.

Valente, M. B. (2016). The conventionality of simultaneity in Einstein's practical chrono-geometry, *Theoria (Spain)* **32**(2): 177–190.

Valente, M. B. (2018). The Conventionality of Simultaneity and Einstein's Conventionality of Geometry, *Kairos. Journal of Philosophy & Science* **20**(1): 159–180.

Vorms, M. (2011). Representing with imaginary models: Formats matter, *Studies in History and Philosophy of Science Part A* **42**(2): 287–295.

Vorms, M. (2012). Formats of Representation in Scientific Theorizing, *Models, Simultations, and Representations*.

Weingard, R. (1972). Relativity and the Reality of Past and Future Events, *The British Journal for the Philosophy of Science* **23**(2): 119–121.

Wesley, S., Earman, J., Glymour, C., Lennox, J., Machamer, P., McGuire, J., Norton, J. and Schaffner, K. (1992). Introduction to the Philosophy of Science, *Hackett Publishing Company*, chapter 5.

Worrall, J. (1989). Structural Realism: The Best of Both Worlds, *Dialectica* **43**(1/2): 99–124.

APPENDIX A

Limits

1. The Classical Limit of Minkowski Spacetime

To show that a definition of relativistic simultaneity converges to the Newtonian one, we need to develop some conceptual apparatus, after which the mathematics will be fairly simple. We first need a clear definition of Newtonian spacetime structure. Following Malament (1986, 2012), we take a classical spacetime to be a collection $\big(M, t_a, h^{ab}, \nabla_a\big)$ with M a four-dimensional, smooth, connected $C^\infty-$manifold such that $h^{ab}t_b = 0$ and $\nabla_a t_b = \nabla_a h^{bc} = 0$. The tensor fields t_a and h^{ab} play the roles of (degenerate) temporal and spatial metrics, respectively. As the field t_a is closed, it follows by the Poincaré lemma that t_a is locally exact. Since we are only interested in limits of Minkowski spacetime, we can assume that M is diffeomorphic to $\mathbb{R}^4$ and thus t_a is globally exact, i.e. there is a globally defined time function $t : M \to \mathbb{R}$ such that $t_a = \nabla_a t$. The level sets of the time function[1] will define what we take to be the *Newtonian hypersurfaces of simultaneity*, viewed as a set of points that all occur at the same moment in absolute time. A vector ξ in classical spacetime is spacelike if $t_a\xi^a = 0$ and timelike otherwise. The spacelike vectors turn out to be precisely the ones tangent to the Newtonian hypersurfaces of simultaneity (Malament 1986, pp. 183-184).

Our approach to the classical limit of a relativistic spacetime will follow Malament (1986), who begins with a one-parameter family of relativistic spacetimes $\big(\mathbb{R}^4, g_{ab}(\lambda)\big)$ for $\lambda \in (0, 1]$ such that the following two conditions are[2] satisfied:

[1]Different choices of time function will produce different 0-points for the time function, but the level sets will be the same and so will the time lapsed between them.

[2]Note that the limit is a function of the entire one-parameter family of spacetimes. As such, we are dealing with the classical limit of a family of spacetimes and therefore Minkowski spacetime can have different classical limits depending on the family of spacetimes one chooses. Nothing hinges on the particular choice of classical limit.

- There is some closed field t_a such that[3] $g_{ab}(\lambda) \to t_a t_b$ as λ tends to zero, and
- $\lambda g^{ab}(\lambda)$ converges to $-h^{ab}$ for some field h^{ab} of signature $(0, 1, 1, 1)$.

Our concern will be with Minkowski spacetime, where M is $\mathbb{R}^4$ and so we can use global Cartesian coordinates to illustrate the limit-procedure (this will amount to a specific choice of one-parameter set of spacetimes and thus to a choice of classical limit). Note first that the above two conditions can be explicitly satisfied by letting[4]:

$$g_{ab}(\lambda) = \text{diag}(1, -\lambda, -\lambda, -\lambda) \text{ and } g^{ab}(\lambda) = \text{diag}(1, -\frac{1}{\lambda}, -\frac{1}{\lambda}, -\frac{1}{\lambda})$$

Letting $t_a = (1, 0, 0, 0)$ and $h^{ab}(\lambda) = \text{diag}(0, 1, 1, 1)$, it follows immediately that $g_{ab}(\lambda) \to t_a t_b$ and $\lambda g^{ab}(\lambda) \to -h^{ab}$ as λ tends to zero. Note that h^{ab} has signature $(0, 1, 1, 1)$ and that t_a can be realised as the deriative of the scalar function $t : (t, x, y, z) \mapsto t$ as desired. Lastly, Malament (1986) shows that such a one-parameter family $\big(M, g_{ab}(\lambda)\big)$ will converge on a classical spacetime $\big(M, t_a, h^{ab}, \nabla_a)$ as λ tends to zero.

Our strategy for understanding the classical limit of a relativistic simultaneity relation is now as follows. At a given point, we will conceive of a relativistic simultaneity relation as determining a surface of events that are simultaneous with that point and which contains only spacelike or lightlike vectors. Our aim is to show that this surface converges towards a Newtonian hypersurface of simultaneity as λ tends to zero[5]. To do this, it will be expedient to characterise a hypersurface as the orthogonal complement under the Minkowski metric to a given vector field ξ, and then show that all the tangents of Newtonian Hypersurfaces of simultanety come closer and closer to the orthogonal complement of ξ as λ tends to zero (this is reasonable since in a classical spacetime, the subspace of spacelike vectors at any point is three-dimensional

[3]E.g. a $(2, 3)$—rank tensor $\Lambda^{ab}_{cde}(\lambda)$ converges to α^{ab}_{cde} if for all tensors β^{cde}_{ab} the one-parameter set of reals $\Lambda^{ab}_{cde}\beta^{cde}_{ab}$ converges to the real $\alpha^{ab}_{cde}\beta^{cde}_{ab}$. This procedure will also allow talk of convergence of covariant derivatives since every derivative is associated with a symmetric tensor field C^a_{bc} (see Malament (2012, Proposition 1.7.3)).

[4]Think of λ as $\frac{1}{c^2}$.

[5]Note that since the underlying manifold is $\mathbb{R}^4$ throughout, we can superpose Newtonian hypersurfaces of simultaneity onto Minkowski spacetime.

(Malament 1986, p. 183)). Note again that since the underlying manifold stays the same we can superpose a Newtonian structure onto Minkowski spacetime.

