

Symmetries and invariances in classical physics

Katherine Brading* and Elena Castellani†

December 17, 2005

Abstract

Symmetry, intended as invariance with respect to a transformation (more precisely, with respect to a transformation group), has acquired more and more importance in modern physics. This Chapter explores in 8 Sections the meaning, application and interpretation of symmetry in classical physics. This is done both in general, and with attention to specific topics. The general topics include illustration of the distinctions between symmetries of objects and of laws, and between symmetry principles and symmetry arguments (such as Curie's principle), and reviewing the meaning and various types of symmetry that may be found in classical physics, along with different interpretative strategies that may be adopted. Specific topics discussed include the historical path by which group theory entered classical physics, transformation theory in classical mechanics, the relativity principle in Einstein's Special Theory of Relativity, general covariance in his General Theory of Relativity, and Noether's theorems. In bringing these diverse materials together in a single Chapter, we display the pervasive and powerful influence of symmetry in classical physics, and offer a possible framework for the further philosophical investigation of this topic.

Contents

1 Introduction

The term 'symmetry' comes with a variety of ancient connotations, including beauty, harmony, correspondence between parts, balance, equality, proportion, and regularity. These senses of the term are clearly related to one another; the concept of symmetry used in modern physics arose out of this family of ideas. We are familiar with the approximate symmetries of physical objects that

*Department of Philosophy, 100 Malloy Hall, University of Notre Dame Notre Dame, Indiana 46556. E-mail: kbrading@nd.edu

†Department of Philosophy, University of Florence, via Bolognese 52, 50139, Firenze, Italy. E-mail: elena.castellani@unifi.it

we find around us – the bilateral symmetry of the human face, the rotational symmetry of a snowflake turned through 60° , and so forth. We may define a symmetry of a given geometric figure as the invariance of that figure when equal component parts are exchanged under a specified operation (such as rotation). The development of the algebraic concept of a group, in the nineteenth century, allowed a generalization and refinement of this idea; a precise mathematical notion of symmetry emerged which was applicable not just to physical objects and geometrical figures, but also to mathematical equations – and thus, to what is of particular interest to us, the laws of physics expressed as mathematical equations. The group theoretical notion of symmetry is the notion of invariance under a specified group of transformations. ‘Invariance’ is a mathematical term: something is invariant when it is left unaltered by a given transformation. This mathematical notion is used to express the notion of physical symmetry that we are interested in, i.e. invariance under a group of transformations. This is the concept of symmetry that has proved so successful in modern science, and the one that will concern us in what follows.

We begin in Section 2 with the distinction between symmetries of objects and of laws, and that between symmetry principles and symmetry arguments. This section includes a discussion of Curie’s principle. Section 3 discusses the important connection between symmetries, as studied in physics, and the mathematical techniques of group theory. We offer a brief history of how group theoretical techniques came to be of such central importance in twentieth century physics. With these considerations in mind, Section 4 offers an account of what is meant by symmetry in physics, and a taxonomy of the different types of symmetry that are found within physics. In Section 5 we discuss some applications of symmetries in classical physics, beginning with transformation theory in classical mechanics, and then turning to Einstein’s Special and General Theories of Relativity (see Section 6). We focus on the roles and meaning of symmetries in these theories, and this leads into the discussion of Noether’s theorems in Section 7. Finally, in Section 8, we offer some concluding remarks concerning the place, role and interpretation of symmetries in classical physics.

2 Symmetries of objects and of laws

That we must distinguish between symmetries of objects versus symmetries of laws can be seen as follows. It is one thing to ask about the geometric symmetries of certain objects – such as the 60° rotational symmetry of a snowflake and the approximate bilateral symmetry of the human face mentioned above – and the *asymmetries* of objects – such as the failure of a chair to be rotationally symmetric. It is another thing to ask about the symmetries of the laws governing the time-evolution of those objects: we can apply the laws of mechanics to the evolution of our chair, considered as an isolated system, and these laws are rotationally invariant (they do not pick out a preferred orientation in space) even though the chair itself is not. Re-phrasing the same point, we should distinguish

between symmetries of *states* or *solutions*, versus symmetries of laws. Having distinguished these two types of symmetry we can, of course, go on to ask about the relationship between them: see, for example, current discussions of Curie’s principle, referred to in Section ??, below.

2.1 Symmetry principles and symmetry arguments

It is also important to distinguish between *symmetry principles* and *symmetry arguments*. The application of symmetry principles to laws was of central importance to physics in the twentieth century, as we shall see below in the context of Einstein’s Special and General Theories of Relativity. Requiring that the laws – whatever their precise form might be – satisfy certain symmetry properties, became a central methodological tool of theoretical physicists in the process of arriving at the detailed form of various laws.

Symmetry arguments, on the other hand, involve drawing specific consequences with regard to particular phenomena on the basis of their symmetry properties. This type of use of symmetry has a long history; examples include Anaximander’s argument for the immobility of the Earth, Archimedes’s equilibrium law for the balance, and the case of *Buridan’s ass*.¹ In each case the associated argument can be understood as an example of the application of the Leibnizean Principle of Sufficient Reason (PSR): if there is no sufficient reason for one thing to happen instead of another, then nothing happens (i.e. the initial situation does not change). There is something more that the above cases have in common: in each of them PSR is applied on the grounds that the initial situation has a certain *symmetry*.² The symmetry of the initial situation implies the complete equivalence between the offered alternatives. If the alternatives are completely equivalent, then there is no sufficient reason for choosing between them and the initial situation remains unchanged. Arguments of this kind most frequently take the following form: a situation with a certain symmetry evolves in such a way that, in the absence of an asymmetric cause, the initial symmetry is preserved. In other words, a breaking of the initial symmetry cannot happen without a reason: an asymmetry cannot originate spontaneously. This style of argumentation is also to be found in recent discussions of ‘Curie’s principle’, the principle to which we now turn.

2.2 Curie’s principle

Pierre Curie (1859-1906) was led to reflect on the question of the relationship between physical properties and symmetry properties of a physical system by his studies on the thermal, electric and magnetic properties of crystals, since these properties were directly related to the structure, and hence the symmetry, of the crystals studied. More precisely, the question he addressed was the following: in a given physical medium (for example, a crystalline medium) having specified symmetry properties, which physical phenomena (for example, which

¹For a discussion of these examples, see Brading and Castellani (2003, ch. 1, Section 2.2).

²In the first case rotational symmetry, in the second and third bilateral symmetry.

electric and magnetic phenomena) are allowed to happen? His conclusions, systematically presented in his 1894 work ‘Sur la symétrie dans les phénomènes physiques’, can be summarized as follows:³

- (a₁) When certain causes produce certain effects, the symmetry elements of the causes must be found in their effects.
- (a₂) When certain effects show a certain dissymmetry, this dissymmetry must be found in the causes which gave rise to them.⁴
- (a₃) In practice, the converses of these two propositions are not true, i.e., the effects can be more symmetric than their causes.
- (b) A phenomenon may exist in a medium having the same characteristic symmetry or the symmetry of a subgroup of its characteristic symmetry. In other words, certain elements of symmetry can coexist with certain phenomena, but they are not necessary. What is necessary, is that certain elements of symmetry do not exist. Dissymmetry is what creates the phenomenon.

Conclusion (a₁) is what is usually called *Curie’s principle* in the literature. Conclusion (a₂) is logically equivalent to (a₁); the claim is that symmetries are necessarily transferred from cause to effect, while dissymmetries are not. Conclusion (a₃) clarifies this claim, emphasizing that since dissymmetries need not be transferred from cause to effect, the effect may be more symmetric than the cause.⁵ Conclusion (b) invokes a distinction found in all of Curie’s examples, between the ‘medium’ and the ‘phenomena’. We have a medium with known symmetry properties, and Curie’s principle concerns the relationship between the phenomena that can occur in the medium and the symmetry properties – or rather, ‘*dissymmetry*’ properties – of the medium. Conclusion (b) shows that Curie recognized the important function played by the concept of dissymmetry – of *broken symmetries* in current terminology – in physics.

In order for Curie’s principle to be applicable, various conditions need to be satisfied: the cause and effect must be well-defined, the causal connection between them must hold good, and the symmetries of both the cause and the effect must also be well-defined (this involves both the physical and the geometrical properties of the physical systems considered). Curie’s principle then furnishes a necessary condition for given phenomena to happen: only those phenomena can happen that are compatible with the symmetry conditions stated by the principle. Curie’s principle has thus an important methodological function: on

³For an English translation of Curie’s paper, see Curie (1981); some aspects of the translation are misleading.

⁴Curie uses the term *dissymmetry* in his paper, as was current at his time. The sense is the same of that of *symmetry breaking* in modern terminology, which is today often identified with the sense of *asymmetry*. To be more precise one should distinguish between the result of a symmetry-breaking process (*broken symmetry*), the absence of *one* of the possible symmetries compatible with the situation considered (*dissymmetry*, as it was called in the nineteenth century literature, notably by Louis Pasteur in his works on molecular dissymmetry), and the absence of *all* the possible symmetries compatible with the situation considered (*asymmetry*).

⁵Note that for some authors conclusion (b) is a principle on its own. Radicati (1987) goes further, describing conclusions (a₁), (a₂) and (b) as three different principles: Curie’s first, second and third principle, respectively.

the one hand, it furnishes a kind of selection rule (given an initial situation with a specified symmetry, only certain phenomena are allowed to happen); on the other hand, it offers a falsification criterion for physical theories (a violation of Curie's principle may indicate that something is wrong in the physical description).

Such applications of Curie's principle depend, of course, on our accepting its truth, and this is something that has been questioned in the literature, especially in relation to spontaneous symmetry breaking. Different proposals have been offered for justifying the principle. Curie himself seems to have regarded it as a form of causality principle, and the question in the recent literature has been whether the principle can be demonstrated from premises that include a definition of "cause" and "effect". In this direction it has become current of late to understand the principle as following from the invariance properties of *deterministic* physical laws. The seminal paper for this approach is Chalmers (1970), which introduces the formulation of Curie's principle in terms of the relationship between the symmetries of earlier and later states of a system, and the laws connecting these states. This "received view" can be criticized for offering a reformulation that is significantly different from Curie's intentions (so that the label 'Curie's principle' is a misnomer), and for resting on an assumption that may undermine the interest and importance of the view, as we discuss in the following brief remarks.⁶

The received view, by concerning itself with temporally ordered cause and effect pairs (or states of systems), offers a diachronic or dynamic analysis. In fact, Curie himself focusses on synchronic or static situations, concerning the compatibility of different phenomena occurring at the same time, rather than the evolution of one state of a system into another state. In other words, the 'cause-effect' terminology used by Curie is *not* intended to indicate a temporal ordering of phenomena being considered. This is clear from his examples, and also from the fact that discussion of the laws – so central to the diachronic version – is absent from Curie's own analysis. That the diachronic version has come to have the label 'Curie's principle' therefore misrepresents Curie's original principle and his discussion of that principle.

