Representation Theorems for Geometry and Spacetime
J Butterfield for 28 January 2020

1. The Idea of a Representation Theorem

A representation theorem has the shape: all items satisfying some abstract description A are identical to, or isomorphic to, a concrete object O; with the identification or isomorphism often having a well-controlled freedom F, e.g. a uniqueness up to a certain transformation, T. Typically, F is a Lie group of transformations, and $T \in F$.

The general themes:
(A): The idea of A as more cognitively secure than O (because in a preferred language, and-or having a clearer operational language, and-or being self-evident or at least more self-evident). Then the idea is: the representation theorem vindicates the apparently cognitively insecure O.
(B) The theorem also shows that what seemed a generalization (going from O to A) is in fact not. And this is enlightening, even reassuring.
(C) Also the freedom shown in F, explains, i.e. explains away/makes one comfortable with, the redundancy or artefacts one can/might discern, and find suspicious, in O.
(D) Often we do NOT have a representation theorem in the above sense! That is: the items satisfying abstract description A form a large class of non-isomorphic objects, even of the same type as O. Then the aim/benefit is to study the variety of the Os, and how that variety encodes information about A. Think after all of representation theory: A the abstract group, each O a specific representation. E.g. a unitary representation on a complex vector space. The variety of the Os in this example is their falling into different equivalence classes under the equivalence relation of unitary equivalence.

2. Examples

1. *Hilbert's axiomatization of Euclidean geometry, 1899*: —
The main philosophical themes here are:

(i): the rise of formal methods, and interest in axiomatization and its metatheory, across allof pure mathematics, from 1850 to 1930+;

(ii): Hilbert's axioms are in the traditional Greek, i.e. synthetic, style, and the representation theorem can be regarded as vindicating analytic geometry, especially Cartesian coordinates.

(iii): continuity, which distinguishes $\mathbb{R}$ from the rationals, is a second order notion, i.e. can only be expressed by quantifying not only over objects (the usual variables: $x, y, z, ...$), but also over properties: saying e.g. $\forall P$.

(iv): the tradition of such theorems continued, e.g. work by Tarski's school in 1950s. Cf. Tarski (1959), 'What is elementary geometry?'

(v): the commitment to spatial points: cf. example 4 below.

A = Hilbert's axioms governing a domain of points, and a three-place relation of Betweenness $B(x, y, z)$ and a four-place relation of Congruence $C(w, x, y, z)$.[1]

$O = \mathbb{R}^2$, with its Euclidean metric.

F = the affine group of translations, rotations and (spatially constant) dilations acting on $O = \mathbb{R}^2$. These three kinds of element of F reflect the freedom (given a choice of Cartesian coordinates for the plane: such a choice being a bijection from the geometric plane to $\mathbb{R}^2$), respectively:

to shift origin, to rotate axes, and to change the unit of length.

(For example, the axioms for Congruence, and more generally the axioms, do not dictate a unit of length).

Thus for theme (C): the redundancy/artefacts in O is the choice of an origin and direction of axes, and length unit, for Cartesian coordinates in $\mathbb{R}^2$.

This style of axiomatization, and associated representation theorems, has been extended to the classical non-Euclidean geometries (elliptic, hyperbolic and spherical), and even to arbitrary Riemannian manifolds. (At least: so it is said. Cf. Belot (2011, 11 note 10, and 32, note 64, citing Mundy (1992, especially Section 3 and Appendix)).

2. *The Robb synthetic axiomatization of special relativity, 1914*:—
Arthur Robb (1914, 1921, 1936) did for the kinematics of special relativity, i.e. for Minkowski spacetime, what Hilbert did for Euclidean space. But using just a single 2-place predicate representing the relation of causal connectibility (by a signal at most as fast as light) among spacetime points.

Cf John Winnie's exposition (1977) for a detailed modern account, Mundy (1986) for a simplification; and e.g. Zeeman (1974) for an example from the mathematical literature of the same theme, viz. 'causality determines all the structure'.

The main philosophical themes here turn on the consequence that all of Minkowski geometry is determined by (in philosophers' jargon: supervenient upon; in logicians' jargon: implicitly defined by) the extension of (philosophers' jargon: facts about) the relation of causal connectibility . This consequence:

(i): revived the idea (in Reichenbach et al.) of a 'causal theory of time';

(ii): refuted the idea that distant simultaneity was not merely frame-dependent, but also merely conventional: more precisely: it refuted this idea provided that the facts of causal connectibility among spacetime points were not conventional.