We have already seen what it means for a one-parameter family of spacetimes to have a classical limit, but before we can prove Proposition 1, we need to consider what it means for a relation to have a classical limit. Take $(\mathbb{R}^4, g_{ab}(\lambda))$ to have classical limit in the sense of Malament (1986). Take Sim to be binary on $(\mathbb{R}^4, g_{ab}(1))$ (Minkowski spacetime). If Sim foliates spacetime into spacelike hypersurfaces, then there is ξ timelike such that for any χ everywhere tangent to the foliation associated with Sim, we have $g_{ab}(1)\xi^a\chi^b = 0$. If Sim foliates spacetime into backwards (or forwards or entire) lightcones, there is ξ lightlike (just choose $\xi = \chi$) such that $g_{ab}(1)\xi^a\chi^b = 0$. To take the limit of Minkowski spacetime, we embedded it into the family $(\mathbb{R}^4, g_{ab}(\lambda))$ of spacetimes. Likewise, to take the limit of the relation Sim, we will embed Sim into a family of relations such that $Sim(\lambda)$ is a binary relation on the spacetime $(\mathbb{R}^4, g_{ab}(\lambda))$ for each λ. The question is: how do we "transport" Sim from Minkowski spacetime to $(\mathbb{R}^4, g_{ab}(\lambda))$ for $\lambda \neq 1$? In either case, we use ξ. To transpose Sim into a one-parameter family $Sim(\lambda)$ associated with each spacetime in $(\mathbb{R}^4, g_{ab}(\lambda))$, let $\chi(\lambda)$ be everywhere tangent to the foliation associated with $Sim(\lambda)$. Then we define $Sim(\lambda)$ by the requirement that $g_{ab}(\lambda)\xi^a\chi^b(\lambda) = 0$ for each λ. This means that we take Sim to be essentially characterised by ξ, and insist that as the metric changes with λ, the simultaneity surfaces associated with Sim will remain orthogonal to ξ. Last, we say that Sim has a classical limit if the following two requirements are fulfilled:

- We can choose a one-parameter family of vector field $\chi(\lambda)$ such that $g_{ab}(\lambda)\xi^a\chi^b(\lambda) = 0$ for all λ and such that it tends to $\chi(0)$ for some $\chi(0)$ as λ tends to zero, and
- $g_{ab}(\lambda)\chi^a(\lambda)$ tends to $t_a t_b \chi^a(0)$ as λ tends to zero.

It is reasonable to assume that any putative simultaneity relation has a classical limit since the Newtonian approximation must be recovered in the cases where all speeds are small compared to the speed of light. Specifically, the first condition amounts to the assertion that we can choose a tangent vector field for each $Sim(\lambda)$ in such a way that the family $\chi(\lambda)$ has a limit for λ tending to zero. The second assertion is a compatibility requirement between the convergence properties of the metric and of the tangent vector field. We choose the metrics $g_{ab}(\lambda)$ to have the limit $t_a t_b$, and the family of tangents $\chi(\lambda)$ to have some limit $\chi(0)$, but now we ask in addition that these two limits are compatible. That is, $g_{ab}(\lambda)\chi^a(\lambda)$ must converge on $t_a t_b \chi^a(0)$. Having assumed that there is *some* limit, we go on to prove that this limit is indeed the Newtonian foliation into simultaneity slices associated with the family of spacetimes $\big(\mathbb{R}^4, g_{ab}(\lambda)\big)$. We are now in position to prove Proposition 1:

Proposition 1. Given Minkowski spacetime, and given a putative simultaneity relation that has some classical limit with respect to a Newtonian spacetime, the following holds: if the simultaneity relation consists of a foliation into spatial hypersurfaces, then it converges in the classical limit to the Newtonian hypersurfaces of simultaneity. A similar conclusion holds if the simultaneity relation consists of a foliation into backwards (or forwards or indeed entire) lightcones.

Proof. let ξ and η be vector fields such that ξ is everywhere orthogonal to the simultaneity slices associated with our putative simultaneity relation *Sim* and let η be everywhere tangent to a Newtonian hypersurface of simultaneity, i.e. $t_a\eta^a = 0$. To get an immediate sense of why this is true, reason as follows. $g_{ab}(\lambda)$ tends to $t_a t_b$ as λ tends to zero implying that $g_{ab}(\lambda)\xi^a\eta^b$ will tend to $t_a t_b \xi^a \eta^b = t_a\xi^a t_b\eta^b$. But by assumption, $t_a\xi^a$ is some real number and $t_b\eta^b$ is zero so $g_{ab}(\lambda)\xi^a\eta^b$ will tend to $t_a t_b \xi^a\eta^b = t_a\xi^a t_b\eta^b = 0$ as desired. This argument shows that as the speed of light tends to infinity, the angle between ξ and *any* vector field that is everywhere tangent to a Newtonian hypersurface of simultaneity will tend to $90°$ since the inner product $g_{ab}(\lambda)\xi^a\eta^b$ tends to zero. But recall that our putative simultaneity relation was defined as the orthogonal complement to ξ, so in the limit the relation will converge on Newtonian simultaneity. The case where *Sim* foliates spacetime into spacelike hypersurfaces corresponds to ξ timelike, and the case where *Sim* foliates spacetime into backwards lightcones corresponds to $\xi = \eta$ tangent to a Newtonian hypersurface. By the above, $g_{ab}(\lambda)\eta^a\eta^b$ will tend to zero, so in the limit an arbitrary vector field everywhere tangent to a foliation into Newtonian hypersurfaces of simultaneity "becomes null" and hence tangent to the surface of the lightcone.

In the case where Sim foliates spacetime into spacelike surfaces, this can be done in more explicit terms. Let $\chi(\lambda)$ be everywhere tangent to the foliation associated with $Sim(\lambda)$ for each λ. We are assuming that Sim has a classical limit and the task is to show that this limit is a foliation into Newtonian simultaneity slices. By construction, $g_{ab}(\lambda)\xi^a\chi^b(\lambda) = 0$ for every λ and by the second point in the definition of what it means for Sim to have a classical limit, we get that $g_{ab}(\lambda)\xi^a\chi^b(\lambda)$ converges on $t_a t_b \xi^a \chi^b(0)$ as λ tends to zero. Combining these, we get $t_a t_b \xi^a \chi^b(0) = t_a \xi^a t_b \chi^b(0) = 0$. But ξ is $g_{ab}(1)$-timelike (and so, $g_{ab}(\lambda)$-timelike) and hence also Newtonian timelike. However, the latter means that $t_a \xi^a \neq 0$ so $t_b \chi^b(0)$ must be zero, which means that $\chi^b(0)$ is tangent to a foliation of Newtonian surfaces of simultaneity. This is what we wanted. $\square$

APPENDIX B

Affine Geometry

Malament (1977) proved a uniqueness theorem showing that given certain assumptions about the meaning of simultaneity, the Einstein (1905) simultaneity relation is the only one possible. However, some research such as Sarkar and Stachel (1999) has rejected his assumption of temporal symmetry in the form of a lack of time orientation, and therefore his conclusion too. In an unpublished set of notes, Malament (2009) sets out the basis for a response, by proving another uniqueness result that drops the assumption of temporal symmetry but answers a slightly different question echoing a proposal of Stein (1991):

> There is one slightly delicate point to be noted: Malament's discussion, which is concerned with certain views of Grünbaum, follows the latter in treating space-time without a distinguished time-orientation. To obtain Malament's conclusion for the (stronger) structure of space-time with a time-orientation, one has to strengthen somewhat the constraints he imposes on the relation of simultaneity; it suffices, for instance, to make that relation (as in the text above) relative to a state of motion (i.e., a time-like direction), rather than — as in Malament's paper — to an inertial observer (i.e., a time-like line). (Stein 1991, pp 153)

Since Stein's proposal, and thus Malament's response, prima facie involves changing the question from simultaneity relative to the world line of an observer to simultaneity with respect to an inertial frame, we must ask whether the original concern is ameliorated. I argue that it is possible to interpret Malament's response as answering the original question with the help of a simple corollary.