Is the diachronic version interesting and important, nevertheless? The account can be understood as an application of PSR in which we pay careful attention to whether the laws provide a "sufficient reason" for a symmetry to be broken as a system evolves from its initial to final state by means of those laws. The reformulation of the diachronic version by Earman (2004) has the strong merit of being precise, and thereby enabling a proof that if the initial state possesses a given symmetry, and the laws deterministically preserve that symmetry, then the final state will also possess that symmetry. However, things are not so simple as they might seem because the proof takes a state with a *given* symmetry. Specifying the symmetries of a state requires, in general, recourse to a background structure – such as space or spacetime, or the space of

⁶For detailed discussion see Brading and Castellani (2006). The "received view" that we attribute first to Chalmers is developed in Ismael (1997) and Earman (2004). See also Earman (this vol., ch. 15, Section 2.3).

solutions. In some cases, the required structure may seem trivial or minimal, but nevertheless the dynamics of the system will not be independent of this structure (consider the examples of the spatial or spatiotemporal structure or, more strongly still, the space of solutions). This has the consequence that, in general, the structures on which the symmetries of a state and the symmetries of the dynamics depend are not independent of one another, and any appearance to the contrary in the “proof” needs to be handled with caution. Indeed, we think that answering the question of whether the diachronic version is interesting and important depends in part upon investigating this lack of independence and the role it plays in the proof, something which has yet to be provided in the literature on the diachronic version of ‘Curie’s principle’.

Both Curie’s original version of his principle, and the diachronic version, begin with the symmetries of states of physical systems. In contemporary physics, focus has shifted to symmetries of laws, and the significant connection between symmetries of physical systems and symmetries of laws has to do not with symmetries of *states* of those systems, but with symmetries of *ensembles of solutions*.⁷ The symmetries of a dynamical equation are not, in general, the symmetries of the individual solutions (let alone states), but rather the symmetries of the whole set of solutions, in the sense that a symmetry of a dynamical equation transforms a given solution into another solution. Considering this relationship between laws and solutions leads to an alternative version of Curie’s principle, which we propose here.⁸ As with the diachronic version of Curie’s principle, our proposal departs from Curie’s original proposal, but our contention is that it remains true to the main motivation behind Curie’s original investigation. In this version we seek to unite two things:

1. We understand Curie’s motivating question to be ‘which phenomena are physically possible?’, and his suggestion to be that we can use symmetries as a guide towards answering this question; and
2. We go beyond Curie in making use of symmetries of laws, something about which he said nothing, but which has become a central concern in contemporary physics.

Combining these two ingredients, a “modern” version of Curie’s principle would then simply state that *the symmetries of the law (equation) are to be found in the ensemble of its solutions*. This version expresses Curie’s basic idea – that “symmetry does not get lost (without a reason)” – in virtue of the fact that the symmetry of the law is to be found in the ensemble of solutions. The fact that this is how we define the relationship between symmetries and laws does not render it empty of significance with respect to Curie’s motivating

⁷By “solution” here we mean a temporally extended history of a system, the “state” of a system being a “solution at an instant”.

⁸Notice that this version does not involve the temporal evolution from cause to effect (as in the diachronic version), nor is it restricted to a state of a system at a given instant or during a certain temporal period (as in the synchronic version); rather, it concerns the structure of an ensemble of solutions, considered as a whole.

question. On the contrary, the point is that we can *use* the symmetries of the law as a guide to finding solutions, i.e. to determining which phenomena are physically possible, when not all the solutions are known. We can ask, following Curie, ‘What phenomena are possible?’, and we can use the connection between the symmetries of the law and the symmetries of the ensemble of solutions as a guide to finding the physically possible phenomena. Thus, what is on the one hand a definitional statement (that the symmetries of the law (equation) are to be found in the ensemble of the solutions) comes on the other hand to have epistemic bite when we don’t know all the solutions. This, we believe, is true to Curie’s motivating question, as expressed in item (1), above.

3 Symmetry and group theory: early history

Group theory is the powerful mathematical tool by means of which the symmetry properties of theories are studied. In this section, we begin with the definition of a group, and outline the origins of this notion in the mathematics of algebraic equations. We then turn our attention to the manner in which group theory was applied first to geometry and then to physics in the course of the nineteenth century.

3.1 The introduction of the group concept and the first developments of group theory

A *group* is a family \mathcal{G} of elements $g_1, g_2, g_3 \dots$ for which there is defined a multiplication that assigns to every two elements g_i and g_j of the group a third element (their product) $g_k = g_i g_j \in \mathcal{G}$, in such a way that the following requirements hold:⁹

- $(g_i g_j) g_k = g_i (g_j g_k)$ for all $g_i, g_j, g_k \in \mathcal{G}$ (associativity of the product);
- there exists an *identity element* $e \in \mathcal{G}$ such that $g_i e = e g_i$ for all $g_i \in \mathcal{G}$;
- for all $g_i \in \mathcal{G}$ there exists an *inverse element* $g_i^{-1} \in \mathcal{G}$ such that $g_i g_i^{-1} = g_i^{-1} g_i = e$.

The concept of a group was introduced by Évariste Galois in the short time he was able to contribute to mathematics (born in 1811, he died as a result of a duel in 1832) in connection with the question of the resolution of equations by radicals.¹⁰ The resolvent formulas for cubic and quartic equations were

⁹The concept of a group can be weakened by relaxing these conditions (for example, dropping the inverse requirement leads to the concept of a monoid, and retaining only associativity leads to the concept of a semigroup). The question that then arises is whether the full group structure, or some weaker structure, is related to the symmetry properties of a given theory.

¹⁰That is, in terms of a finite number of algebraic operations – addition, subtraction, multiplication, division, raising to a power and extracting roots – on the coefficients of the equations

found by the mathematicians of the Renaissance,¹¹ while the existence of a formula for solving the general equations of the fifth and higher degrees by radicals remained an open question for a long time, stimulating developments in algebra. In particular, the studies of the second half of the eighteenth century focussed on the role played in the solution of equations by functions invariant under permutation of the roots, so giving rise to the theory of permutations. In J. L. Lagrange's *Réflexions sur la résolution algébrique des équations*, the most influential text on the subject, some fundamental results of permutation theory were obtained.¹² Lagrange's text served as a basis for successive algebraic developments, from P. Ruffini's first proof in 1799 of the impossibility of solving n th degree equations in radicals for $n \geq 5$, to the seminal works by A. L. Cauchy, N. H. Abel, and finally Galois.¹³

Galois's works¹⁴ marked a turning point, providing answers to the open questions in the solutions of equations by using new methods and algebraic notions, first of all the notion of a *group*. This notion was introduced by Galois in relation to the properties of the set of permutations of the roots of equations (the permutations constituting what he named a 'group'), together with other basic notions of group theory such as "subgroup", "normal subgroup", and "simple group".¹⁵ By characterizing an equation in terms of its "degree of symmetry", determined by the permutation group of the roots preserving their algebraic relations (later known as the *Galois group* of the equation), Galois could transform the problem of the resolution of equations into that of studying the properties of the permutation groups involved. In this way he obtained, among other things, the necessary and sufficient conditions for solving equations by radicals.

Galois's achievements in group theory, first brought to publication by Joseph Liouville in 1846, were collected and expanded in Camille Jordan's 1870 *Traité des substitutions et des équations algébriques*. Jordan's *Treatise*, the first systematic textbook on group theory, had a decisive influence on the application of this new theory, including its application to other domains of mathematical science, such as geometry and mathematical physics.

¹¹The resolvent formula for a quadratic equation was known since Babylonian times. A historical survey on the question of the existence of resolvent formulas for algebraic equations is in Yaglom (1988, pp. 3 f.).

¹²Among other results, the so-called Lagrange's theorem which states – in modern terminology – that the order of a subgroup of a finite group is a divisor of the order of the group.

¹³Cauchy (1789–1857) generalized Ruffini's results in 1815; Abel (1802–1829) published in 1824 a proof of the impossibility of solving the quintic equation by radicals and in 1826 the paper *Démonstrations de l'impossibilité de la résolution algébrique des équations générales qui passent le quatrième degré*.

¹⁴A few "m'emoires" submitted to the Académie des Sciences, three brief papers published in 1830 in the Académie's 'Bulletin', and some letters, among which is the last one written to his friend Auguste Chevalier in the night before the fatal duel.

¹⁵See Yaglom (1988, pp. 9 f.), for details.

3.2 Applications of group theory: the contributions of Klein and Lie

3.2.1 Projective geometry, the theory of invariants and group theory: Klein and Lie's starting point

In the same year as the appearance of Jordan's *Treatise*, Sophus Lie and Felix Klein, two young mathematicians who were to become the key figures in extending the domain of application of group theory, moved for a period from Berlin to Paris to enter into contact with the French school of mathematics. Lie and Klein had just written a joint paper investigating the properties of some curves in terms of the groups of projective transformations leaving them invariant. In fact they were drawn to Paris mostly by their interest in *projective geometry*, the science founded by J. V. Poncelet to study the properties of figures preserved under central projections. Projective geometry had become, at the time, a particularly fruitful research field for the combination of algebraic and geometrical methods based on the notion of invariance. The *theory of invariants* was itself a flourishing branch of mathematics, centered on the systematic study of the invariants of "algebraic forms". Using the theory of invariants, the English algebraist Arthur Cayley¹⁶ had recently clarified the relationship between Euclidean and projective geometry, showing the former to be a special case of the latter. Before leaving for France, Klein had tried to extend Cayley's results, based on the possibility of defining a distance (a "metric") in terms of a quadratic form defined on the projective space, to the case of non-Euclidean geometries.

While in Paris Lie and Klein became acquainted not only with Jordan, but also with the expert in differential geometry Gaston Darboux, who stimulated their interest in the relations between differential geometry and projective geometry.

3.2.2 Klein's Erlangen Program

The question of the relations between the different contemporary geometrical systems particularly interested Klein. He aimed at obtaining a unifying foundational principle for the various branches into which geometry had apparently recently separated. In this respect, he fruitfully combined (a) the application of the theory of invariants to the study of geometrical properties, with (b) his and Lie's idea of applying algebraic group theory to treating also geometrical transformations. The new group theoretical conception of a geometrical theory which resulted was announced in his famous *Erlangen Program*, as it became known following the inaugural lecture entitled 'Comparative Considerations on Recent Geometrical Research'¹⁷ that the 23-year-old Klein delivered when entering, in 1872, as a professor on the staff of the University of Erlangen. Guided by the idea that geometry is in the end a unity, Klein's solution to the problem posed

¹⁶Cayley was one of the three members of the 'invariant trio', as the French mathematician Hermite dubbed them, the other two being James Sylvester, inventor of most of the terminology of the theory including the word 'invariant', and George Salmon.

¹⁷'Vergleichende Betrachtungen über neuere geometrische Vorschungen'.

by the existence of different geometries was to propose a general characterization of a geometrical theory by using the notion of invariance under a transformation group (i.e., the notion of symmetry). According to his characterization a geometry is defined, with respect to a given domain (the plane, the space, or a given “manifold”) and a group of transformations acting on it, as the science studying the invariants under the transformations of the group. Each specific geometry is thus determined by the characterizing symmetry group (for example, planar Euclidean geometry is determined by the group of affine transformations acting on the plane), and the interrelations between geometries can be described by the relations between the corresponding groups (for example, the equivalence of two geometries amounts to the isomorphism between the corresponding groups).

With Klein’s definition of a geometry, geometrical and symmetrical properties become very close: the symmetry of a figure, which is defined in a given “space”¹⁸ the “geometrical” properties of which are preserved by the transformations of a group \mathcal{G} , is determined by the subgroup of \mathcal{G} leaving the figure invariant. The new group theoretical techniques prompted a transition from an inductive approach (familiar from the nineteenth century classifications of crystalline forms in terms of their visible – and striking – symmetry properties)¹⁹ to a more abstract and deductive approach. This is the procedure formulated in Weyl’s classic book on symmetry (*Symmetry*, 1952) as follows:

Whenever you have to do with a structure-endowed entity Σ try to determine its group of automorphisms, the group of those element-wise transformations which leave all structural relations undisturbed. [...] After that you may start to investigate symmetric configurations of elements, i.e. configurations which are invariant under a certain subgroup of the group of all automorphisms (Weyl, 1952, p. 144, emphasis in original).