A = Robb's axioms governing a domain of points, and just one single binary pred-

[1] Saying 'domain of points' is anachronistic: it follows a later tradition, but Hilbert and various formalizations that followed him had various geometric figures determined by a finite set of points, e.g. straight lines, line segments and polygons as values of individual variables. Cf. Tarski (1959, footnote 3).

icate $After(x, y)$ representing causal connectibility. *Synthetic* chronogeometry!

$0 =$ Minkowski spacetime, $Mink$: identified with $\mathbb{R}^4$ with the Minkowski metric.

$F =$ the Poincare group, augmented by (spacetime-constant) dilations, acting on $O = Mink$. As in Example 1: the various kinds of element of F—spacetime translations, spatial rotations, boosts and dilations—reflect the freedom, given a choice of inertial coordinates (a choice of Lorentz chart) for spacetime, respectively: to shift spacetime origin, to rotate spatial axes, to boost, and to change the unit of length.

This example is clearly a cousin of Hilbert's axiomatization of geometry.

Two broad kinds of philosophical doubt, about:

(i) the existence of abstract mathematical objects such as real numbers ('against Platonism') and

(ii) about the existence of spatial or spacetime points,

prompt two further programmes: the effort to write down physical theories like those usually cast in terms of numbers and spacetime:

(i): without reference to any abstract objects, especially real numbers as the values of physical fields (e.g. of a scalar gravitational potential): the *nominalism*, the 'mathematical atheism', of Field (1980): cf. Example 3.

(ii): without reference to spacetime or spatial points: *relationism*, as vs. *substantivalism*—in one sense out of many ... cf. Example 4.

3. *Field's synthetic axiomatization of Newtonian gravitation, 1980*:—
The 'pure mathematical atheist', Hartry Field (1980) proposes to 'vindicate' pure mathematics as literally false—there are no pure mathematical numbers, functions etc.!—though it is instrumentally useful in physics. (That pure mathematics is, not merely useful, but *indispensible*, to state physical theories is the core of the *indispensability argument* for mathematical platonism: it is advocated by e.g. Quine.).

Field aims to show that pure mathematics is a conservative extension, in logicians' sense, of our physical theories. So he is committed to formulating an adequate physics without any reference to numbers, functions etc. But he is happy to accept spatial and spacetime points (he is a 'substantivalist'), and relations among them.

His example is: Newtonian gravitational theory in a Newtonian/Galilean spacetime setting. This is *usually* formulated with: a gravitational potential ϕ which is a scalar (and so has values $\in \mathbb{R}$!); and a mass density ρ which is also a scalar: they are coupled by Poisson's equation (an equation of values $\in \mathbb{R}$!).

NB: for simplicity, I shall set aside the dynamics (the self-gravitation!) of ρ. This is usual, even in discussions of curved spacetime, aka: *Newton-Cartan* formulations of Newtonian gravitational theory. So just think of ρ as *given* across spacetime: and as determining the timelike curves of free-falling particles via the prescription that a particle's 4-acceleration is $-\nabla\phi$. (Newton-Cartan formulations trade in ϕ for a curved connection, in such a way that Poisson's equation becomes an analogue—a close analogue!—of Einstein's general relativistic field equations.)

So Field aims to state the 'same content' as these usual formulations—in 'synthetic form'.

So we expect him to use e.g. a three-place predicate representing Betweenness of values of ϕ amongst three spacetime points x, y, z: call it $B_\phi(x, y, z)$. Similarly, we expect a predicate representing Betweenness of values of ρ: call it $B_\rho(x, y, z)$. And we expect, of course:

a synthetic axiomatization of Newtonian/Galilean spacetime in Hilbert-Robb style; using e.g. a 'Hilbertian' spatial betweenness predicate holding between (*absolutely!*) simultaneous spacetime points x, y, z: call it say, $B_{space}(x, y, z)$;

synthetic axioms governing the predicates that encode mass density and gravitational potential, and their relations to Newtonian spacetime.

Field (1980) presents such a *nominalization* of Newtonian gravitational theory. In order to argue that:

(i) his nominalization is adequate; i.e. captures the (atheistically acceptable!) content of the usual (naughtily platonistic!) formulation(s);

(ii) the usual formulations' pure mathematical objects are 'only useful, but not indispensable':

he proves a *representation theorem*. Namely: Every model of his axiom system is isomorphic to a model of a usual formulation. More precisely ...

A = Field's nominalized formulation of Newtonian gravitational theory; with its many models (of course: non-isomorphic, since the distributions of ϕ, ρ can vary).

