

Relativity theory between structural and dynamical explanations

Mauro Dorato
Department of Philosophy
University of Rome 3
Via Ostiense 234
00146, Rome, Italy
dorato@uniroma3.it

Phenomena like rod contractions and clocks retardations have attracted the attention of philosophers since the publication of Einstein's original paper on special relativity (SR). Very often, the main question that has been raised about such relativistic effects is: are they real? Of course the answer depends on what one means by the metaphysically appealing but philosophically unsuitable adjective "real" in our context. If "real" means "measurable", then the answer ought to be an uncontroversial "YES", as every experimental physicist working at CERN knows. If "real" means "dynamical", then I guess that the vast majority of physicists and philosophers would answer with a resounding "NO". Can't we explain the relativistic effects as a projection of four-dimensionally conceived rods and clocks onto different, arbitrarily chosen inertial frames of Minkowski spacetime?

In various papers, Harvey Brown and some co-workers (especially Oliver Pooley, a former, brilliant student of his) have been arguing for some years that length contractions and clock dilations, so far often regarded as purely *kinematical* effects, need a *dynamical*, presumably quantum explanation in terms of Lorentz covariant laws, an explanation that does not require the postulation of a privileged inertial frame. To my knowledge, Brown's *Physical Relativity. Space-time structure from a dynamical perspective* is the first book-length defence of this interesting thesis (Brown 2005), resurrected by John S. Bell (1976) some decades ago, but running through last century like an underground river that every now and then emerges to the surface.

The *Leitmotiv* running through the book, which for the most part presents previously unpublished material, is Einstein's famous distinction between two types of physical theories, the "Principle-theories" and the "Constructive theories". The former, like thermodynamics, are empirically based on general characteristics of empirical processes; the latter, like statistical mechanics, build up, in Einstein's word "a picture of the more complex phenomena out of the materials of a relatively simple formal scheme" (p.71). This distinction runs deep into questions pertaining to the nature of scientific explanations, but Brown stays away from either methodological or metaphysical issues, in order to concentrate solely on the historical-foundational problems raised by the relativistic effects.

Historical considerations are particularly important from Brown's perspective, as they are central to explain why most physicists, partially influenced by a wrongheaded picture of Einstein's contribution to SR *vis à vis* Hendrik Lorentz's, are still convinced that contractions and dilations are kinematical effects. As chapter 5 convincingly shows, Einstein's decision to favour the formulation of SR as a Principle theory was a consequence of the shaky ground on which the newly born quantum theory was resting in 1905 (for a similar claim, see Rynasiewicz 2005). In view of the fact that the light quantum hypothesis, put forward by Einstein in the same year, had already convinced him that Maxwell's equation "could not claim to exact validity", the light Postulate of SR had to be formulated independently of any theory about the constitution of light (p.77). For similar reasons, Brown argues, the other pillar of SR, the Relativity Principle, was originally presented by Einstein in such a way as to transform Lorentz's (and Poincaré's) problem – explaining the Relativity Principle by deriving it from Lorentz covariant dynamical laws governing the inner structure of rods and clocks – into a postulate.

In various parts of the book, Brown shows how Einstein, however, did not like the view of rods and clocks as "primitives" of the theory, and thought of them as "moving atomic

configurations”: according to Einstein, SR should ideally be formulated as a constructive theory, in the same sense in which statistical mechanics “builds up” the thermal phenomena. It then follows that the distance between Lorentz and the other “trailblazers” on the one hand (George FitzGerald, Joseph Larmor, Henri Poincaré), and Einstein on the other, is not at all as vast as it has been pictured for years from the 60ies onward. Brown’s contribution in favour of regarding the theory change before and after 1905 as being not radically discontinuous can hardly be overestimated, in the same sense in which we should not underestimate the advantage of formulating SR by using the “exemplar” of thermodynamics in Kuhn’s sense. The Relativity Principle, constraining as it does the behaviour of clocks and rods, will not easily be refuted by future physics, and will enjoy the same stability as the Principles of thermodynamics.

Parallel to the attempt at showing the importance of a dynamical underpinning of SR is Brown’s more disputable thesis that no other explanation of contractions and dilations is available: “a moving rod contracts and a moving clock dilates because of what it is made of and not because of the nature of their spatiotemporal environment” (p.8). If this claim amounted to the reasonable thesis – already advanced by Robert Di Salle (1995) – that the Minkowski metric is but a “codification” of the behaviour of rods and clocks, and cannot in any way regarded as a *cause* of such a behaviour, I would wholeheartedly agree, for many of the reasons provided by Brown in chapter 8.