In this way, the symmetry classifications could be extended to figures in “spaces” different from the plane and space of common experience.

Klein himself contributed to the classification of symmetry groups of figures with his works on discrete groups; in particular, he studied the transformation groups related to the symmetry properties of regular polyhedra, which proved to be useful in the solution of algebraic equations by radicals.

3.2.3 Lie’s theory of continuous groups

After 1872, while Klein was concerned especially with *discrete* transformation groups, Lie devoted all his research work to building the *theory of continuous transformation groups*, the results of which were systematically collected in his three-volume *Theorie der Transformationsgruppen* (I: 1888, II: 1890, and III: 1893), written with the collaboration of F. Engel. Lie’s interest in continuous groups arose in relation to the theory of differential equations, which he took

¹⁸A set of points endowed with a structure.

¹⁹A classic textbook in this respect is Shubnikov–Koptsik (1974). See also Section ??, below.

to be ‘the most important discipline in modern mathematics’. By the time he was in Paris, Lie had begun to study the theory of first-order partial differential equations, a theory of particular interest because of the central role it played in the formulation given by W. R. Hamilton and C. G. Jacobi to mechanics.²⁰ His project was to extend to the case of differential equations Galois’s method for solving algebraic equations: that is, using the knowledge of the ‘Galois group’ of an equation (the symmetry group formed by the transformations taking solutions into solutions) so as to solve it or reduce it to a simpler equation. Thus Lie’s guiding idea was that continuous transformation groups could, in the solution of differential equations, play a role analogous to that of the permutation groups used by Galois in the case of algebraic equations.

Lie had already considered continuous groups of transformations in some earlier geometrical works. In his studies with Klein on special kinds of curves (called by them ‘W-curves’), he had examined transformations that were continuously related in the sense that they were all generated by repeating an *infinitesimal transformation*.²¹ The relevance of infinitesimal transformations to continuous groups of transformations was to become a central point in his studies of *contact transformations*, so called because they preserved the contact or tangency of surfaces. Lie had started to investigate contact transformations in association with geometrical reciprocities implied in his “line-to-sphere mapping”, a mapping between a line geometry and a sphere geometry that he had discovered while in Paris.²² When he turned to considering first-order partial differential equations, he soon realized that they admitted contact transformations as symmetry transformations (i.e. transformations taking solutions into solutions). Thus contact transformations could form the “Galois group” of first-order partial differential equations. This motivated him to develop the invariant theory of contact transformations, which represented the first step of his general theory of continuous groups.

Lie’s crucial result, allowing him to pursue his program, was the discovery that to each continuous transformation group could be assigned what is today called its *Lie algebra*. Lie showed that the infinitesimal generators of a continuous transformation group obey a linearized version of the group law, involving the commutator bracket (or Lie bracket); this linearized law then represents the structure of the algebra. In short (and in modern terminology): we describe the elements (transformations) of a continuous group (now called a *Lie group*)²³ as functions of a certain number r of continuous parameters a_l ($l = 1, 2, \dots, r$). And these group elements can be written in terms of a corresponding number

²⁰For details on classical mechanics we refer the reader to Butterfield (this vol., ch. 1) and the references therein.

²¹See on this part Hawkins (2000, Section 1.2). According to Hawkins (p. 15), with the works of Lie and Klein on W-curves ‘for the first time not only is a continuous group the starting point for an investigation, but also for the first time in print we have the idea that infinitesimal transformations are a characteristic and useful feature of continuous systems of transformations’.

²²See Hawkins (2000, Section 1.4).

²³For a precise definition of this and other terms in this paragraph, see Butterfield (this vol., ch. 1, Section 3) and Harvey (this vol., ch. 14).

r of infinitesimal operators X_l , the *generators* of the group, which satisfy the “multiplication law” represented by the Lie brackets

$$[X_s, X_t] = c_{st}^q X_q ,$$

so forming what is called the *Lie algebra* of the group. The coefficients c_{st}^q are constants characterizing the structure of the group and are called the *structure constants* of the group.²⁴

Thanks to this sort of result, the study and classification of continuous groups could be conducted in terms of the corresponding Lie algebras. This proved to be extremely fruitful in the successive developments, not only algebraic and geometrical, but also physical. With regard first of all to the physics of Lie’s time, Lie had arrived at the correspondence between continuous groups and Lie algebras by reinterpreting, in the light of his program for solving differential equations, the results obtained by Poisson and especially Jacobi about the integration of first-order partial differential equations arising in mechanics.²⁵ His achievements were thus of great relevance to the solution of the dynamical problems discussed by his contemporaries.

But it is in twentieth century physics, with the works of such figures as Hermann Weyl, Emmy Noether and Eugene Wigner (just to recall the central figures who first contributed to the applications of Lie’s theory to modern physics), that the theory of Lie groups and Lie algebras acquired a fundamental role in the description of physical phenomena. Today, the applications of the theory that originated from Lie’s works include the whole of theoretical physics, of both the large and the small: classical and quantum mechanics, relativity theories, quantum field theory, and string theory.²⁶

4 What are symmetries in physics? Definitions and varieties

4.1 What is meant by ‘symmetry’ in physics

We can understand intuitively the generalization of the scientific notion of symmetry from physical or geometric objects to laws, as follows. We write down our law as a mathematical equation, and appearing in this equation will be various mathematical objects and operators. For a particular group of transformations, these objects and operators transform according to rules that may be fixed either by the mathematical nature of the object or operator concerned, or (where the mathematics does not fix the transformation rules) by our specification. If the “form” of the equation is preserved when we transform each of the objects

²⁴For more details, see Butterfield (this vol., ch. 1, Sections 3.2 and 3.4).

²⁵For details see Hawkins (2000, Section 2.5).

²⁶In this volume, see especially t’Hooft (this vol., ch. 7), Harvey (this vol., ch. 14) and Belot (this vol., ch. 2).

and operators appearing in our equation by any element of the group, then we say that the group is a *symmetry group* of the equation.

More precisely, what we mean by the symmetry transformations of the laws in physics can be formulated in either of the following ways, which are equivalent in the sense that they pick out the same set of transformations:

- (1) Transformations, applied to the independent and dependent variables of the theory in question, that leave the form of the laws unchanged.
- (2) Transformations that map solutions into solutions.

Symmetry transformations may be viewed either *actively* or *passively*. From the passive point of view we re-describe *the same physical evolution* in two different coordinate systems.²⁷ That is, we transform the independent and dependent variables, as in (1). If the description in the original set of coordinates is a solution of given equations, then the new description in the new set of coordinates is a solution *of the same equations*. (If the transformation is *not* a symmetry transformation, then the new description in the new coordinates will not, in general, be a solution of the same equations, but rather of *different* equations.) The mapping of one solution into another solution of the same equations, by means of a symmetry transformation, leads to the *active* interpretation of such transformations. On this interpretation, the two solutions are viewed as *different physical evolutions* described in *the same coordinate system*. Thus, formulation (2) lends itself naturally to an active interpretation.

The ‘form of the law’ in (1) means the functional form of the law, expressed in terms of the independent and dependent variables. A transformation of those variables will, in general, lead to an expression whose functional form differs from that of the original expression (x goes to x^2 , for example). At this point it will be helpful to say a few words about “invariance” and “covariance”. Let the reader beware that there is no unanimity over how these terms are used in discussing the laws of physics, especially in the philosophy of physics literature. Often, the term ‘invariant’ is reserved for objects, and ‘covariant’ is used for equations or laws. However, this is a product of a more fundamental distinction, which when understood correctly allows for the application of the notion of invariance to laws as well.

We think that the discussion of Ohanian and Ruffini (1994, Section 7.1) is very useful, and that it nicely distils much of the best of what can be found in the literature, both in physics and in philosophy of physics. The upshot is as follows. We may say that an equation is *covariant* under a given transformation when its form is left unchanged by that transformation. This is the notion at work in Definition 1. In a way, it is rather weak: given an equation that is not covariant under a given transformation, we can always re-write it so that it becomes covariant. On the other hand, this re-writing may involve the introduction of new functions of the variables, and it is the physical interpretation of these new

²⁷By ‘coordinates’ here we are referring to generalized coordinates; in general, one coordinate for each degree of freedom of the system.

quantities that allows covariance to gain physical significance. We will have more to say about this for the specific case of *general covariance* and Einstein’s General Theory of Relativity in Section ?? below.

Invariance of an equation, as characterized by Ohanian and Ruffini, is a stronger requirement than covariance. Not only should the form of the equation remain the same, but so too should the values of any non-dynamical quantities, including “constants” such as the speed of light. By “non-dynamical quantities” we mean all those objects which appear in the equations yet which do not themselves satisfy equations of motion. We here enter the muddy waters of how to distinguish between “absolute” and “dynamical” objects, as discussed by Anderson (1967).²⁸

In both cases (covariance and invariance), the associated transformations – when actively construed – take solutions into solutions. When using formulation (2), it is important to be clear about what is meant by a solution. This does *not* mean a solution-at-an-instant, i.e. an instantaneous state of a system; rather, it means an entire history, i.e. possible time-evolution, of the system in question.²⁹

4.2 Varieties of symmetry

Symmetries in physics come in a number of different varieties, distinguished by such terms as ‘global’ and ‘local’; ‘internal’ and ‘external’; ‘continuous’ and ‘discrete’. In this Section we briefly review this terminology and the associated distinctions.

The most familiar are the global spacetime symmetries, such as the Galilean invariance of Newtonian mechanics, and the Lorentz invariance of the Special Theory of Relativity. Global spacetime symmetries are intended to be valid for all the laws of nature, for all the processes that unfold in the spacetime. Symmetries with this universal character were labelled ‘geometric’ by Wigner (see 1967, especially p. 17).

This universal character is not shared by some of the symmetries introduced into physics during the twentieth century. Most of these were of an entirely new kind, with no roots in the history of science, and in some cases expressly introduced to describe specific forms of interactions – whence the name ‘dynamical symmetries’ due to Wigner (1967, see especially pp. 15, 17–18, 22–27, 33).

The various symmetries of modern physics can also be classified according to a second distinction: that between global and local symmetries. The terms ‘global’ and ‘local’ are used in physics, and in philosophy of physics, with a variety of meanings. The distinction intended here is between symmetries that depend on constant parameters (global symmetries) and symmetries that depend on arbitrary smooth functions of space and time (local symmetries). While Lorentz invariance is an example of a global symmetry, the gauge symmetry of

²⁸See also Section ??, below. One difficulty in tackling the literature on this issue is the variety of uses and meanings attaching to the common terminology of covariance, principle of covariance, invariance, absolute and dynamical objects, and so forth.

²⁹The distinction is important in, for example, our discussion of Curie’s principle, Section ?? above.

classical electromagnetism (an internal symmetry)³⁰ and the diffeomorphism invariance in General Relativity (a spacetime symmetry) are examples of local symmetries, since they are parameterized by arbitrary functions of space and time.³¹ Recalling Wigner's distinction, Lorentz invariance is a geometric symmetry, applying to all interactions, whereas the gauge symmetry of electromagnetism concerns the electromagnetic interaction specifically and is therefore a dynamical symmetry.