O = (a model of) Newtonian gravitational theory in a usual formulation; with its many (again, of course: non-isomorphic) models. With, in each model, some choices of: spatial origin (absolutely at rest, or inertially moving in the Newtonian/Galilean cases, respectively), and spatial axes, and units for spatial and temporal length in $\mathbb{R}^4$.

F = the Newtonian/Galilean symmetry group for spacetime structure, augmented by (spacetime-constant) dilations, acting on O. As in Examples 1 and 2: the various kinds of element of F—spacetime translations, spatial rotations, boosts and dilations—reflect the freedom, given a choice of inertial coordinates for Newtonian/Galilean spacetime, respectively: to shift spacetime origin, to rotate spatial axes, to boost, and to change the unit of length.

(There is a subtle discussion to be had about whether F should also be taken as the *dynamical* symmetry group. But I again set aside issues about dynamics!)

The main philosophical themes are:

(a): Is it really so that we should reject mathematical platonism? More specifically, about our topic of representation theorems: is A really more cognitively secure than the platonistic O? (Cf. theme (A) in Section 1 above).

(b): Can this nominalizing strategy be carried out for other physical theories? What about theories with a curved spacetime? Indeed with a dynamical spacetime? Even if we do not adopt Field's mathematical atheism requiring heterodox treatment of the field-content, there is (so far as I know) no synthetic 'Hilbert-Robbian' characterization of them ...And anyway: what about theories cast in very different terms, of a state space: like Lagrangian or Hamiltonian mechanics, or field theory?

(c): Field's dialectical position is subtle: as follows. He the mathematical atheist speaks of the 'forbidden fruit' (i.e. the allegedly non-existent), numbers etc. in his representation theorem, i.e. in his attempt to prove to us that we do not need to refer to them to do science, and so should not believe in them. This is certainly subtle: maybe it is unstable. It also raises the interesting technical, and dialectically relevant, question. What about the mathematical commitments of the *methods* by which he proves the representation theorem, rather than the *content* of it, i.e. what is explicitly mentioned by the statement of the theorem? Do those *methods* commit him to mentioning 'forbidden fruit'? If so, is that mention dialectically legitimate?

(d): Field uses (and of course cites) a tradition of *representation theorems* in the foundations of empirical measurement in general (Suppes, Luce Raiffa): which are taken to justify a *numerical* measure of physical quantity (and its associated unit and scale conventions) starting from suitable non-numerical properties and relations of empirical objects. For example, they have a theorem about mass measurement, with:

A = axioms for Ôbeing more massive thanÕ as a partial order on some objects (usually assumed to be finite in number); also with an operation of mereological fusion (think Ôweighing togetherÕ) on objects.

$O = \mathbb{R}$, considered as values of the objects' masses. The 'isomorphism' is now an *embedding*: i.e. the function that makes the assignment of real-number masses to the objects. There is a natural zero, the mass of the empty object. But no natural unit. And all masses are to be positive, let us say. So F = linear maps from $\mathbb{R}^+$ to $\mathbb{R}^+$; so F is itself given by $\mathbb{R}^+$. So here, as to Section 1's initial theme (C): the redundancy or artefacts in O is the choice of a unit of mass. So there is a philosophical literature about such theorems.

The representation theorem of this kind that, so far as I know, is deepest, mathematically and conceptually, is the *Helmholtz-Lie theorem* that the free mobility of rigid bodies requires a space of constant curvature. But I set it aside: (sad to say, it has dropped out of the philosophical literature ...).

(e): The philosophical literature has also pursued questions in logic raised by Field's programme: e.g. about the relation between representation theorems and proofs of conservativity; and about the need to invoke scecond-order logic to characterize continuity.

(f): Should we be substantivalists about spacetime?! This question, and the mention in (d) of *embedding* a finite number of empirical objects into a pure mathematical structure (in (d): into $\mathbb{R}$), prompts our final example...

4. *Relationism a la Mundy: 1980s*:–

The question, 'Should we be substantivalists about spacetime?!' is of course enormous. It runs from the Leibniz-Clarke Correspondence of 1710, through e.g. Mach, Barbour, and Earman, to debates about the required basic spatiotemporal ingredients of quantum gravity. But here, we pick off a tiny fragment, which is adjacent to the topic of Hilbert-Robbian representation theorems: though it is a matter of embedding, rather than isomorphism.