However, as Brown knows, the correctness of this claim by itself does not entail that such a behaviour is caused by something else. We don’t need to consider spacetime as the cause of a Lorentz deformation to appreciate the fact that the view that regards entities as four-dimensional worldtubes exemplifying the structure of Minkowski spacetime might yield a satisfactory explanation of the reason why the shape of a rod is different in different inertial frames, without having to introduce any dynamical hypotheses. Considering that Brown

himself elsewhere seems to have partially endorsed this view¹, it is surprising that its consequences for his main problem – why do physical entities conform to Minkowski geometry? – in the book have not received sufficient attention. This drawback is probably a consequence of the fact that Brown never discusses the nature of physical explanations in general, and does not defend a distinctive view about it. Sometimes we find him tacitly endorsing a deductive-nomological model of explanation, as when he claims that for the first time in history the general theory of relativity explains inertia (p.141), while throughout the book he seems to be defending a causal/dynamical account of relativistic effects, against a structural view of explanation of the type defended by R.I.G. Hughes (1989 and 1993).

In spite of the fact that a summary is not going to do justice to the wealth of historical and conceptual details to be found in the book, the following précis may help the reader to grasp its structure.

In chapter 1, which is an overview of the whole book, we are introduced to our historical journey, which begins with the Irish physicist George F. FitzGerald who, in a little note published in *Science* in 1889, tried to explain the null result of Michelson and Morley's 1887 experiment. By following an idea of Oliver Heaviside's, FitzGerald "feigned the hypothesis" that motion could have an effect on the electrostatic forces acting within the body by distorting its shape (not necessarily by contracting it).

In chapter 2 we find a condensed criticism of the thesis that contractions and dilations can be understood only *via* the usual Lorentz transformation of spacetime coordinates, and therefore in a pure kinematic fashion. There are two remarkable arguments in this chapter. The first concern the nature of inertia, the other the role of the relativity of simultaneity in explaining the relativistic effects.

¹ See Brown and Pooley (2005, p.79), where we read that explanations of contractions in terms of 4-dimensionality and the structure of Minkowski spacetime are "perfectly acceptable explanations (perhaps the only acceptable explanations) of the explananda in question." I thank Oliver Pooley for having reminded me of this passage.

The former argument is important, because it illustrates Brown's tendency to look for explanations where they might simply be unnecessary. For instance, he wonders why force-free particles conspire to follow rectilinear paths without being able to influence each other and adds that "anyone who is not amazed by this conspiracy has not understood it" (p.15). I must belong to those who have not understood inertia: isn't going in straight paths (or following geodesics in lorentzian manifolds) simply part of the definition of inertial motion, namely a correspondence rule or meaning postulate linking a geometric notion (being straight or minimizing distance) with a physical concept (being force free)? Brown sees this objection, of course, but then adds that what has empirical content is "the claim that inertial coordinate systems exist, applicable to all free bodies in the universe" (p.15)². But even if the law of inertia is an empirical existence claim, it would still not follow that the "conspiracy" above is something that needs to be explained, and is explained by the general theory of relativity (GR). In chapter 9 we read that "GR is the first in the long line of dynamical theories, based on that profound Aristotelian distinction between natural and forced motions of bodies, that explains inertial motion (p. 141). The explanation is obtained via the vanishing of the covariant divergence of the stress-energy tensor (*ibid.*).

It seems to me that Brown's view of explanation neglects the fact that scientific revolutions are often characterized by the fact that phenomena previously thought to be needing a causal explanation are declared to be "natural", and the transition from the Aristotelian to the Newtonian physics shows this transformation of a problem into a postulate *a propos* of the question "why do bodies continue to move?". If inertial motion is the new natural state of bodies in Aristotle sense, it does not need to be explained by impetus theories or by other causal paraphernalia of the Aristotelian tradition. In the new Newtonian framework, what needs to be explained is rather the deviation from inertial motion, *via* the introduction of

² Newton was well aware that such an existence claim could be false for ordinary bodies, and this is the main reason why he postulated absolute space; after the elimination of absolute space, for the law of inertia to have empirical meaning, it could be sufficient to have moving bodies that are approximately inertial.