In the next section, we formulate and provide a proof of Malament's original 1977 result. Section 2 presents and proves Malament's second 2009 result and discusses whether it can be interpreted as an effective response to the critique of Sarkar and Stachel (1999). Section 3 offers a discussion of the physical relevance of Malament's overall strategy.

1. Malament's First Theorem

Both of Malament's uniqueness results are formulated in the framework of Minkowskian affine geometry, so let us first give the basic definitions.

1.1. Affine and Minkowski Geometry. Though the theory of finite dimensional vector spaces is native to any student of physics, vector spaces are not ideal for the modelling of the physical spacetime continuum. This is because vector spaces contain the algebraically priviledged element "0", whereas the target system contains no points of intrinsic priviledge. The solution for the working physicist is to only endow with physical meaning those statements that pertains to "vectors between points" and not those about the points themselves. The structure thus envisioned is in reality not a vector space but rather what is called "affine space". These affine spaces will be the topic of this section[1]. We start with the central definition,

Definition 4 (Affine Space). An *affine space* is an ordered touple $\big(A, V, +\big)$ where A is a non-empty set of points, V is a vector space and $+$ is a map from $A \times V$ to A such that

(AS1) For all $p, q \in A$ there is unique $v \in V$ such that $q = p + v$, and

(AS2) For all $p \in V$ and $v, u \in V$ we have $(p + v) + u = p + (v + u)$.

A couple of comments are in order. First, note how we equivocate between two different addition-operations: V is a vector space and therefore comes equipped with a binary operation called "+" taking two vectors to a third vector. On the other hand, the affine space comes with a binary operation taking a point and a vector to a point. Both operations are denoted by the symbol "+". In the formulation

[1] I follow the unpublished Malament (2009).

of requirement (II), this equivocation is explicit. Second, this structure really does formalise the intuitive picture layed out above. The set of points A contains the correlates of spacetime points whereas the vectors of V "points" from one point to another. This struture also succesfully gets rid of any priviledged points: even though both A and V will equal $\mathbb{R}^4$ in the following, $A = \mathbb{R}^4$ as sets whereas $V = \mathbb{R}^4$ as vector spaces. As 0 is priviledged *algebraically*, the underlying set $\mathbb{R}^4$ simply does not have the requisite structure to priviledge any particular point—in that sense, "0" is just a name. When $q = p + v$, we write $v = \vec{pq}$, and we take the *dimension* of the affine space $\big(A, V, +\big)$ to be the dimension of V. Hence, in what follows we will be working with four-dimensional affine space.

The following proposition lists certain central properties of affine spaces and we provide a proof to get a sense of how this new piece of machinery works,

Proposition 2 (Malament Proposition 2.2.1). Let $\big(A, V, +\big)$ be an affine space, and $p, q, r \in A$ be points. Then,
(i) $\vec{pp} = 0$ or, equivalently, $p + 0 = p$
(ii) if $\vec{pq} = 0$ then $p = q$
(iii) $\vec{qp} = -\vec{pq}$
(iv) $\vec{pq} + \vec{qr} = \vec{pr}$

Proof. (i) Let $p \in A$ be a point. By AS1 we have a unique vector $u \in V$ such that $p+u = p$. Using AS2 once and AS1 twice we get: $p+(u+u) = (p+u)+u = p+u = p$. Invoking the uniqueness clause in AS1, we have $u + u = u$ and thus $u = 0$ yielding $p + 0 = 0$. This is equivalent to the claim $\vec{pp} = 0$ since the vector $\vec{pp}$ *by definition* is the unique vector such that $p + \vec{pp} = p$, but that simply means that $\vec{pp} = u = 0$.
(ii) Let $p, q \in A$ be points and assume $\vec{pq} = 0$. By definition, this means that $q = p+\vec{pq} = p+0$, or equivalently that $q = p+0$. But from (i) we have that $p+0 = p$ so combining we get $q = p + 0 = p$ as desired.
(iii) Let p, q be points. By AS1 there is a unique vector v such that $q = p + v$ and a unique vector u such that $p = q + u$. Combining and using AS2 we get $q = (q + u) + v = q + (u + v)$. Since $q = q + 0$, the uniqueness clause in AS1 yields

$u + v = 0$. But by definition $u = \vec{qp}$ and $v = \vec{pq}$ so $\vec{qp} + \vec{pq} = 0$ or equivalently $\vec{qp} = -\vec{pq}$ as desired.
(iv) Let p, q, r be points. By AS1 there are unique vectors v, u, w such that $q = p + v$, $r = q + u$ and $r = p + w$. Combining and using AS2 yields $r = q + u = (p + v) + u = p + (v + u)$. But we already have that $r = p + w$, so by the uniqueness clause in AS1 we get $v + u = w$. By definition, $v = \vec{pq}$, $u = \vec{qr}$ and $w = \vec{pr}$ so $\vec{pq} + \vec{qr} = \vec{pr}$ as desired. □

Since Malament defines simultaneity relative to a timelike line, we need to prove a couple of results regarding lines in affine spaces. But first we need to define the notion of *affine subspace*,

Definition 5 (Affine Subspace). Let $(A, V, +)$ be an affine space. If $W \subset V$ is a subspace and $p \in A$ is a point, we define the *affine subspace of A through p determined by W* as $p + W := \{q \in A \mid \exists v \in W \textit{ such that}\, q = p + v\}$.

And we say that $p + W$ is a *line* just in case W is one-dimensional.

Proposition 3 (Malament Problem 2.2.1). Let $(A, V, +)$ be an affine space. For all points $p, q \in A$ and subspaces $W \subset V$, the following are equivalent
(i) q belongs to $p + W$
(ii) p belongs to $q + W$
(iii) $\vec{pq} \in W$
(iv) $p + W = q + W$ as an equality of sets
(v) $p + W \cap q + W \neq \emptyset$

Proof. (i) → (ii): if q belongs to $p + W$, there must be some vector $w \in W$ such that $q = p + w$. As W is a subspace it is closed under negation and so $-w \in W$. So $p = q - w$ is in $q + W$ as desired.
(ii) → (iii): if p belongs to $q + W$, there must be some vector $winW$ such that $p = q + w$. But $-w \in W$ so $q = p - w$ is in $p + W$ as desired.
(iii) → (iv): let $w \in W$. By definition, $q = p + \vec{pq}$ so $p + w = q - \vec{pq} + w$. But then if $\vec{pq} \in W$, we have $\vec{pq} + w \in W$ so $p + w$ is in $q + W$. For the converse, let $w \in W$

and consider $q + w$. By definition, we get $q + w = p + \vec{pq} + w$ so if $\vec{pq} \in W$, $q + w$ is in $p + W$ as desired.