The gauge symmetry of classical electromagnetism is an internal symmetry because the transformations of the vector potential occur in the internal space of the field system, rather than in spacetime. The gauge symmetry of classical electromagnetism can seem to be no more than a mathematical curiosity, specific to this theory; but with the advent of quantum theory the use of internal degrees of freedom, and the related internal symmetries, became fundamental.³²

The translations, rotations and boosts of the inhomogeneous Lorentz group are all examples of *continuous symmetries*, for which any finite symmetry transformation can be built up of infinitesimal symmetry transformations. In contrast with the continuous symmetries we have the *discrete symmetries* of charge conjugation, parity, and time reversal (CPT), along with permutation invariance. Thus, Newtonian mechanics and classical electrodynamics are invariant under parity (left-right inversion) and under time reversal (roughly: the laws hold for a sequence of states evolved in the backwards time direction just as they hold for the states ordered in the forwards direction). Classical electrodynamics is also invariant under charge conjugation, so long as we correctly implement the associated transformations of the electric and magnetic fields. Finally, there is a sense in which classical statistical mechanics is permutation invariant: the particles postulated are identical to one another, and their permutation takes a solution into a solution. However, the power and significance of the discrete symmetries achieves its full force only in quantum theory, and for discussion of this we refer the reader to Harvey (this vol., ch. 14).

In Section ?? below, we discuss some of the interpretative issues associated with these different varieties of symmetry in classical physics.

³⁰For more on gauge and internal symmetries, see the following paragraph.

³¹We discuss the local symmetry of General Relativity further in Section ?? below. See also Belot (this vol., ch. 2).

³²For interpretative issues associated with gauge symmetry in classical electromagnetism, see Belot (1998). Gauge symmetries came to prominence with the development of quantum theory. For symmetries in quantum theory see Harvey (this vol., ch. 14). The term 'gauge symmetry' itself stems from Weyl's 1918 theory of gravitation and electromagnetism. For discussions of all these aspects of gauge symmetry, see Brading and Castellani, 2003.

5 Some applications of symmetries in classical physics

5.1 Transformation theory in classical mechanics

As we have seen, Lie’s interest in continuous groups arose in relation to his studies of the theory of first-order partial differential equations, which played a central role in the formulation given by Hamilton and Jacobi to mechanics. The *transformation theory of mechanics* based on this formulation is indeed one of the first examples of a systematic exploitation in physics of the invariance properties of dynamical equations. These symmetries are exploited according to the following strategy: the integration of the equations of motion is simplified by transforming – by means of symmetry transformations – the original dynamical system into another system with fewer degrees of freedom.

Historically, the road to the possibility of applying the above ‘transformation strategy’ to solving dynamical problems was opened by the works of J. L. Lagrange and L. Euler. The Euler-Lagrange analytical formulation of mechanics, grounded in the seminal *Mécanique Analytique* (1788) of Lagrange, expressed the laws of motion in a form which was covariant (cf. Section ??) under all coordinate transformations. This meant one could more easily choose coordinates to suit the dynamical problem concerned. In particular, one hoped to find a coordinate system containing “cyclic” (a.k.a. “ignorable”) coordinates. The presence of ignorable coordinates amounts to a partial integration of the equations: if all the coordinates are ignorable, the problem is completely and trivially solved. The method was thus to try to find (by applying coordinate transformations leaving the dynamics unchanged) more and more ignorable variables, thus transforming the problem of integrating the equations of motion into a problem of finding suitable coordinate transformations.

The successive developments in the analytical approach to mechanics, from Hamilton’s “canonical equations of motion” to the general transformation theory of these equations (the theory of canonical transformations) obtained by Jacobi, presented many advantages of the “transformation strategy” point of view. For further details we refer the reader to Butterfield’s chapter of this volume, along with classic references such as Lanczos ([1949], 1962) and Whittaker ([1904], 1989). Butterfield (this vol., ch. 1), by expounding the theory of symplectic reduction in classical mechanics, thoroughly illustrates the strategy of simplifying a mechanical problem by exploiting a symmetry. This strategy is also the main subject of Butterfield (2006), focussing on how symmetries yield conserved quantities according to Noether’s first theorem (see Section ??, below), and thereby reduce the number of variables that need to be considered in solving a problem.

We end these brief remarks on symmetry and transformation theory in classical mechanics by emphasizing two points.

First, we note that a problem-solving strategy according to which a dynamical problem (equation) is transformed into another equivalent problem (equa-

tion) by means of a symmetry might be seen as an example of the application of Curie's principle in its modern version (see here Section ??): by transforming an equation into another equivalent equation using a specific symmetry we may arrive at an equation which we can solve; the solution of the new equation is related to the unknown solution of the old equation by the specific symmetry; that is, we thereby arrive at an equivalent solution.

Second, we emphasize that in all these developments the invariance properties of the dynamical equations, though undoubtedly important, were considered exclusively in an instrumental way. That is, canonical transformations were studied only for the purpose of solving the dynamical problem at hand. The equations were given, and their invariance properties were investigated to help find their solutions. The formulation of Einstein's Special Theory of Relativity at the beginning of the twentieth century brings an inversion of this way of thinking about the relationship between symmetries and physical laws, as we shall see in the following section.

5.2 Symmetry principles as guides to theory construction

The principle of relativity, as expressed by Einstein in his 1905 paper announcing the Special Theory of Relativity, asserts that

The laws by which the states of physical systems undergo changes are independent of whether these changes are referred to one or the other of two coordinate systems moving relatively to each other in uniform motion.³³

It further turns out that these coordinate systems are to be *inertial* coordinate systems, related to one another by the Lorentz transformations comprising the inhomogeneous Lorentz group.

The principle of relativity thus stated meets the conditions listed above in Section ?? for a *symmetry* principle:

- The Lorentz transformations, applied to the independent and dependent variables of the theory, leave the form of the laws as stated in one inertial system unchanged on transformation to another inertial coordinate system.
- The Lorentz transformations map a solution, given relative to an inertial coordinate system, into another solution.

This principle was explicitly used by Einstein as a guide to theory construction: it is a principle that must be satisfied whatever the final details of the theory.³⁴ Indeed, using just the principle of relativity and the light postulate, Einstein derives various results, including the Lorentz transformations.

³³Miller's (1981) translation, p. 395.

³⁴For discussion of the principle/constructive theory distinction in Einstein, see Brown (2006, ch. 6) and Howard (2007).

As noted above, this represents a reversal in the priority that, since the time of Newton, had been given to the relativity principle versus the dynamical laws. Huygens used the relativity principle as a basic postulate from which to derive dynamical results, but in Newton the relativity principle, initially presented in his manuscripts as an independent postulate, is relegated in the *Principia* to a corollary.³⁵ From then until Einstein, the relativity of inertial motion is seen as a consequence of the particular laws under consideration, and something that could turn out to be false once the details of the laws of some particular interaction are known. Similarly for classical physics in general, symmetries – such as spatial translations and rotations – were viewed as properties of the laws that hold as a consequence of those particular laws. With Einstein that changed: symmetries could be postulated prior to details of the laws being known, and used to place restrictions on what laws might be postulated. Thus, symmetries acquired a new status, being postulated independently of the details of the laws, and as a result having strong heuristic power. As Wigner wrote, Einstein’s papers on special relativity ‘mark the reversal of a trend’: after Einstein’s works, ‘it is now natural to try to derive the laws of nature and to test their validity by means of the laws of invariance, rather than to derive the laws of invariance from what we believe to be the laws of nature’ (Wigner, 1967, p. 5).

The methodology that had served Einstein well with the Special Theory of Relativity (STR) also had a role in his development of the General Theory (GTR), for which he used various different principles as restrictions on the possible form that the eventual theory might take.³⁶ One of these was, so Einstein maintained, an extension of the principle of relativity found in STR to include coordinate systems that are in accelerated motion relative to one another, implemented by means of the requirement that the equations of his new theory be generally covariant. Einstein was seeking a “Machian” solution to the challenge of Newton’s bucket, which he took to require that there be no preferred reference frames. Thus, in his 1916 review article Einstein wrote that ‘*The laws of physics must be of such a nature that they apply to systems of reference in any kind of motion.* Along this road we arrive at an extension of the postulate of relativity’ (emphasis in original).

The questions of whether or not the principle of general covariance (a) makes any arbitrary smooth coordinate transformation into a symmetry transformation, and (b) is a generalization of the principle of relativity, have been much discussed. The answer to (b) is a definitive ‘no’, but there is less consensus at present about the answer to (a).³⁷ In the following section we take up discussion of (a). Here we close with a few brief remarks concerning (b).

³⁵In fact, it does not follow from Newton’s three laws of motion – we must further assume the velocity independence of mass and force. See Barbour (1989, Section 1.2).

³⁶Primarily the following: the principle of relativity, later (in 1918) distinguished from what Einstein referred to as ‘Mach’s principle’; the principle of equivalence; and the principle of conservation of energy–momentum.

³⁷See for example Torretti (1983, pp. 152–4); Norton (1993), who also discusses the relationship with the principle of equivalence; Anderson (1967).

Even if general covariance in GTR is a symmetry principle, it is not an extension of the relativity principle. That is to say, general covariance says nothing about the observational equivalence of distinct reference frames.³⁸ As already noted, the thought that general covariance might provide such a principle was, for Einstein, connected with his attempts to provide a “Machian” resolution to the challenge of Newton’s bucket, and with his principle of equivalence. However, the principle of equivalence *does not* imply the observational equivalence of reference frames in arbitrary states of motion (Einstein never thought that it did), and Einstein eventually realized that GTR does not vindicate a solution to Newton’s bucket that depends only on the relative motion of *matter*.³⁹

Whatever the subtleties of whether, and to what extent, general covariance is a symmetry principle, it is clear that it had enormous heuristic power, not just for Einstein in his development of GTR, but also beyond. Think for example Hilbert’s work on the axiomatization of physics (see Corry, 2004, and references therein), and Weyl’s attempts to construct a unified field theory (see O’Raifeartaigh, 1997, for an English translation of Weyl’s 1918 paper ‘Gravitation and Electricity’, and see also Weyl, 1922). In all these cases, general covariance provided a powerful tool for theory construction. In the following Section we discuss further the significance and interpretation of general covariance in GTR.⁴⁰

6 General covariance in General Relativity

In the preceding Section we noted the role of the principle of general covariance as a guide to theory construction. In this Section we turn our attention to a number of further issues relating to general covariance in GTR that have received attention in the philosophical literature. We begin with the issue, raised in the preceding section, concerning the status of arbitrary smooth coordinate transformations as symmetry transformations. We then discuss various characteristics associated with general covariance, including those pointed to by Einstein’s so-called ‘hole argument’, before turning to the issue of whether or not general covariance has physical content.⁴¹ We postpone discussion of Noether’s theorems to Section ??, below.

6.1 General covariance and arbitrary coordinate transformations as symmetry transformations

Does the principle of general covariance make any arbitrary smooth coordinate transformation into a symmetry transformation? One way to approach this

³⁸For further discussion see, for example, Norton (1993) and Torretti (1983, Section 5.5).

³⁹For a clear and concise discussion, see Janssen (2005).

⁴⁰For detailed presentation of the Special and General Theories of Relativity, see Malament (this vol., ch. 3). See also Rovelli (this vol., ch 12, Section 4).

⁴¹See also Belot (this vol., ch. 2).

question is to consider active rather than passive transformations (see Section 4.1, above), and to compare the situation in GTR with that in STR.