forces (the causes of such a deviation). With GR, however, free fall becomes the *new* natural state of bodies, for the simple reason that it is locally identifiable with inertial motion: it is for this very reason that gravitational forces (*qua* causes) in GR are dispensable. Brown is correct in remarking that the fact that spinless test particles follow geodesics “can be derived by the vanishing of the covariant divergence of the stress energy tensor associated with the body” (p.161). But I would not regard this as an “explanation” of inertial motion (is Brown relying on the nomological-deductive model of explanation here?) for two reasons. First, inertial motion and free fall may be regarded as standards against which we explain the rest of motions (they are “natural motions” in the sense briefly illustrated above). Second, since the association between the geometrical concept of “geodesic” and the physical notion of “being force free” is a precondition for applying the mathematical model described by Einstein field equations to the physical world, the explanation in question seems to me circular.

As to the second argument of the chapter, Brown claims that the relativistic effects do not presuppose the relativity of simultaneity (p. 29, note 43), because rods and clock protocols are defined with respect to a single frame. That is true of course, but if measuring the length of any rod in any frame presupposes the simultaneity of its ends, why is the relativity of simultaneity insufficient to explain Lorentz’s deformations? The standard line of argument, according to which if simultaneity were not relative we would not have rods contractions, should have not been confined to a footnote, since *it* tends to push the explanation of the relativistic effects toward kinematics. With respect to this problem, Brown claims that unless spacetime has a metric or a scale factor, the cross sections of the rod relative to two different frames could not be compared, so that the relativity of simultaneity is neither necessary nor sufficient for the contraction. But if we agree that “contraction” is a purely relative phenomenon (according to Brown there is no privileged frame), and if we endorse the

innocuous claim that the laws governing the behaviour of the constituents of rods and clocks are Lorentz covariant, it is not clear to me what else we have to explain.

In chapter 3 Brown traces an historical sketch leading from Galileo's formulation of the Relativity Principle to Newton's. By imagining that, in order to investigate the constraints of dynamics, a fictitious German physicist living in the seventeenth century (Albert Keinstein) postulated rather than deriving the Relativity Principle (like Newton did), Brown expresses two main ideas: (1) the Relativity Principle has non-trivial implications for the nature of dynamical interactions (p.35); (2) the connection between kinematics and dynamics is much closer than it is usually thought (p. 40).

After an examination of the kinematics of Michelson and Morley's celebrated experiment from a contemporary perspective, in chapter 4 we come back to FitzGerald and Heaviside, then to Lorentz, Larmor's book *Aether and Matter* (Larmor 1900) and Poincaré, namely to those "trailblazers" who attempted to explain Michelson and Morley's null result before Einstein. The main aim of the chapter is to show that (i) none of such attempts really hinged on the existence of an ether, that (ii) the latter is only a peg for the electromagnetic field, and that (iii) the role of the effect of motion on the molecular forces responsible for the contraction should not have been abandoned after the abandonment of the ether. Brown also raises a point that is very probably unfamiliar to most readers: neither FitzGerald nor Lorentz really put forth a contraction hypothesis, but simply the conjecture that the dimensions of the body are altered in a certain ratio (p. 45-55).

He then concludes: "shape deformation produced by motion is far from the proverbial riddle wrapped in mystery" (p. 52). But one could ask: how can purely relative "motion" suffice for relativistic effects to occur, given that Brown assumes that there is no need of a privileged frame? If inertial motion is purely relative, and deformations effects are a function of velocity, one would expect that *velocity*, in Brown's understanding of the notion, has *more*

than a relative significance. But if the amount of deformation depends on the arbitrary choice of the measuring frame, why should we grant the deformation a dynamic significance, rather than a simple perspectival significance? If relative to frame f the deformation is $F(f)$, relative to frame f' is $F(f')$, relative to f'' is $F(f'')$ and so on, the implication that there is *no matter of fact* as to the 3 dimensional shape of the body is quite natural, since length or shape in SR are not invariant notions. However, objections to arguments like these in the book are neglected.