(iv) $\to$ (v): since $0 \in W$, we have $p \in p + W$ and $q \in q + W$. As none of the affine spaces are empty, the implication is immediate.

(v) $\to$ (i): for some $w, w' \in W$, we have $p + w = q + w'$. But then $q = p + (w - w')$ so as W is a subspace, $q \in p + W$ as desired. ☐

We state and prove one more result about affine subspaces,

Proposition 4 (Malament Problem 2.2.2). Let $\big(A, V, +\big)$ be an affine space. If $p, q \in A$ are points and $p + W$ and $q + U$ are lines and $w \in W$ and $u \in U$ both non-zero, then $p + W$ and $q + U$ intersect if and only if $\vec{pq}$ is a linear combination of u and w.

Proof. If $p + W \cap q + U \neq \emptyset$, there must be $u' \in U$ and $w' \in W$ such that $p + w' = q + u'$. By definition, we have $q = p + \vec{pq}$ so $p + w' = p + (\vec{pq} + u')$. By the uniqueness clause of AS1, $w' = \vec{pq} + u'$ or equivalently $\vec{pq} = w' - u'$. But W and U are one-dimensional so there are numbers $\alpha, \beta \in \mathbb{R}$ such that $w' = \alpha w$ and $u' = \beta u$. Hence, $\vec{pq} = w' - u' = \alpha w - \beta u$ as desired.

Conversely, if $\alpha, \beta \in \mathbb{R}$ and $\vec{pq} = \alpha w + \beta u$ we get, by definition, $q = p + \alpha w + \beta u$ or equivalently $q - \beta u = p + \alpha w$ as desired. ☐

We will routinely need the fact that two points in an affine space defines a unique line,

Definition 6 (Line through two points). Given two points $p, q \in A$, we define *the line through p and q* by $L(p, q) := \{p + t\vec{pq} \mid t \in \mathbb{R}\}$.

For the line to be well-defined, we need to show that it is the only line through p and q,

Proposition 5 (Unique line). $L(p, q)$ is the only line containing p and q.

Proof. Clearly the line we have defined as $L(p, q)$ contains p and q so assume that $r + V$ is any line containing the p and q. We show that $L(p, q) = r + V$. By assumption, the lines $r + V$ and $L(p, q)$ intersect so by Proposition 4 there must be numbers $\alpha, \beta \in \mathbb{R}$ such that $\vec{pr} = \alpha\vec{pq} + \beta v$. Hence, $r = p + \alpha\vec{pq} + \beta v$. But since $p \in r + V$, there must be $v_0 \in V$ such that $p = r + v_0$. Combining we get $r = r + v_0 + \alpha\vec{pq} + \beta v$ so $v_0 + \alpha\vec{pq} + \beta v = 0$, or eqivalently, $v_0 + \beta v = -\alpha\vec{pq}$. But then $\vec{pq} \in V$ so $L(p, q) = p + V$. We can now restate the basic assumption as $r+V \cap p+V \neq \emptyset$. But then Proposition 3 (iv) immediately implies that $L(p, q) = r+V$ as desired. □

So far, we have described affine space, but for the purposes of special relativity, we will need more than flat affine space. Specifically, we need affine space equipped with a Lorentz-signature, generalised inner product. Technically, one simply equips the underlying vector space with a generalised inner product,

Definition 7 (Generalised Inner Product). A *generalised inner product* $\langle , \rangle$ on a vector space V is a map taking pairs of vectors to a single vector, satisfying

(IP1) For all $u, v \in V$, $\langle u, v\rangle = \langle v, u\rangle$

(IP2) For all $u, v, w \in V$, $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$

(IP3) For all $r \in \mathbb{R}$ and $u, v \in V$, $\langle u, rv\rangle = r\langle u, v\rangle$

(IP4) For all non-zero $u \in V$, there is a $v \in V$ such that $\langle u, v\rangle \neq 0$.

The reason this is a "generalised" inner product is the lack of requirement that $\langle v, v\rangle > 0$ whenever $u \neq 0$. We call the collection of an affine space together with an inner product (we drop the qualifier "generalised") a *metric affine space*. Generalised inner products are classified by *signature*:

Definition 8 (Signature). Let V be a vector space and let $\langle , \rangle$ be an inner product on V. The *signature* of $\langle , \rangle$ is an ordered pair of integers (n^+, n^-) where n^+ is the maximal possible dimension of a subspace U of V, where for all $u \in U$ $\langle u, u\rangle > 0$ and n^- is the maximal possible dimension of a subspace U of V where for all $u \in U$ $\langle u, u\rangle < 0$.

And for reference we state the following proposition without proof:

PROPOSITION 6 (Uniqueness of Metric Affine Spaces). Two metric affine spaces are isomorphic if and only if they have the same signature and dimension.

In what follows we will be interested in metric affine spaces with the signature $(1, n-1)$ for $n \geq 2$. This particular signature is called *Lorentz signature*. For the purposes of special relativity, the relevant dimension is four, but there is no reason to specialise the analysis to this case. In what follows, fix a metric affine space $\big(A, V, \langle , \rangle, +\big)$ of Lorentz signature with a dimension of at least two. The inner product generates a classification of vectors:

DEFINITION 9 (Classification of Vectors). Let $v \in V$ be a vector. We say that v is,
timelike if $\langle v, v\rangle > 0$
null or *lightlike* if $\langle v, v\rangle = 0$
spacelike if $\langle v, v\rangle < 0$ and
causal if v is either null or timelike.

We now turn to Malament's theorems.

1.2. Malament's First Theorem. Let $\big(\mathbf{A}, \langle\ ,\ \rangle\big)$ be an n-dimensional $(n \geq 2)$ Minkowskian space with underlying vector space V and point set A. Before turning to Malament's result, we review the standard relation of simultaneity relative to an inertial observer and make some observations that will be of use later. To this end, let L be a timelike line in A (think of L as the world-line of an inertial observer) and let us follow Malament in denoting the standard simultaneity-relation relative to L by Sim_L such that $(p, q) \in Sim_L$ iff $\overrightarrow{pq} \perp L$, where orthogonality is defined with respect to the Minkowskian inner product $\langle\ ,\ \rangle$. Further, we follow Malament in denoting the unique line containing the points $p, q \in A$ by $L(p, q)$. Sections 1 and 2 present the contents of Propositions 3.4.1 and 3.4.2 from[2] Malament 2009. Lemma 1 and proposition 7 figure in Malament 2009 as exercises for the reader, corollary 1 is mine and I have divided the proofs of Malament's two main theorems into a number

[2]In presenting Malament's results, I have taken the liberty to make minor changes when I felt it made the material more accessible.

of Lemmas—again to ease the acquisition. Let $f : A \to L$ be the function taking $p \in A$ to the (unique) point $q \in L$ such that $(p, q) \in Sim_L$. That this function is well-defined is the content of a lemma:

LEMMA 1. The function $f : A \to L$ taking points in A to their orthogonal projection onto L is well-defined.