In STR, a Lorentz transformation – actively construed – picks up the matter fields and redistributes them with respect to the spacetime structure encoded in the metric. The principle of relativity holds for such transformations because the evolution of the matter fields in the two cases (related by the Lorentz transformation) are observationally indistinguishable: no observations, in practice or in principle, could distinguish between the two scenarios. In GTR, active general covariance is implemented by active diffeomorphisms on the spacetime manifold (see Rovelli, this vol., ch. 12, Section 4.1). These involve transformations of not just the matter fields, but also the metric field, in which both are redistributed with respect to the spacetime manifold. Once again, the “two cases” are observationally “indistinguishable”, but this time the reason generally given is that the “two cases” are in fact just one case.⁴²

Why should we accept that there are two genuinely distinct cases when considering the Lorentz transformations in STR, and only one case for the diffeomorphisms of GTR? One approach would be to claim that a crucial difference between the two is that a Lorentz transformation can be implemented on an effectively isolated *sub-system* of the matter fields, producing an observably distinct scenario in which, nevertheless, the evolution of the sub-system in question is indistinguishable assuming no reference is made to matter fields outside that subsystem. For example, in Galileo’s famous ship experiment we consider two observably distinct scenarios – one in which the ship is at rest with respect to the shore, and one in which it moves uniformly with respect to the shore – and we notice that the behaviour of physical systems within the cabin of the ship does not distinguish between the two scenarios.⁴³ No analogue of the Galilean ship experiment can be generated for the general covariance of GTR.⁴⁴

The importance of symmetry transformations being implementable to produce observationally distinct scenarios has been emphasized by Kosso (2000). On this view, the observational significance of symmetry transformations rests on a combination of *two* observations being possible in principle. First, it must be possible to confirm empirically the implementation of the transformation – hence the importance of being able to generate an observationally distinguishable scenario through the transformation of a subsystem. Second, we must be able to observe that the subsequent internal evolution of the subsystem is unaffected. That we cannot meet the first of these requirements for arbitrary smooth

⁴²See also Section ??, penultimate paragraph.

⁴³This implementation can be only approximate, relying on the degree to which the sub-system in question can be isolated from the “external” matter fields.

⁴⁴One suggestion might be that we perform a transformation T which is the identity outside some region R , and which differs from the identity within that region. This will not achieve the desired result. The two scenarios must have observationally distinct consequences, at least in principle. In the case of Galileo’s ship, if we allow the subsystem to interact with other matter once again, we will see that in one case the ship crashes into rocks (for example), while in the other it suffers no such collision. Thus, we have observational distinguishability in principle. The transformation T does not produce a scenario which any future events could enable us to distinguish from the original.

coordinate transformations in GTR marks a difference between these and the Lorentz transformations.⁴⁵

On this approach, while the field equations of GTR take the same form for any choice of coordinate system, this is not sufficient for arbitrary coordinate transformations to be symmetries. In addition, the actively construed transformations must have a physical interpretation – we must be transforming one thing with respect to something else. When we perform a diffeomorphism, we get back the same solution, not a new solution, for we are not re-arranging the matter fields with respect to the metric.

We stress that this is only one way to approach the issue of whether general covariance should be understood as a symmetry principle in GTR. A contrasting position may be found in Anderson (1967, Section 10-3), who argues that we must understand Einstein as viewing general covariance as a symmetry requirement, and attempts to spell out the conditions under which it can function as such.

6.2 Characteristics of generally covariant theories

Any generally covariant theory will possess certain characteristics that are philosophically noteworthy. First, there will be a *prima facie* problem with causality and determinism within the theory, and second, there will be constraints on the specification of the initial data. Einstein recognized aspects of the first characteristic while he was searching for his theory of gravitation, maintaining from 1913 through until the fall of 1915 that his so-called ‘hole argument’ provided grounds for concluding that no generally covariant theory could be physically acceptable.

In the ‘hole argument’, Einstein considers a region of spacetime in which there are no matter fields (the “hole”), and then shows that in a generally covariant theory no amount of data about the values of the matter and gravitational fields outside the hole is sufficient to uniquely determine the values of the gravitational field inside the hole. From this, Einstein concluded that no generally covariant theory could be physically acceptable.⁴⁶

The context to bear in mind here is that Einstein was searching for a theory in which the matter fields plus the field equations would uniquely determine the metric.⁴⁷ In the summer of 1915 Einstein lectured on relativity theory in Göttingen where his audience included David Hilbert. If we assume that Einstein’s presentation included a version of his ‘hole argument’, then we can reasonably infer that Hilbert was quick to reinterpret the issue that the ‘hole argument’ points to, and to present the problem raised for generally covariant theories in terms of whether such theories permit well-posed Cauchy problems.⁴⁸

⁴⁵Indeed, this result applies generally to local versus global symmetries. See also Brading and Brown (2004).

⁴⁶For presentation and discussion of the ‘hole argument’, see Norton (1984, pp. 286–291, and 1993, Sections 1-3), Stachel (1993), and Ryckman (2005, Section 2.2.2). See also Rovelli, this vol., ch. 12, Section 4.1.1

⁴⁷For more on Einstein’s (mis)appropriation of Mach’s principle, see Barbour (2005).

⁴⁸Brading and Ryckman (2007); see also Brown and Brading (2002, especially Section IV).

In the years immediately following the advent of GTR, Hilbert played a central role in spelling out the problems of causality and determinism faced by any generally covariant theory. He pointed out that in any such theory, including GTR, there will be four fewer field equations than there are variables, leading to a mathematical underdetermination in the theory. As Hilbert stressed, the Cauchy problem is not well-posed: given a specification of initial data, the field equations do not determine a unique evolution of the variables.

We can see the connection between the underdetermination problem and general covariance as follows. For the Cauchy problem to be well-posed, we must be able to express the second time derivatives of the metric in terms of the initial data (plus the further spatial derivatives that can be calculated from the initial data). However, if we re-express the 10 (source-free) Einstein field equations $G_{\mu\nu} = 0$ so as to explicitly display all the terms containing the second time derivative of the metric, we see that we have ten equations for six unknowns $g_{ij,00}$, the remaining four second time derivatives $g_{\mu 0,00}$ failing to appear in the equations.⁴⁹ This is a direct consequence of general covariance: we can always make a coordinate transformation in the neighborhood of the initial data surface such that the metric components and their first derivatives are unchanged, while the second time derivatives $g_{\mu 0,00}$ vanish on that surface. Thus the field equations, which must be valid in all coordinate systems, cannot possibly contain information on the second time derivatives. The initial data do not determine the metric uniquely: there are four arbitrary functions $g_{\mu 0,00}$ that we are free to choose.

Today, it is customary to assert from the outset that solutions of Einstein's field equations differing only in the choice of these four arbitrary functions are physically equivalent.⁵⁰ But here we should note that this "gauge freedom" interpretation of general covariance leads to problems of its own.⁵¹ For example, within this framework the observables of the theory must be "gauge invariant" quantities, but such quantities have (to date) turned out to be far removed from anything "observable" in the operational sense. The gauge freedom interpretation of general covariance is sometimes accompanied by the view that this freedom – and therefore general covariance itself – lacks physical content. We turn to consider this issue in Section ??, below.

In our explanation of the underdetermination problem, above, we noted that the Einstein equations provide ten equations for the six unknowns $g_{ij,00}$. The other face of the underdetermination problem is therefore an *overdetermination* problem with respect to the $g_{ij,00}$, and what this means is that there will be constraints on the specification of the data on the initial hypersurface. This is the second characteristic of all generally covariant theories that we mentioned in our opening remarks of the current subsection. Indeed, the presence of constraint equations is a feature shared with other theories with a local symmetry structure, such as electromagnetism. Philosophically, the significance lies in the

⁴⁹See Adler, Bazin and Schiffer (1975, ch. 8) for details of the over- and under-determination issues.

⁵⁰Recall the discussion of Section ??, above.

⁵¹See Belot (this vol., ch. 2).

relationship between the theory and the initial data. In the seventeenth century Descartes wrote a story of a world created in a state of disorder from which, by the ordinary operation of the laws of nature, a world seemingly similar to our own emerged.⁵² This image of the world emerging from an initial chaos has a long history, of course, but the emergence of order by means of the operation of the laws of nature offered a novel twist to the tale. It involves the separation of initial conditions, which could be anything, from the subsequent law-governed evolution of the cosmos. In modern terms, this is a theory without constraints: the theory determines *which* properties of a system must be specified in order to give adequate initial data, but we are then free to assign whatever *values* we please to these properties; the equations of the theory are used to evolve that data forwards in time. A theory with constraints, by contrast, contains two types of equations: constraint equations that must be satisfied by the initial data, as well as evolution equations. In GTR, four of the ten field equations connect the curvature of the initial data hypersurface with the distribution of mass–energy on that hypersurface, and the remaining six field equations are evolution equations. To sum up, in a theory with constraints, the initial “disorder” cannot be so disordered after all, but must itself satisfy constraints set down by the laws of the theory.

6.3 Does general covariance have any physical significance?

As we saw in Section ??, Einstein treated general covariance as a symmetry principle guiding the search that produced his General Theory of Relativity. There is no doubt that general covariance proved a useful heuristic for Einstein, but there remains an ongoing dispute over whether general covariance in fact has any physical significance. The issue was forcefully raised by Kretschmann already in 1917. The thrust of the argument, which continues to reverberate today, is that any theory can be given a generally covariant formulation given sufficient mathematical ingenuity, and therefore the principle of general covariance places no restrictions on the physical content of a theory. Indeed, Norton (2003) begins his discussion of the issue by claiming that this negative view of general covariance has become mainstream, before going on to give an alternative viewpoint (see below).

It seems clear to us that the characteristic features of generally covariant theories discussed above may, in some theories at least (including GTR), be far from trivial, and that the mainstream view – which would indeed render these issues trivial – should be opposed. Those wishing to oppose the mainstream view adopt a two-step general strategy: first, show under what conditions general covariance places a restriction on the physical content of a theory; and second, demonstrate what those implications for physical content consist in. Thus, the general mathematical point that any theory can be put into generally covari-

⁵²Written around 1633, *Le Monde* was not published in Descartes’s lifetime. For an English translation see Descartes (1998). The “order out of disorder” story is in the Treatise on Light, chs. 6 and 7. Whether the ordinary operation of the laws of nature was sufficient to bring order out of chaos became a much-disputed issue.

ant form is conceded, but the implication that general covariance is therefore necessarily physically vacuous is resisted by attention to the manner in which general covariance is implemented in a given theory or class of theories.