After a variation on the theme (chapter 6), involving a very interesting discussion on the nature of a clock, the rejection of the non-conventionality of simultaneity, and a derivation of the so-called Ignatowski transformations (which don't presuppose the light principle), in chapter 7 we return to the main historical narrative. Brown sketches the contributions given by Hermann Weyl, and by Wolfgang Pauli and Arthur Eddington (in the twenties) on dynamical explanations. We then read about W.E.G. Swann (30s and 40s), and L. Jánossi and Bell's contribution in the seventies. Most readers will not be familiar with Jánossi's (1971) attempt at explaining in a detailed way the "mechanism of Lorentz deformation", the idea being that external forces necessary to boost the system eventually die down, and the system relaxes back to equilibrium after the elastic disturbances due to such forces (p. 124).

In chapter 8 Brown faces the greatest challenge to his approach: the possibility of giving a "structural" explanation of Lorentz's "deformations" and clock dilations (see Hughes 1989, 1993 and Clifton 1998). It is in particular in this chapter that the decision of not taking stance on the nature of explanation in physics is detrimental to the force of some of his arguments. For instance, why should all explanations be dynamical? Brown neglects the possibility that the relativistic "deformations" could be given different explanations, one in terms of the four-dimensional structure of Minkowski spacetime, and one in terms of dynamical phenomena (provided that the latter are really to be counted in). There is a striking analogy with the example of a balloon left in the cabin of an aircraft before take off, made famous by Wes

Salmon (Salmon 1990). The forward movement of the balloon during take off can be explained *both* by the Principle of Equivalence *and* by causal mechanisms due to the pressure gradient of the air that are a consequence of the acceleration of the aircraft. Exactly as the two explanations in this case complete and don't exclude each other, I submit that in Brown's case the dynamical and the geometrical explanations might well coexist, a possibility that Brown does not consider.

According to Brown, Friedman's reading the theory of relativity as a theory about the four-dimensional manifold "puts the cart before the horse" by reverting the order of explanation: Brown argues that Minkowski spacetime has the geometrical structure it has in virtue of the Lorentz covariance of the physical laws obeyed by, and exemplified by physical systems, and not *viceversa*. Yuri Balashov and Michel Janssen (2003) defended the opposite viewpoint: "the geometrical structure is the *explanans* ... and the invariance of forces the *explanandum* (quoted in the book, p. 133). But why not solve this dispute by a conventionalist move? That is, why not claim that this question about the order of explanation in the end is a pseudo-question, since part of what it means for Minkowski spacetime to have a certain geometrical structure is simply that the alleged laws are Lorentz invariant?

Of course, at this point one can always ask: *why* does such a stipulation hold? And here, if I had to take a stand, I would go along Brown's line and claim that in SR the geometric structure depends on the physics and laws of the system. But note that given the new, dual nature of the metric tensor in GR, in the latter theory this separation of geometry from physics cannot be done, so that the question about the order of explanation seems to arise only for SR.

After interesting and informative discussions of the possibility of offering geometrical explanations of phenomena *via* configuration spaces of Riemannian geometry, geometrical readings of quantum theory *à la* Jeeva Anandan, and Costantin Carathéodory's approach to thermodynamics as the analogue of an abstract approach to SR, we land into the final, ninth

chapter. Here, following Eddington (1923), we are presented with two different ways of building GR: one starts from rods and clocks and arrives at matter *via* the metric tensor; the other starts from matter (energy tensor) and gets to the metric field. Brown brilliantly defends the latter, less familiar way to the theory, thereby showing that the dynamical interpretation of SR is consistent with GR (p.151).

The final section³ of the ninth chapter is particularly important, as it points toward the fact that once we endorse a dynamical underpinning of SR, we must also solve the conflict between gravitational phenomena and quantum mechanics. And here one may wonder with Brown whether a constructive account of the relativistic effects even just in terms of quantum field theory can ever be given, given its complexity: “I am unaware of a systematic derivation of the stable structure of even the hydrogen atom within relativistic quantum field theory” (p.148, n. 55). Especially one question seems missing in this respect: given the lack of a quantum theory of measurement, how can we presently hope to give a dynamical, quantum explanation of the contractions of classically conceived rods? Don’t we need to solve the measurement problem first?

Brown is silent on this question, which, however, suggests an interesting analogy between a dynamical approach to SR and quantum mechanics, which is not considered by the author, despite his admiration for John Bell. One could argue that there has to be a theory of measurement in terms of quantum mechanics, in the same sense in which there has to be a quantum theory of the solid structure of rods undergoing changes in shape, and much for the same reason. Phenomenologically, both QM and STR work quite well by taking classical objects as primitive and given (apparatuses in QM, rods in STR), but we might want to know why Born’s rule is so successful in the same sense in which we want to know why rigid bodies undergo measurable contractions in moving frames.