PROOF. For $o \in A$ we need to show, 1) that there is a point $r \in L$ such that $(o, r) \in Sim_L$ and 2) that this point is unique. We start with existence:

Existence: Let $o \in A$ be arbitrary and let $L(p, q)$ be a timelike line in A. If $x \in L$, then $\overrightarrow{ox} = \overrightarrow{op} + \overrightarrow{px} = \overrightarrow{op} + a\overrightarrow{pq}$ for some $a \in \mathbb{R}$. But then $\langle \overrightarrow{ox}, \overrightarrow{pq} \rangle = \langle \overrightarrow{op}, \overrightarrow{pq} \rangle + a \langle \overrightarrow{pq}, \overrightarrow{pq} \rangle$. Since L is timelike, we have $\langle \overrightarrow{pq}, \overrightarrow{pq} \rangle > 0$ so the expression $g(a) = \langle \overrightarrow{op}, \overrightarrow{pq} \rangle + a \langle \overrightarrow{pq}, \overrightarrow{pq} \rangle$ defines a polynomial of degree 1. Hence, g has a root by the fundamental theorem of algebra[3]. Choosing the value of a such that $g(a) = 0$, we can go back and define a vector $\overrightarrow{ox} = \overrightarrow{op} + a\overrightarrow{pq}$, which will be orthogonal to L. Since $x \in L$, this establishes existence.

Uniqueness: Let $o \in A$ and assume that there are points $r, s \in L$ such that $\overrightarrow{or} \perp L$ and $\overrightarrow{os} \perp L$. Since $\overrightarrow{or} + \overrightarrow{rs} = \overrightarrow{os}$, it follows that $\overrightarrow{or} + \overrightarrow{rs} \perp L$, but then $\langle \overrightarrow{or}, \overrightarrow{rs} \rangle + \langle \overrightarrow{rs}, \overrightarrow{rs} \rangle = 0$. By hypothesis $\langle \overrightarrow{or}, \overrightarrow{rs} \rangle = 0$, so it follows that $\langle \overrightarrow{rs}, \overrightarrow{rs} \rangle = 0$. But then $r = s$ by proposition 2.2.1 in Malament (2009), which establishes uniqueness. $\square$

Note the reasonableness of this result: any point in space-time is simultaneous with some point on the world-line of our inertial observer—a fair candidate for a necessary condition for being appropriately interpreted as a relation representing the simultaneity judgements of that particular observer. It is equally clear that Sim_L is an equivalence relation: reflexivity and symmetry are obvious, and for transitivity let $L = L(a, b)$ and assume $(p, q) \in Sim_L$ and $(q, r) \in Sim_L$. Since $\overrightarrow{pr} = \overrightarrow{pq} + \overrightarrow{qr}$, we have: $\langle \overrightarrow{pr}, \overrightarrow{ab} \rangle = \langle \overrightarrow{pq}, \overrightarrow{ab} \rangle + \langle \overrightarrow{qr}, \overrightarrow{ab} \rangle = 0 + 0 = 0$ so $(p, r) \in Sim_L$. We will return to the question of pre-theoretical intuitions about the meaning of "simultaneity" in section 3 but at least prima facie it seems reasonable that any relation interpreted as simultaneity-for-L should be an equivalence relation. Before we will consider the

[3]Note that this does not establish uniqueness since a could depend on the choice of p and q.

possibility of other simultaneity-relations, we turn to the interplay between Sim_L and the symmetries of $\big(\mathbf{A}, \langle\ ,\ \rangle\big)$. Formally, we will take a symmetry of $\big(\mathbf{A}, \langle\ ,\ \rangle\big)$ to be an isometry[4] $\phi : A \to A$ and say that an L-isometry is an isometry that also preserves L, in that $\phi[L] = L$. The point is that Sim_L is preserved under every L—isometry in the sense that $(p, q) \in Sim_L$ iff $(\phi(p), \phi(q)) \in Sim_L$. (Malament 2009, p. 59) classifies all L—isometries but for our purposes, we only need to pay attention to the case where ϕ is a so-called temporal reflection with respect to a hyperplane orthogonal to L. For future reference, we write down the action of ϕ explicitly: If $o \in L$ and $p \in A$ proposition 3.1.1 in (Malament 2009) yields unique vectors $\overrightarrow{v} \parallel L$ and $\overrightarrow{w} \perp L$ such that $p = o + \overrightarrow{v} + \overrightarrow{w}$. Now, $\phi(p) = o - \overrightarrow{v} + \overrightarrow{w}$. That Sim_L is preserved under temporal reflections of this type is the content of the next proposition:

Proposition 7. Sim_L is preserved under temporal reflections with respect to hyperplanes orthogonal to L.

Proof. We need to show that $(p, q) \in Sim_L$ iff $\big(\phi(p), \phi(q)\big) \in Sim_L$ whenever ϕ is a temporal reflection with respect to a hyperplane orthogonal to $L = L(a, b)$.

Claim: Let $o \in L$ and $p, q \in A$. $\overrightarrow{pq} \perp L$ iff there are vectors v, w, w' with $v \parallel L$ and $w, w' \perp L$ such that $p = o + v + w$ and $q = o + v + w'$. **Proof of Claim:** From proposition 3.1.1 in (Malament 2009) there are vectors $v, v', w,'$ with $v, v' \parallel L$ and $w, w' \perp L$ such that $p = o + v + w$ and $q = o + v' + w'$. It remains to show that $v = v'$. We have that $\overrightarrow{pq} = \overrightarrow{po} + \overrightarrow{oq} = -(\overrightarrow{v} + \overrightarrow{w}) + (\overrightarrow{v'} + \overrightarrow{w'}) = (\overrightarrow{v'} - \overrightarrow{v}) + (\overrightarrow{w'} - \overrightarrow{w})$ and thus $\langle \overrightarrow{pq}, \overrightarrow{ab} \rangle = \langle \overrightarrow{v'} - \overrightarrow{v}, \overrightarrow{ab} \rangle + \langle \overrightarrow{w'} - \overrightarrow{w}, \overrightarrow{ab} \rangle$ but the latter equals $\langle \overrightarrow{v'} - \overrightarrow{v}, \overrightarrow{ab} \rangle$ by hypothesis. This means that $\langle \overrightarrow{pq}, \overrightarrow{ab} \rangle = 0$ iff $\langle \overrightarrow{v'} - \overrightarrow{v}, \overrightarrow{ab} \rangle = 0$. But by proposition 3.1.5 in (Malament 2009) (the"wrong way schwarz inequality"), this holds iff $\overrightarrow{v'} - \overrightarrow{v} = \overrightarrow{0}$ or equivalently $\overrightarrow{v'} = \overrightarrow{v}$. This proves **claim**.