For example, Anderson (1967), Ohanian and Ruffini (1994), Norton (2003), and Earman (2006) each attempt to explain under what conditions the purely mathematical feature of general covariance comes to have physical bite.⁵³ Anderson distinguishes between the symmetries of a theory (which have physical significance) and the covariance group of the equations (which need not). Anderson is the classic reference for the distinction between “absolute” and “dynamical” objects,⁵⁴ and in this terminology the covariance group of the equations of a theory becomes a symmetry group if and only if the theory contains no absolute objects. Ohanian and Ruffini (1994) appeal to the distinction they make between *invariance* and *covariance* of the equations of a theory.⁵⁵ Covariance, they agree, is a mathematical feature (perhaps simply an artefact of the particular formulation of the theory at hand); but we require not only the covariance of the equations, but also that for any objects (with one or more components) appearing in the theory that are nevertheless independent of the state of matter (such as the speed of light, Planck’s constant, etc.), their value should be unchanged by the general coordinate transformations. Norton (2003) emphasizes the role of physical considerations in fixing the content of a theory such that this restricts the formal games that we can play. Earman (2006) begins by taking pains to emphasize the distinction between the ‘mere co-ordinate freedom’ (associated with arbitrary coordinate transformations, passively construed) and ‘the substantive demand that diffeomorphism invariance is a gauge symmetry of the theory at issue’. That is to say, he reminds us that the issue at stake is not our ability to re-write a theory in generally covariant form (it is conceded that this is something we can always do, given sufficient mathematical ingenuity), but the relationship between the physical situations that are related by diffeomorphisms, i.e. by (active) point transformations (see Section 6.1, above). ‘Substantive general covariance’ holds when diffeomorphically related models of the theory represent different descriptions of the same physical situation. The claim is that GTR satisfies substantive general covariance whereas generally covariant formulations of such theories as STR need not, and the goal is to show that this requirement provides demarcation between theories in which general covariance represents a physically significant property of the theory, and those in which it does not.⁵⁶

⁵³See also Norton (1993, especially Section 5), and Rovelli (this vol., ch. 12, Section 4.1.3).

⁵⁴It has proved difficult to make the distinction between absolute and dynamical objects precise, but the intuitive idea is clear enough. Dynamical objects satisfy field equations and interact with other objects, whereas absolute objects are not affected by the dynamical behaviour of other fields appearing in the theory. For a careful and detailed treatment of Anderson’s approach, and the counter-examples that have been raised, see Pitts (2006). The conclusion of this paper is that Anderson’s intuition can be made sufficiently precise to cope with all counter-examples that have appeared in the literature to date (including one due to Pitts himself), but that there is another example, due to Geroch, that Pitts has been unable to resolve. The debate goes on!

⁵⁵See Section ??.

⁵⁶One important tool for distinguishing genuine ‘gauge theories’ from those in which the

Thus, Anderson, Ohanian and Ruffini, Norton, and Earman each seek to add bite to the “merely mathematical” requirement of general covariance by placing conditions on the manner in which it is implemented in the theory. Once these requirements are added, various consequences follow for the content of the theory, such as that the metric be a dynamical object. In each case, the aim is to elevate general covariance as implemented in GTR to a symmetry principle.⁵⁷

Considerations of the significance of general covariance in theories of gravitation led to the formulation of three theorems important for the general interpretation of symmetries in physics. These theorems are due to Emmy Noether and Felix Klein, and will be discussed in the following section.

7 Noether’s theorems

Any discussion of the significance of symmetries in physics would be incomplete without mention of Noether’s theorems. These theorems relate symmetry properties of theories to other important properties, such as conservation laws.

Within physics, the term ‘Noether’s theorem’ is most frequently associated with a connection between global continuous symmetries and conserved quantities. Familiar examples from classical mechanics include the connections between: spatial translations and conservation of linear momentum; spatial rotations and conservation of angular momentum; and time translations and conservation of energy. In fact, this theorem is the first of two theorems presented in her 1918 paper ‘Invariante Variationsprobleme’.⁵⁸

Before stating the two theorems, we begin with the following cautionary remark. The connection between *variational symmetries* (connected to the invariance of the action, and in terms of which Noether’s theorems are formulated) and *dynamical symmetries* (concerning the dynamical laws, which is the topic of our discussion here) is subtle (see Olver, 1993, ch. 4). Noether herself never addressed the connection, and never used the word ‘symmetry’ in her paper. She discusses integrals mathematically analogous to (but generalizations of) the action integrals of Lagrangian physics, and uses variational techniques and group theory to elicit a pair-wise correspondence between variational symmetries of the integral and a set of identities.

Noether then proves two theorems, the first for the case where the variational symmetry group depends on constant parameters, and the second for the case where the variational symmetry group depends on arbitrary functions of the variables.⁵⁹ In the following statement of her theorems we use the term ‘Noether symmetry’ to refer to a symmetry of the field equations for which the change in the action arising from the infinitesimal symmetry transformation is

local symmetry in question is merely formal is Noether’s second theorem; see Section 7, below.

⁵⁷Brown and Brading (2002) attempt to analyze in more detail, by means of Noether’s theorems (see Section ??, below), what additional conditions must be added to general covariance in order to arrive at specific aspects of the content of GTR.

⁵⁸For an English translation see Noether (1971).

⁵⁹See Brading and Brown, 2007.

at most a surface term. Using the terminology of Section ??, the first type of symmetry then corresponds to a *global* dynamical symmetry, and the second to a *local* one. We state the theorems in a form appropriate to Lagrangian field theory; Noether's own statement of the theorems involves no such specialization. For discussion of the first theorem in the context of finite-dimensional classical mechanics see Butterfield (this vol., ch. 2, Section 2.1.3). We state the theorems so that we can refer back to them to characterize the conceptual content, but for discussion of the mathematical detail of their derivation and content we refer the reader elsewhere – see especially Olver (1993) and Barbashov and Nesterenko (1983).

We can state Noether's two theorems, for a Lagrangian density L depending on the fields $\phi_i(x)$ and their first derivatives, as follows.

Noether's first theorem

If a continuous group of transformations depending smoothly on ρ constant parameters ω_k ($k = 1, 2, \dots, \rho$) is a Noether symmetry group of the Euler-Lagrange equations associated with a Lagrangian $L(\phi_i, \partial_\mu \phi_i, x^\mu)$, then the following ρ relations are satisfied, one for every parameter on which the symmetry group depends:⁶⁰

$$\sum_i E_i^L \xi_i^k = \partial_\mu j_k^\mu. \quad (1)$$

On the left-hand side we have a linear combination of Euler expressions,

$$E_m^L \equiv \frac{\partial L}{\partial \phi_m} - \partial_\mu \left(\frac{\partial L}{\partial \phi_{m,\mu}} \right) \quad (2)$$

where

$$E_m^L = 0 \quad (3)$$

are the Euler-Lagrange equations for the field ϕ_m . (The ξ_i^m depend on the particular symmetry transformations and fields under consideration, and the details are not important for our current purposes.)

On the right-hand side we have the divergence of a current, j_k^μ . When the left-hand side vanishes, the divergence of the current is equal to zero, and this expression can be converted into a conserved quantity subject to certain conditions. Thus, Noether's first theorem gives us a connection between global symmetries and conserved quantities.⁶¹

Noether's second theorem

If a continuous group of transformations depending smoothly on ρ arbitrary functions of time and space $p_k(x)$ ($k = 1, 2, \dots, \rho$) and their first derivatives is

⁶⁰Note that we are using the Einstein summation convention to sum over repeated greek indices.

⁶¹This theorem is widely discussed. See especially Barbashov and Nesterenko (1983); Doughty (1990). We refer the reader to Butterfield (this vol., ch. 2, and 2006) for further discussion of Noether's first theorem in the context of finite-dimensional classical mechanics.

a Noether symmetry group of the Euler-Lagrange equations associated with a Lagrangian $L(\phi_i, \partial_\mu \phi_i, x^\mu)$, then the following ρ relations are satisfied, one for every function on which the symmetry group depends:

$$\sum_i E_i^L a_{ki} = \sum_i \partial_\nu (b_{ki}^\nu E_i^L). \quad (4)$$

The a_{ki} and b_{ki}^ν depend on the particular transformations of the fields in question, and while again the details need not concern us here, we note for use below that while the a_{ki} arise even when the symmetry transformation is a global transformation, the b_{ki}^ν occur only when it is *local*.⁶² What we have here, essentially, is a dependency between the Euler expressions and their first derivatives. This dependency holds as a consequence of the *local* symmetry used in deriving the theorem. In the case when all the fields are dynamical (i.e. satisfy Euler-Lagrange equations) it follows that not all the field equations are independent of one another. This *formal underdetermination* is characteristic of theories with a local symmetry structure.⁶³

As Hilbert recognized in the context of generally covariant theories of gravitation, the underdetermination is independent of the specific form of the Lagrangian.⁶⁴ In the case of General Relativity, once we specify the Lagrangian and substitute it into (??), we arrive at the (contracted) Bianchi identities.

For Noether herself, the impetus for the paper arose from the discussions over the status of energy conservation in generally covariant theories between Hilbert, Klein and Einstein, during which Hilbert commented that energy conservation for the matter fields no longer has the same status in generally covariant theories as it had in previous (non-generally covariant) theories, because it follows independently of the field equations for the matter fields. Noether's two theorems can be used to support this conjecture (see Brading, 2005). The discussion over the status of energy conservation in General Relativity continues, the root of the issue being that energy–momentum cannot, in general, be defined locally.⁶⁵

⁶²Once again, the reader is referred to Brading and Brown (2007) for further details.

⁶³Whether the dependencies expressed by the second theorem are trivial or not depends on the status of the fields with respect to which the local symmetry holds. It is in this way that Noether's second theorem can be used as a tool in the attempt to demarcate 'true gauge theories' from theories where the local symmetry is a 'mere mathematical artefact' (see Section 6.3 above, and Earman, 2006). For a 'true gauge theory' the dependencies have significant physical implications.

⁶⁴Hilbert (1915).

⁶⁵The energy–momentum conservation law in General Relativity is formulated in terms of the vanishing of the *covariant* divergence of the energy–momentum tensor associated with the matter fields. Alternatively, we can express this in terms of the vanishing of the coordinate divergence of the energy–momentum of the matter fields plus that of the gravitational field. The latter term falling under the divergence operator is not uniquely defined and, pertinent the issue of non-localizability, may vanish in some coordinate systems and not in others. We can understand this coordinate dependence by reflecting on the equivalence principle, according to which partitioning the inertial-gravitational field to obtain a division between inertial and gravitational forces is itself a coordinate-dependent issue. For further discussion see, for example, Misner, Thorne and Wheeler (1970), pp. 467–8, and Wald (1984), p. 70. See also Malament (this vol., ch. 3).

Today, the significance of Noether’s results lie in their generality. Many of the specific connections between global spacetime symmetries and their associated conserved quantities were known before Noether’s 1918 paper, and both Einstein and Hilbert anticipated some aspects of the second theorem in their investigations of energy conservation during and after the development of GTR.⁶⁶ However, her systematic treatment allows us to understand that these relations do not rely on the detailed dynamics of a particular theory, but in fact follow from the structure of Lagrangian theories and significantly weaker stipulations than the full dynamics of the theory. For example, general covariance leads to energy conservation in GTR given satisfaction of the gravitational field equations, but *independently* of the detailed form of those equations, and *independently* of the fields equations for the matter fields (indeed, independently of whether the matter fields satisfy Euler-Lagrange equations at all).⁶⁷ Noether’s theorems are a powerful tool for investigating the structure of theories – which assumptions are required to generate which aspects of the theory, and so forth.⁶⁸

It is worthwhile mentioning a third theorem, connected with Noether’s two theorems and derived in the same context (i.e. the study of generally covariant theories of gravitation and conservation of energy) by Felix Klein (1918). We call it the ‘Boundary theorem’ for reasons associated with its method of derivation.⁶⁹ As with Noether’s second theorem, the Boundary theorem concerns local symmetries, and results in a series of identities (termed the ‘cascade equations’ by Julia and Silva (1998)).⁷⁰ We state here a simplified version of the Boundary theorem in which the action is left unchanged by an infinitesimal symmetry transformation (i.e. we do not allow for the possibility of a surface term).⁷¹

The Boundary theorem (restricted form)

If a continuous group of transformations depending smoothly on ρ arbitrary functions of time and space $p_k(x)$ ($k = 1, 2, \dots, \rho$) and their first derivatives is a Noether symmetry group⁷² of the Euler-Lagrange equations associated with a Lagrangian $L(\phi_i, \partial_\mu \phi_i, x^\mu)$, then the following three sets of ρ relations are satisfied, one for every parameter on which the symmetry group depends:

⁶⁶On Einstein, see Janssen (2005), pp. 75–82; and see Sauer (1999) on Hilbert.