³ There are two appendixes to the book, one devoted to the change of attitudes in Einstein’s view of covariance, the other to special relativity and non-locality.

But notice the limits of the analogy: if QM is linear and complete, it does not explain what we see (at least without further assumptions, like decoherence, which in my opinion don't solve the measurement problem). In the case of SR, instead, the need for a constructive explanation is not so dramatic, as it is not spurred by outright conflicts with our experience or with the experiments. And one could even put forth an account of the relativistic effects that point toward a dissolution of the need to explain. "Length at a given time" and "clock's readings in a given space" are non-invariant notions, in the same sense in which the particular projection of the four-dimensional electromagnetic field into an electric and a magnetic field depends on a particular choice of a 3+1 splitting of space-time. Would it make sense to ask why is the electric field the way it is in this frame and it is different in another frame?

Once again, the impression that a more thorough discussion of the *Erlangen-Program's* view that the objective and the invariant are one and the same thing would have strengthened the book: if there is no such thing as "length at a time", why do we need a dynamically based explanation for the relative change in length of rigid bodies? Examples like that illustrated by Bell, in which a rope linking two bodies accelerating away from an inertial frame eventually breaks, might have helped Brown's cause.

In any case, I want to make it clear that by focussing on my criticism of parts the book, I have tried to pay tribute to its inspiring and thought-provoking content, which radically changes our historical understanding of the transition between Lorentz's and Einstein's so-called "research programs". Critical of the mainstream understanding of SR, the book infuses new life into what is today still an "heretical" way of looking at the theory, thereby contributing to deepen our understanding of it, even for those who disagree with its main thesis. The author shows a remarkable familiarity with the technical literature, and with both the history and the foundations of the special and general theory of relativity. I conclude with no hesitation that this significant contribution to the tradition of historical-critical analysis of

scientific theories *à la* Mach is an absolute “must” for physicists and philosophers interested in the special and general theory of relativity, as well as in the nature of scientific explanation.

References

- Balashov Y. and Janssen M. (2003), “Critical notice: Presentism and Relativity”, *British Journal for the Philosophy of Science*, 54, pp. 327-346.
- Bell J. (1976), “How to teach special relativity”, *Progress in Scientific Culture*, 1, reprinted in Bell J. (1987), *Speakable and unspeakable in Quantum Mechanics*, Cambridge University Press, pp. 52-62.
- Brown H. (2005), *Physical Relativity. Spacetime Structure from a dynamical perspective*. Oxford University Press, Oxford, xiii + 225 pp., ISBN 0-19-927583-1, £ 32.00 (hardback)
- _____ and Pooley O. (2005), “Minkowski Spacetime: a Glorious Non-Entity”, in Dieks D. (ed.) *The Ontology of Spacetime*, Elsevier, pp. 67-89.
- Clifton R. (1998), *Scientific Explanation in Quantum Theory*, in <http://philsci-archive.pitt.edu/archive/00000091/00/explanation-in-QT.pdf>
- DiSalle R. (1995), “Spacetime theory as physical geometry”, *Erkenntnis*, 42, pp. 317-337.
- Eddington A. (1923), *The Mathematical Theory of Relativity*, Cambridge University Press.
- Hughes R.I.G. (1989), *The Structure and Interpretation of Quantum Mechanics*, Harvard University Press.
- _____ (1993), “Theoretical Explanation”, *Midwest Studies in Philosophy* 18, pp. 132-153.
- Jánossi L. (1971), *Theory of Relativity based on Physical Reality*, Akadémie Kiadó, Budapest.
- Larmor J. (1900), *Aether and matter*, Cambridge University Press, Cambridge.
- Rynasiewicz R. (2005), “The Optics and Electrodynamics of ‘On the electrodynamics of moving bodies’”, *Annalen der Physik*, 14, Supplement, pp. 38–57.
- Salmon, W. (1990) *Four Decades of Scientific Explanation*. University of Minnesota Press.

Acknowledgments

I am highly indebted to Harvey Brown, Michel Ghins and Oliver Pooley for their generous comments to a previous version of this review. I also thank Massimo Pauri for useful conversations on some of the topics mentioned in the paper