Now, recall that the action of ϕ is $o + \overrightarrow{v} + \overrightarrow{w} \mapsto o - \overrightarrow{v} + \overrightarrow{w}$ so we can argue as follows: $\overrightarrow{pq} \perp L$ iff $\overrightarrow{v} = \overrightarrow{v'}$ by **claim.** But the latter is obviously equivalent to $-\overrightarrow{v} = -\overrightarrow{v'}$, which in turn holds iff $\overrightarrow{\phi(p)\phi(q)} \perp L$ again by **claim**, so $(p, q) \in Sim_L$ iff $(\phi(p), \phi(q)) \in Sim_L$ as desired. □

[4]A Minkowski space-time isometry is an isometry of affine spaces satisfying the extra requirement that the Minkowski inner product is preserved. See Malament (2009, pp 13, 59).

In summary, the standard simultaneity-relation Sim_L is an equivalence relation, satisfies the condition that for all $p \in A\ \exists! q \in L$ such that $(p, q) \in Sim_L$ and is preserved under all $L - isometries$. The question is whether any appropriate simultaneity-relation *ought* to satisfy these criteria. For now, we follow Malament in simply going with the assumption that any adequate simultaneity-relation $S_L \subseteq A \times A$ will[5]. meet the following list of requirements (relative to a fixed timelike line L):

- **S1:** S_L is an equivalence relation.
- **S2:** For all $p \in A\ \exists! q \in L$ such that $(p, q) \in S_L$.
- **S3:** S_L is preserved under all L—isometries and in particular under temporal reflections with respect to hyperplanes orthogonal to L.

The goal is now to prove that $S_L = Sim_L$. The following facts about Sim_L shall be convenient to refer to in proofs:

- **(i):** For all $p \in A$, $(p, f(p)) \in Sim_L$, and
- **(ii):** For all $p, p' \in A$, $(p, p') \in Sim_L$ iff $f(p) = f(p')$.

Note that the result follows immediately if we could show that S_L had to satisfy **(ii)**—this is indeed the case, but to prove it, we need to first show that S_L satisfies **(i)**. This is the content of lemma 2:

LEMMA 2. Let L be a timelike line in A, and let S_L be a two-place relation on A that satisfies $S1$ and $S2$, and is L—invariant. It follows that for all $p \in A$, $\big(p, f(p)\big) \in S_L$.

PROOF. Let L and $S \subseteq A$ be as above and pick $p \in A$ arbitrarily. By $S2$ there exists unique $q \in L$ such that $(p, q) \in S_L$—we show that $q = f(p)$. Let $\phi : A \to A$ be the temporal reflection with respect to the hyperplane orthogonal to L through p and $f(p)$. We have already seen that ϕ is an L-isometry and since $p = f(p) + \overrightarrow{f(p)p}$ and $q = f(p) + \overrightarrow{f(p)q}$, we have $\phi : p \mapsto p$ and $\phi : q \mapsto f(p) - \overrightarrow{f(p)q} = f(p) + \overrightarrow{qf(p)}$. It follows from L-invariance that $(\phi(p), \phi(q)) = (p, f(p) + \overrightarrow{qf(p)}) \in S_L$ but since $f(p)$

[5] Malament (2009) drops the subscript, but I prefer to include it to emphasise that simultaneity is in relation to a timelike line or, more colloquially, to an inertial observer.

and q are both on L it follows that $f(p)+\overrightarrow{qf(p)} \in L$ so that by the uniqueness clause in $S2$ we get $\overrightarrow{qf(p)} = \overrightarrow{0}$ or equivalently that $f(p) = q$ as desired. □

We are now in a position to show that S_L satisfies **(ii)** above:

LEMMA 3. Let L be a timelike line in A, and let S_L be a two-place relation on A that satisfies $S1$ and $S2$, and is L—invariant. Then the following holds: for all $p, p' \in A$, $(p, p') \in S_L$ iff $f(p) = f(p')$.

PROOF. Let $(p, p') \in S_L$. By lemma 2 $(p, f(p)), (p', f(p')) \in S_L$. But S_L is symmetric and transitive so $(f(p), f(p')) \in S_L$ so that $f(p) = f(p')$ by the uniqueness clause in $S2$. Conversely, if $f(p) = f(p')$, then $(p, f(p')) \in S_L$ by lemma 2 and since $(p', f(p')) \in S_L$, we get $(p, p') \in S_L$ by symmetry and transitivity of S_L. □

We are now ready to prove the main result and as promised the proof is immediate:

THEOREM 5. Let L be a timelike line in A, and let S_L be a two-place relation on A that satisfies $S1$ and $S2$, and is L—invariant. Then $S_L = Sim_L$.

PROOF. By lemma 3 $(p, p') \in S_L$ iff $f(p) = f(p')$, but the latter holds iff $(p, p') \in Sim_L$. □

2. Changing the Question? Malament's Second Theorem

As we saw in section 1 the crux of Sarkar and Stachel (1999)'s critique of Malament (1977) is the dependence in the proof of theorem 5 on the invariance of S_L under temporal reflections. Clearly, including a temporal orientation as part of the fundamental geometry shrinks the class of spacetime symmetries since temporal reflections will no longer count (and one can check that no other symmetries are suddenly introduced by this move). This in turn means that the notion of invariance under the symmetry-group is weakened accordingly to the point where a uniqueness result can no longer be established[6]. This raises the question of whether any set

[6]See Malament (2009, p. 61) for an example of a symmetry-invariant simultaneity relation satisfying **S1** and **S2** which nevertheless differs from Sim_L. (Sarkar and Stachel 1999) also give an example, but the framework differs from the present.

of plausible extra assumptions suffices to re-establish uniqueness of the simultaneity relation and Malament (2009) shows one way in which this can be done. First, let an L-isometry $\phi : A \to A$ be an $(L, \uparrow)$—isometry if ϕ also preserves temporal orientation, i.e. if for all timelike vectors $\overrightarrow{pq}$ we have $\langle \overrightarrow{pq}, \overrightarrow{\phi(p)\phi(q)} \rangle > 0$. It follows directly that the temporal reflections playing a crucial role in the proof of lemma 2 are not $(L, \uparrow)$—isometries. The idea now is to remedy this weakening of the symmetry-group by defining simultaneity not relative to a single observer but relative to a set of co-moving observers. To this end, we need the definition of a *frame*:

Definition 10 (Frame). A *frame* $\mathcal{L}$ is a set of parallel timelike lines in A which is maximal in the sense that for all $p \in A \; \exists! L \in \mathcal{L}$ such that $p \in L$.