⁶⁷See Brading and Brown (2007).

⁶⁸For a discussion of this in the case of general covariance, see Brading and Brown (2002).

⁶⁹The Boundary theorem also appears in the work of Hermann Weyl, specialized to the case of his unified field theory (see Weyl, 1922, p. 287–289; the first appearance was in the 1919 third edition), and was published in a non-theory-specific form by Utiyama (1956, 1959).

⁷⁰As with Noether’s second theorem, the Boundary theorem is a useful tool in the attempt to demarcate ‘true gauge theories’ from theories where the local symmetry is a ‘mere mathematical artefact’, through inspection of the identities that result from the theorem, and through the physical significance – or otherwise – of these identities.

⁷¹For further details of the Boundary theorem, including the generalization that allows for a surface term, see Brading and Brown (2007).

⁷²The Boundary theorem is here stated in a restricted form such that the Noether symmetry group must belong to the restricted class of such groups associated with an *invariant* action.

$$\sum_i \partial_\mu (b_{ki}^\mu E_i^L) = \partial_\mu j_k^\mu \quad (5)$$

$$\sum_i (b_{ki}^\mu E_i^L) = j_k^\mu - \sum_i \left[\partial_\nu \left(\frac{\partial L}{\partial(\partial_\nu \phi_i)} b_{ki}^\mu \right) \right] \quad (6)$$

$$\left(\frac{\partial L}{\partial(\partial_\mu \phi_i)} b_{ki}^\nu \right) + \left(\frac{\partial L}{\partial(\partial_\nu \phi_i)} b_{ki}^\mu \right) = 0. \quad (7)$$

Once again, the b_{ki}^ν depend on the particular transformations of the fields in question, the details of which need not concern us here. The first identity is connected to the existence of *superpotentials* associated with local symmetries.⁷³ The second equation can be used to investigate the relationship between a field and its sources. For example, in the case of classical electromagnetism, we can investigate the relationship between the local gauge symmetry of the theory and the condition that:

$$j^\mu = \partial_\nu F^{\mu\nu}, \quad (8)$$

i.e. that Maxwell's equations with dynamical sources hold. Using the case of classical electromagnetism as our example once again, the third equation becomes the condition that the electromagnetic tensor be antisymmetric (showing the relationship between this condition and the local gauge symmetry of that theory):

$$F^{\mu\nu} + F^{\nu\mu} = 0. \quad (9)$$

These remarks have been necessarily brief, and the reader is referred to Barbashov and Nesterenko (1983), along with Brading and Brown (2003 and 2007), for detailed derivations and discussion of these results. The identities of the Boundary theorem and of Noether's two theorems are not all independent of one another, and which is most useful depends on the context and the question under consideration. As with Noether's theorems, the Boundary theorem holds independently of the specific details of the dynamical equations, and together they allow us to investigate structural features of our theories that are associated with the symmetry properties of those theories.

8 The interpretation of symmetries in classical physics

In what follows, we begin with 'Wigner's hierarchy', which has become the canonical view of the relationship between symmetries, laws and events. We supplement this with a brief discussion of the connection between symmetry and

⁷³See, for example, Trautman (1962, p. 179).

irrelevance, and how this bears on the interpretation of the various symmetries described in Section ??, above.

The general interpretation of symmetries in physical theories can adopt a number of complementary approaches. We can ask about the different *roles* that various symmetries play; about the epistemological, ontological or other *status* that various symmetries have; and about the significance of the structures left invariant by symmetry transformations. We end with some remarks on each of these issues.

8.1 Wigner's hierarchy

The starting point for contemporary philosophical discussion of the status and significance of symmetries in physics is Wigner's 1949 paper 'Invariance in Physical Theory', along with his three later papers published in 1964.⁷⁴ In these papers, Wigner makes the distinction mentioned above (see Section ??) between geometrical and dynamical symmetries, which we will return to below. He also presents his view of the *hierarchy* of physical knowledge, according to which symmetries are viewed as properties of laws:

There is a strange hierarchy in our knowledge of the world around us. Every moment brings surprises and unforeseeable events – truly the future is uncertain. There is, nevertheless, a structure in the events around us, that is, correlations between the events of which we take cognizance. It is this structure, these correlations, which science wishes to discover, or at least the precise and sharply defined correlations. ... We know many laws of nature and we hope and expect to discover more. Nobody can foresee the next such law that will be discovered. Nevertheless, there is a structure in the laws of nature which we call the laws of invariance. This structure is so far-reaching in some cases that laws of nature were guessed on the basis of the postulate that they fit into the invariance structure. ... This then, the progression from events to laws of nature, and from laws of nature to symmetry or invariance principles, is what I meant by the hierarchy of our knowledge of the world around us. (Wigner, 1967, pp. 28–30).

This view of symmetries, as properties of laws, has become canonical.

8.2 Symmetry and irrelevance

There is a general property of laws, or of the underlying events, to which symmetries are connected: the *irrelevance* of certain quantities that might otherwise

⁷⁴Wigner's papers can be found in the collection *Symmetries and Reflections* (Wigner, 1967).

be thought to have physical significance.⁷⁵ In Section ?? we outlined the variety of symmetries found in physics, and in each case the symmetry is associated with a property that is deemed irrelevant for the purposes of describing the law-governed behaviour of a system. For example, left-right symmetry means that whether a system is left-handed or right-handed is irrelevant to its law-governed evolution. Famously, this symmetry is violated in the weak interaction: the law-governed behaviour of systems turns out to be sensitive to handedness for certain processes (see Pooley, 2003).

In Section ?? we characterized the distinction between global and local symmetries mathematically, in terms of the dependence on constant parameters and arbitrary functions of time (and space) respectively. The physical meaning of this distinction can be understood through the associated properties that are deemed *irrelevant*. A *global* symmetry reflects the irrelevance of absolute values of a certain quantity: only relative values are relevant. So in Newtonian mechanics, for example, spatial translation invariance holds and absolute position is irrelevant to the behaviour of systems.⁷⁶ Only relative positions matter, and this is reflected in the structure of the theory through the equations being invariant under global spatial translations – the equations do not depend upon, or invoke, a background structure of absolute positions.

A global symmetry is a special case of a local symmetry. A *local* symmetry reflects the irrelevance not only of absolute values, but furthermore of relative values specified at-a-distance: only *local* relative values (i.e. relative values specified at a point) are relevant. This is reflected in the structure of the theory by the equations of motion not depending upon some background structure that determines relative values at-a-distance (i.e. there is no global background structure associated with the property in question).⁷⁷

8.3 Roles of symmetries

The various different roles in which symmetries are invoked in physics have become much more evident with the advent of quantum theory.⁷⁸ Nevertheless, already with the classification of crystals using their remarkable and varied symmetry properties, we see the powerful *classificatory* role at work. Indeed, it

⁷⁵For an analysis of the connection between symmetry, equivalence and irrelevance, see Castellani (2003).

⁷⁶We are considering here *Newtonian* mechanics, without Newton’s absolute space.

⁷⁷Instead, we require the explicit appearance of a *connection* in our theory, which provides the rules by which two distant objects may be brought together so that comparisons between them may be made locally.

⁷⁸The application of the theory of groups and their representations for the exploitation of symmetries in the quantum mechanics of the 1920s represents a dramatic step-change in the significance of symmetries in physics, with respect to both the foundations and the phenomenological interpretation of the theory. As Wigner emphasized on many occasions, one essential reason for the ‘increased effectiveness of invariance principles in quantum theory’ (Wigner, 1967, p. 47) is the linear nature of the state space of a quantum physical system, corresponding to the possibility of superposing quantum states. For details on the application of symmetries in quantum physics we refer the reader to Harvey (this vol., ch. 14). For philosophical discussions see Brading and Castellani (2003).

was with René-Just Haüy’s use of symmetries in this way that crystallography emerged in 1801 as a discipline distinct from mineralogy.⁷⁹ Furthermore, the *heuristic* and/or *normative* role is clear for the principle of relativity in the construction of both Special and General Relativity (see above, Section ??). The *unificatory role*, so prominent now in the attempts to unify the fundamental forces, was already present (although differing methodologically somewhat) in Hilbert’s attempt to construct a generally covariant theory of gravitation and electromagnetism (see Sauer, 1999) and in Weyl’s 1918 unified theory of gravitation and electromagnetism, for example. Symmetries may also be invoked in a variety of *explanatory* roles. For example, on the basis of Noether’s first theorem (see Section ??) we might say that it is *because* of the translational symmetry of classical mechanics (plus satisfaction of other conditions) that linear momentum is conserved in that theory. Another example would be an appeal to symmetry principles as an explanation, via Wigner’s hierarchy, for (i) aspects of the *form* of the laws, and thereby (ii) why certain events occur and others do not.

8.4 Status of symmetries

Are symmetries ontological, epistemological, or methodological in status? It is clear that symmetries have an important heuristic function, as discussed above (Section ??) in the context of relativity. This indicates a methodological status, something that becomes further developed within the context of quantum theory. We can also ask whether we should attribute an ontological or epistemological status to symmetries.

According to an ontological viewpoint, symmetries are seen as “existing in nature”, or characterizing the structure of the physical world. One reason for attributing symmetries to nature is the so-called geometrical interpretation of spatiotemporal symmetries, according to which the spatiotemporal symmetries of *physical laws* are interpreted as symmetries of *spacetime itself*, the “geometrical structure” of the physical world. Moreover, this way of seeing symmetries can be extended to non-external symmetries, by considering them as properties of other kinds of spaces, usually known as “internal spaces”.⁸⁰ The question of exactly what a realist would be committed to on such a view of internal spaces remains open, and an interesting topic for discussion.

One approach to investigating the limits of an ontological stance with respect to symmetries would be to investigate their empirical or observational status: can the symmetries in question be directly observed? We first have to address what it means for a symmetry to be observable, and indeed whether all symmetries have the same observational status. Kosso (2000) arrives at the conclusion that there are important differences in the empirical status of the dif-

⁷⁹The use of discrete symmetries in crystallography continued through the nineteenth century in the work of J. F. Hessel and A. Bravais, leading to the 32 point transformation crystal classes and the 14 Bravais lattices. These were combined into the 230 space groups in the 1890s by E. S. Fedorov, A. Schönflies, and W. Barlow. The theory of discrete groups continues to be important in such fields as solid state physics, chemistry, and materials science.