The idea (Mundy 1982) is:— Imagine N point-particles in Euclidean space ($\mathbb{R}^3$, or its affine cousin) with their induced relations of distance and angle amongst themselves. The relationist wants to say: the correct physical account is:

'These particles, and their mutual relations of distance and angle can be *embedded* in Euclidean space. This embedding is of course with the usual freedom, viz. up to a translation and rotation. That freedom represents a redundancy in the usual substantivalist formulation i.e. invoking an ambient space'.

And similarly, for the motion of N point-particles (i.e. taking account of velocity/momentum), and for N point-events in Minkowski spacetime.

It is clear that even if this proposal can be made to work to give an anti-substantivalist account of the geometry of point particles in Euclidean space, or point-events in Minkowski spacetime, there will remain much work, to get:

(i) a theory of motion;

(ii) an analogue for field theories: which seem in any case to undercut relationism, since they replace the (allegedly suspicious) void by a plenum, which is presumably relationistically acceptable: (though of course there are issues about patches of space or spacetime where the field vanishes).

But let us focus on the simple case of N point-particles , or N point-events. The comments for these are the same: so I shall say 'particles', for short. There are three main concerns.

(1): Is the relationist formulation, attributing relations of distance and angle among the N particles *parasitic* on the substantivalist formulation? The main problem here is that there are many constraints about such relations, the simplest being the triangle inequality, which are easily understood from a substantivalist perspective as restrictions to the N particles of an ambient geometry of space—but which the relationist must take as 'brute facts' about the particles that prompt the 'illusion' of an ambient space.

NB: looking further afield. This kind of problem—'write down the inequalities among relative distances that are together sufficient to ensure an embedding in to a given space'—has a considerable history: e.g. in Tait's problem of characterizing inertial motion of N noon-interacting particles: (Tait 1884, discussed in Barbour (1999, p. 1000 et seq).

It also has delightful, and insightful, analogues in probability theory: viz. 'write down the inequalities among single and joint probabilities that are together sufficient to ensure they can be given a Kolmogorov (i.e. orthodox classical) representation'. This was addressed by Boole himself, and rediscovered in the philosophical literature about *Bell's theorem*—which characterizes classical probability functions by linear inequalities ('Bell inequalities') among single and joint probabilities.

(2): Embedding N particles only uses up a bounded region R of $\mathbb{R}^3$ (or $\mathbb{R}^4$): so the relationist's reasons for asserting the embedding are equally reasons to assert an embedding into a space that is globally very unlike $\mathbb{R}^3$ (or $\mathbb{R}^4$): by being bounded, or having another global topology; or by being curved beyond the region R. So the relationist strategy fails to capture our commitment to e.g. Euclidean space. This kind of problem is discussed by Earman (1989: 166-170).

(3): What about the role of possibility? The relationist formulation must admit there are various possible configurations of the N particles (all obeying the many constraints listed in (1)). Can they also legitimately appeal to *possible*, not actual, particles? In order, e.g., to address the problems in (2). If so, how? This kind of problem is discussed by Manders (1982: 166-170), Butterfield (1984), Earman (ibid) and Belot (2011).

References

Belot, G. (2011), *Geometric Possibility*, Oxford University Press.
Butterfield, J. (1984), Relationism and possible worlds, *British Journal for Philosophy of Science* **35**: 101-113.
Earman, J. (1989), *World Enough and Spacetime*, MIT Press.
Field, H. (1980), *Science without Numbers*, Oxford: Blackwell.
Manders, K. (1982), On the spacetime ontology of physical theories, *Philosophy of Science*, **49**: 575-590.
Mundy (1982), 'Relational theories of Euclidean space and Minkowski spacetime', *Philosophy of Science*, **50**, 205-226.
Mundy (1986), 'Optical Axiomatization of Minkowski Space-Time Geometry', *Philosophy of Science*, **53**, 1-30.
Mundy (1992), 'Spacetime and isomorphism', *Proceedings of the PSA 1992 conference*, volume 1, pp. 515-527.
Tarski, A. (1959), 'What is elementary geometry?' in L Henkin, P Suppes and A Tarski eds., *The Axiomatic Method, with special reference to geometry and physics*, North Holland: pp16-29; variously reprinted.
Winnie, J. (1977), 'The causal theory of time', in J. Earman, C. Glymour and J. Stachel (eds.), *Foundations of Spacetime theories*, Minnesota Studies in Philosophy of Science volume 8.
Zeeman, C. (1974), 'Causality implies the Lorentz group', *Journal of Mathematical Physics*.