Importantly, whenever $L, L' \in \mathcal{L}$, we have $Sim_L = Sim_{L'}$ since $L \parallel L'$. This means that any frame $\mathcal{L}$ gives rise to a unique standard relation of simultaneity $Sim_{\mathcal{L}}$. Now, define an $(\mathcal{L}, \uparrow)$—isometry to be an isometry $\phi : A \to A$ that preserves temporal orientation and such that $\phi[L] \in \mathcal{L}$ whenever $L \in \mathcal{L}$. The goal is now to prove that if $S_{\mathcal{L}} \subseteq A \times A$[7] is an equivalence relation, satisfies **S2** for some $L \in \mathcal{L}$ and is invariant under the group of $(\mathcal{L}, \uparrow)$—isometries then $S_{\mathcal{L}} = Sim_{\mathcal{L}}$. Analogously to the proof of theorem 5, we do not actually require invariance under all possible $(\mathcal{L}, \uparrow)$—isometries; here we are going to make use of two types, namely translations by an arbitrary vector and reflections in L of a two-plane containing L and some other point $p \in A$. For future reference, we write down their actions explicitly:

- Let $p \in A$ and let $\phi : A \to A$ be a reflection in L of the two-plane containing L and p. Then $\phi(p) = p + 2\overrightarrow{pf(p)} = f(p) + \overrightarrow{pf(p)}$.
- Let $p \in A$ and let $\phi : A \to A$ be translation by $\overrightarrow{v}$. Then $\phi(p) = p + \overrightarrow{v}$.

It will be convenient to argue on Sim_L where $L \in \mathcal{L}$ is such that L satisfies **S2**, but keep in mind that $Sim_L = Sim_{\mathcal{L}}$ by the comment above. As in the proof of theorem 5, the strategy is to show that $S_{\mathcal{L}}$ satisfies **(i)** and **(ii)** from section 1.

Lemma 4. Let $\mathcal{L}$ be a frame and $S_{\mathcal{L}}$ a two-place relation satisfying **S1** and, for some $L \in \mathcal{L}$ satisfies **S2**. Further, suppose $S_{\mathcal{L}}$ is $(\mathcal{L}, \uparrow)$—invariant (in particular

[7]Again, I prefer to keep the subscript to indicate that simultaneity is defined relative to a frame.

invariant under the actions of translations and reflections in some $L \in \mathcal{L}$ of the two-plane containing L and some other (arbitrary) point $p \in A$). Then $\forall p \in A$ we have $(p, f(p)) \in S_{\mathcal{L}}$.

Proof. Take $p \in A$ arbitrary and let L be the line with respect to which $Sim_{\mathcal{L}}$ satisfies **S2**. Then there is unique $q \in L$ such that $(p, q) \in S_{\mathcal{L}}$. Now, if $p \in L$, the conclusion follows directly from the uniqueness clause in **S2** since in that case $f(p) = p$ so $(p, p) = (p, f(p))$ and $(p, p) \in S_{\mathcal{L}}$ by reflexivity. So assume $p \in A \setminus L$ and let the action of $\phi_1 : A \to A$ be reflection in L. Then $\phi_1(p) = f(p) + \overrightarrow{pf(p)}$ and $\phi_1(q) = q$—the latter since $q \in L$. Further, let the action of $\phi_2 : A \to A$ be translation by the vector $\overrightarrow{qp}$. Since the composition of isometries is an isometry, we get $\big((\phi_2 \circ \phi_1)(p), (\phi_2 \circ \phi_1)(q)\big) \in S_{\mathcal{L}}$. But $(\phi_2 \circ \phi_1)(p) = \phi_2(f(p) + \overrightarrow{pf(p)}) = f(p) + \overrightarrow{pf(p)} + \overrightarrow{qp} = f(p) + \overrightarrow{qf(p)}$ and $(\phi_2 \circ \phi_1)(q) = \phi_2(q) = q + \overrightarrow{qp} = p$ so that $\big(f(p) + \overrightarrow{qf(p)}, p\big) \in S_{\mathcal{L}}$. But note that both $f(p)$ and q are on L so that $f(p) + \overrightarrow{qf(p)} \in L$ and thus $f(p) + \overrightarrow{qf(p)} = q$ by the uniqueness clause in **S2**. But the latter implies that $\overrightarrow{qf(p)} = \overrightarrow{f(p)q}$, which in turn implies $f(p) = q$ by proposition 2.2.1 in Malament (2009). Hence $(p, q) = (p, f(p)) \in S_{\mathcal{L}}$ as desired. □

Having established that $S_{\mathcal{L}}$ satisfies **(i)**, we just need to show that $S_{\mathcal{L}}$ also satisfies **(ii)**—this is the content of the next lemma:

Lemma 5. Let $\mathcal{L}$ be a frame and $S_{\mathcal{L}}$ a two-place relation satisfying **S1** and, for some $L \in \mathcal{L}$ satisfies **S2**. Further, suppose $S_{\mathcal{L}}$ is $(\mathcal{L}, \uparrow)$—invariant (in particular invariant under the actions of translations and reflections in some $L \in \mathcal{L}$ of the two-plane containing L and some other (arbitrary) point $p \in A$). Then for any $p, p' \in A$, we have $(p, p') \in S_{\mathcal{L}}$ iff $f(p) = f(p')$.

Proof. Exactly as the proof of lemma 3 with references to lemma 4 instead of lemma 2. □

Now the uniqueness result follows immediately:

Theorem 6. Let $\mathcal{L}$ be a frame and $S_{\mathcal{L}}$ a two-place relation satisfying **S1** and for some $L \in \mathcal{L}$ satisfies **S2**. Further, suppose $S_{\mathcal{L}}$ is $(\mathcal{L}, \uparrow)$—invariant (in particular invariant under the actions of translations and reflections in some $L \in \mathcal{L}$ of the two-plane containing L and some other (arbitrary) point $p \in A$). Then $S_{\mathcal{L}} = Sim_{\mathcal{L}}$.

PROOF. Exactly as the proof of theorem 5 with references to lemma 5 instead of lemma 3. □

The upshot of theorem 6 for our purposes is that it establishes a uniqueness result without recourse to time reversal invariance. But the cost of this result is a change in setup from one inertial observer to an entire set of co-moving observers raising the question of whether such a move is physically warranted. Interestingly though, theorem 6 allows for another interpretation: instead of following Malament in postulating that simultaneity is defined relative to a congruence of co-moving observers and deriving the resulting symmetry group, we can reverse the order and postulate that simultaneity be invariant under translations and reflections of the type featuring in the proof of theorem 6 directly. Mathematically, this changes nothing, but I maintain that this move has a philosophical advantage—invariance of simultaneity under translations and reflections can plausibly be construed as encoding an assumption of homogeneity of the spacetime. Invariance under translations simply means that the simultaneity of two events depends only on the relative location of the events (as opposed to dependence on absolute location). Invariance under reflections of the type discussed above means that, from the perspective of an inertial observer, spacetime has no preferred direction. We formulate this insight in a corollary:

COROLLARY 1. Let L be a time-like line and S_L a two-place relation satisfying **S1** and **S2**. Further, suppose S_L preserves temporal orientation and is invariant under the actions of translations and reflections in L of the two-plane containing L and some other (arbitrary) point $p \in A$. Then $S_L = Sim_L$.

PROOF. As the proof of theorem 6. □

Note that since corollary 1 concludes that $S_L = Sim_L$, the conclusion is that the resulting simultaneity-relation is really defined on the level of congruences of co-moving observers—the upshot is that this is now a consequence and not a postulate.

Formally, in only assuming invariance under translations and reflections, we are postulating a little less than Malament who assumes from the outset that the relation is defined on a frame and is $(\mathcal{L}, \uparrow)$—invariant. Malament's own proof shows that our weaker assumption suffices.

3. On the Physical Significance of Malament's Theorems and the Re-Emergence of Conventionality

Let us first take stock. Invariance under temporal reflection played a key role in the proof of theorem 5 and simultaneity was defined relative to one inertial observer. In theorem 6, simultaneity was defined relative to a maximal set of co-moving observers and the proof crucially relied on invariance under two types of action: translation by a fixed vector and reflections in some $L \in \mathcal{L}$ of the two-plane containing L and some other arbitrary point $p \in A$. The situation is further complicated by the fact that reliance on invariance under these two actions can be interchanged for invariance under isometries of a different type: namely $\mathcal{L}$—isometries ϕ such that for some line $L \in \mathcal{L}$ we have $\phi_{|L} = Id_L$ (the purpose of the last clause is to preserve the temporal orientation)[8] (Malament 2009, pp 61). Anyone who wishes to draw the conclusion that simultaneity is not conventional is tasked with justifying the assumptions underwriting the truth of theorem 6 — i.e. it is not enough to simply undermine one proof. Defining simultaneity with respect to a frame rather than a single timelike line only makes sense under the assumption that co-moving observers ought to agree on all simultaneity judgements. However, this is a non-trivial assumption. Belot makes the observation that näively, one should expect the simultaneity-relation to be an equivalence-relation with three-dimensional, space-like and connected equivalence classes but that the Minkowski spacetime itself does not allow for any such non-trivial relations to be defined[9] (Belot 2010, pp 396). (Giulini 2001) offers a potent algebraic framework for understanding the discussion of invariance of relations on spacetimes,

[8]Note that the L-reflections are of the latter type. Crucially, the class of $\mathcal{L}$—isometries such that for some $L \in \mathcal{L}$, $\phi_{|L} = Id_L$ is larger than the set of reflections and rotations around L. The (perhaps) surprising result is that the "difference" is made up for by translations by a fixed vector.

[9]"Non-trivial" in this context means "different from both the total relation and the diagonal relation".

the details of which I will not go into here, but the upshot for our purposes is this: the symmetry-group of Minkowski spacetime is too rich for any appropriate (non-trivial) relations to be invariant under it. Essentially, what Malament does when he provides, first, a single timelike line, and later, a frame, is to add additional structure to the raw Minkowski spacetime. The effect of adding additional structure is a shrinking of the group of symmetries to the point where a unique, non-trivial and, importantly, invariant relation is admitted (Giulini 2001, pp 657). This highlights a sense in which the concept of simultaneity can be said to be external to special relativity. The dialectic is at once similar to and different from what is described in Belot (2010) as a "symmetry argument". It is similar in the sense that we are given the structure of Minkowski spacetime and wish to extend it by adding a simultaneity relation. It is different in the sense that we know from the outset that no sensible simultaneity-relation is invariant under the full group of Minkowski-symmetries, and this is why we cannot define absolute simultaneity in Minkowski spacetime in the same way as we do in Newtonian physics[10]. Rather, we proceed by strategically shrinking the symmetry-group by introducing extra structure to the spacetime (e.g. a timelike line or a frame) and only then do we consider adding a simultaneity-relation invariant under the reduced symmetry-group. Once a would-be simultaneity-relation is obtained, the question arises of what the appropriate standard of judgement is. Belot writes:

> If, on the other hand, one can show that the proposed extension is invariant. [...] If the extension can be shown to be the unique invariant extension of the sort under consideration, then one has reason to accept the extended structure as a (more or less) adequate representation of the features under investigation. (Belot 2010, pp 395)

In our case, "the feature under investigation" is, of course, the notion of simultaneity. However, Belot offers an interesting caveat. He notes that the degree of

[10]For details see (Giulini 2001, pp 662).

confirmation conferred on the added structure upon learning that it "fits" the symmetries should be in proportion to how certain we are of the adequacy of the initial structure. In plain English: Don't be too impressed with the fact that some extra piece of structure fits snugly with the theory if the theory itself is bogus. In the context of simultaneity in STR, Belot goes on to write:

> This is quite typical: symmetry arguments are of little polemical value in situations where fundamental questions are at stake, since those are the cases in which there will be little agreement as to whether a given structure provides an acceptable point of departure for such an argument. (Belot 2010, pp 397)

And, of course, we find ourselves exactly in this situation. The comments above can be synthesised into this schematic frame for the discussion of simultaneity:

(1) First, we need to agree on the initial structure. This includes deciding whether temporal orientation ought to count as part of the geometrical framework.
(2) Now, the symmetry-group can be computed together with the class of equivalence relations satisfying **S2**. Probably, the initial symmetry-group is too rich to admit non-trivial invariant relations.
(3) Now comes the time to add further structural elements (a timelike line, a frame, etc.) with the effect that the symmetry-group is weakened. Preferably, the extra pieces can be given appropriate physical interpretations (a timelike line is a single observer, a frame is a host of co-moving observers, etc.). The aim is to weaken the symmetry-group to the point where a unique invariant relation is admitted. Probably, multiple ways of doing this are possible.
(4) The points $1-3$ should be cycled through iteratively so that changes can be made to the initial structure (point 1) and the extra structural elements (point 3) in light of the nature of the admitted relations.

Two considerations must be balanced: any bit of structure (Point 1 and Point 3) should be physically motivated while heed should also be paid to the physical reasonableness of the resulting simultaneity relation. The latter implies the re-emergence of an irreducible and more fundamental conventionality, for what does the word "simultaneity" mean when robbed of its näive, pre-relativistic sense? Surely, there is room here for genuine disagreement that prompts those hoping for a mathematical solution to the question of the status of relativistic simultaneity to adopt a more modest stance. Even if it is provably the case that only one binary relation exists satisfying **S1** and **S2** and being appropriately invariant, a more fundamental question remains: "Is this relation *simultaneity*"? To answer this, we need to go back to Newton where the semantics governing the term is arguably necessitated by the physical facts.