⁸⁰See Section 4.2, above, for the varieties of symmetry.

ferent kinds of symmetries. In particular, while global continuous symmetries can be directly observed – via such experiments as the Galilean ship experiment – a local continuous symmetry can have only indirect empirical evidence.⁸¹

The direct observational status of the familiar global spacetime symmetries leads us to an epistemological aspect of symmetries. According to Wigner, the spatiotemporal invariance principles play the role of a prerequisite for the very possibility of discovering the laws of nature: ‘if the correlations between events changed from day to day, and would be different for different points of space, it would be impossible to discover them’ (Wigner, 1967). For Wigner, this conception of symmetry principles is essentially related to our ignorance (if we could directly know all the laws of nature, we would not need to use symmetry principles in our search for them). Such a view might be given a methodological interpretation, according to which such spatiotemporal regularities are presupposed in order for the enterprise of discovering the laws of physics to get off the ground.⁸² Others have arrived at a view according to which symmetry principles function as “transcendental principles” in the Kantian sense (see for instance Mainzer, 1996). It should be noted in this regard that Wigner’s starting point, as quoted above, does not imply exact symmetries – all that is needed epistemologically (or methodologically) is that the global symmetries hold approximately, for suitable spatiotemporal regions, so that there is sufficient stability and regularity in the events for the laws of nature to be discovered.

As this discussion, and that of the preceding Subsections, indicate, the *differences* between various types of symmetry become important before we have ventured very far into interpretational issues. For this reason, much recent work on the interpretation of symmetry in physical theory has focussed not on general questions, such as those sketched above, but on addressing interpretational questions specific to particular symmetries.⁸³

8.5 Symmetries, objectivity, and objects

Turning now to the issue of the structures left invariant by symmetry transformations, the old and natural idea that what is objective should not depend upon the particular perspective under which it is taken into consideration is reformulated in the following group theoretical terms: what is objective is what is invariant with respect to the relevant transformation group. This connection between symmetries and objectivity is something that has a long history going back to the early twentieth century at least. It was highlighted by Weyl (1952),

⁸¹See Section 6.1, above; and Brading and Brown (2003b), who argue for a different interpretation of Kosso’s examples.

⁸²We are grateful to Brandon Fogel for this point, and for the comparison he suggested between this view of spatiotemporal symmetries and the methodological face of Einstein’s notion of separability.

⁸³These include the varieties of gauge invariance found in classical electromagnetism and in quantum theories, along with general covariance in GTR (these being continuous symmetries), plus the discrete symmetries of parity (violated in the weak interaction) and permutation invariance, both of which are found in classical theory but require reconsideration in the light of quantum theory. See Brading and Castellani (2003).

where he writes that ‘We found that objectivity means invariance with respect to the group of automorphisms.’ This connection between objectivity and invariance was discussed particularly in the context of Relativity Theory, both Special and General. We recall Minkowski’s famous phrase ([1908] 1923, p. 75) that ‘Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality’, following his geometrization of Einstein’s Special Theory of Relativity, and the recognition of the spacetime interval (rather than intervals of space and of time) as the geometrically invariant quantity. The connection between objectivity and invariance in General Relativity was discussed by, amongst others, Hilbert and Weyl, and continues to be an issue today.⁸⁴

Related to this is the use of symmetries to characterize the *objects* of physics as sets of invariants. Originally developed in the context of quantum theory, this approach can also be applied in classical physics.⁸⁵ The basic idea is that the invariant quantities – such as mass and charge – are those by which we characterize objects. Thus, through the application of group theory we can use symmetry considerations to determine the invariant quantities and “construct” or “constitute” objects as sets of these invariants.⁸⁶

In conclusion, then, the philosophical questions associated with symmetries in classical physics are wide-ranging. What we have offered here is nothing more than an overview, influenced by our own interests and puzzles, which we hope will be of service in further explorations of this philosophically and physically rich field.

Acknowledgements – We are grateful to the editors, Jeremy Butterfield and John Earman, for their encouragement and detailed comments. We would also like to thank Brandon Fogel, Brian Pitts, and Thomas A. Ryckman for their comments and suggestions.

9 References

R. Adler, M. Bazin, and M. Schiffer (1975), *Introduction to General Relativity*, 2nd edition, McGraw-Hill Book Company, New York, St Louis, San Francisco, Toronto, London, Sydney.

J. L. Anderson (1967), *Principles of Relativity Physics*, New York and London: Academic press.

B. M. Barbashov and V. V. Nesterenko (1983), ‘Continuous symmetries in field theory’, *Fortschr. Phys.* **31**, pp. 535-67.

⁸⁴We saw above (Sections 6.2 and 6.3) some aspects of this debate in the discussion of Einstein’s ‘hole argument’ and of the status of observables in GTR.

⁸⁵See Max Born, reprinted in Castellani (1998).

⁸⁶For further discussion see Castellani (1998), part II.

- J. B. Barbour (1989), *Absolute or relative motion?* Vol. 1 *The discovery of dynamics*, Oxford University Press.
- J. B. Barbour (2005), *Absolute or relative motion?* Vol. 2 *The deep structure of general relativity*, Oxford University Press.
- G. Belot (1998), ‘Understanding Electromagnetism’, *British Journal for the Philosophy of Science* **49**, pp. 531-55.
- G. Belot (this volume), ‘Geometric Mechanics’.
- K. Brading (2005), ‘A note on general relativity, energy conservation, and Noether’s theorems’, in *The Universe of General Relativity (Einstein Studies)*, ed. J. Eisenstaedt and A. J. Kox, Birkhäuser.
- K. Brading and H. R. Brown (2002), ‘General covariance from the perspective of Noether’s theorems’, *Diálogos*, pp. 59-86.
- K. Brading and H. R. Brown (2003), ‘Symmetries and Noether’s theorems’, in K. Brading and E. Castellani (eds.) (2003), pp. 89-109.
- K. Brading and H. R. Brown (2004), ‘Are gauge symmetry transformations observable?’, *British Journal for the Philosophy of Science* **55**, pp. 645-65.
- K. Brading and H. R. Brown (2007), ‘Noether’s theorems, gauge symmetries and general relativity’, manuscript.
- K. Brading and E. Castellani (eds.) (2003), *Symmetry in Physics: Philosophical Reflections*, Cambridge University Press.
- K. Brading and E. Castellani (2006), ‘Curie’s Principle, Encore’, in preparation.
- K. Brading and T. A. Ryckman (2007), ‘Hilbert’s axiomatic method and the foundations of physics: an interpretation of generally covariant physics and a revision of Kant’s epistemology’, manuscript.
- H. Brown (2006), *Physical relativity: spacetime structure from a dynamical perspective*, Oxford University Press.
- J. Butterfield (2004), ‘Between Laws and Models: Some Philosophical Morals of Lagrangian Mechanics; available at Los Alamos arXiv: <http://arxiv.org/abs/physics/0409030>; and at Pittsburgh archive: <http://philsci-archive.pitt.edu/archive/00001937/>.
- J. Butterfield (2006), ‘On Symmetry and Conserved Quantities in Classical Mechanics’, forthcoming in a Festschrift for Jeffrey Bub, ed. W. Demopoulos and I. Pitowsky, Kluwer: University of Western Ontario Series in Philosophy of Science; available at Los Alamos arXiv: <http://arxiv.org/abs/physics/>; and at Pittsburgh archive: <http://philsci-archive.pitt.edu/archive/00002362/>.
- J. Butterfield (this volume), ‘On Symplectic Reduction in Classical Mechanics’.
- E. Castellani (ed.) (1998), *Interpreting Bodies. Classical and Quantum Objects in Modern Physics*, Princeton University Press.
- E. Castellani (2003), ‘Symmetry and Equivalence’, in K. Brading and E. Castellani (eds.) (2003), pp. 321-334.

- A. F. Chalmers (1970), ‘Curie’s Principle’, *The British Journal for the Philosophy of Science* **21**, pp. 133-148.
- L. Corry (2004), *David Hilbert and the Axiomatization of Physics (1898-1918)*, Dordrecht: Kluwer Academic.
- P. Curie (1894), ‘Sur la symétrie dans les phénomènes physiques. Symétrie d’un champ électrique et d’un champ magnétique’, *Journal de Physique*, 3^e série, vol. 3, pp. 393-417.
- P. Curie (1981), ‘On symmetry in physical phenomena’, trans. J. Rosen and P. Copié, *Am. J. Phys.* **49**(4), pp. 17-25.
- R. Descartes (1998), *The World and other writings*, ed. and trans. S. Gaukroger, Cambridge University Press.
- N. Doughty (1990), *Lagrangian interaction*, Addison-Wesley.
- J. Earman (2004), ‘Curie’s Principle and Spontaneous Symmetry Breaking’, *International Studies in Philosophy of Science* **18**, pp. 173-199.
- J. Earman (2006), ‘Two Challenges to the Substantive Requirement of General Covariance’, *Synthese*, in press.
- J. Earman (this volume), ‘Determinism in Modern Physics’.
- J. Harvey (this volume), ‘Symmetries and Invariances in Quantum Physics’.
- T. Hawkins (2000), *Emergence of the Theory of Lie Groups: an essay in the history of mathematics 1869-1926*, New York: Springer.
- D. Hilbert (1915), ‘Die Grundlagen der Physik (Erste Mitteilung), Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen’, *Mathematisch-physikalische Klasse*, pp. 395-407.
- ’t Hooft (this volume), ‘Conceptual Basis of Quantum Field Theory’.
- D. Howard (2007), ‘‘And I Shall not Mingle Conjectures and Certainties’: The Roots of Einstein’s Principle Theories/Constructive Theories Distinction’, manuscript.
- J. Ismael (1997), ‘Curie’s Principle’, *Synthese* **110**, pp. 167-190.
- M. Janssen (2005), ‘Of pots and holes: Einstein’s bumpy road to general relativity’, *Ann. Phys.* **14**, Supplement, pp. 58-85.
- B. Julia and S. Silva (1998), ‘Current and superpotentials in classical gauge invariant theories I. Local results with applications to perfect fluids and general relativity’, gr-qc/9804029 v2.
- F. Klein (1918), ‘Über die Differentialgesetze für die Erhaltung von Impuls und Energie in der Einsteinschen Gravitationstheorie. Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse’, *Nachrichten*, pp. 469-92.
- P. Kosso (2000), ‘The empirical status of symmetries in physics’, *British Journal for the Philosophy of Science* **51**, pp. 81-98.
- C. Lanczos ([1949], 1962), *The Variational Principles of Mechanics*, University of Toronto Press.

- K. Mainzer (1996), *Symmetries of Nature. A Handbook*, New York: de Gruyter.
- D. Malament (this volume), ‘Special and General Relativity Theory’.
- A. I. Miller (1981), *Albert Einstein’s Special Theory of Relativity*, Addison-Wesley Publishing Company.
- H. Minkowski ([1908], 1923), ‘Space and Time’, in *The Principle of relativity. A collection of original memoirs on the special and general theory of relativity*, New York: Dover, pp. 75-91.
- C. W. Misner, K. S. Thorne and J. A. Wheeler (1970), *Gravitation*, New York: W. H. Freeman and Company.
- E. Noether (1971), ‘Noether’s theorem’, *Transport Theory and Statistical Physics* 1, 183-207.
- J. Norton (1984), ‘How Einstein found his field equations: 1912-1915’, *Hist. Stud. Phys. Sci.* **14**, pp. 253-316.
- J. Norton (1993), ‘General covariance and the foundations of general relativity: eight decades of dispute’, *Rep. Prog. Phys.* **56**, pp. 791-858.
- J. Norton (2003), ‘General covariance, gauge theories, and the Kretschmann objection’, in K. Brading and E. Castellani (eds.) (2003), pp. 110-123.
- H. C. Ohanian and R. Ruffini (1994), *Gravitation and Spacetime*, 2nd edition. London, New York: W. W. Norton and company.
- P. J. Olver (1993), *Applications of Lie Groups to Differential Equations*, 2nd edition, New York: Springer.
- L. O’Raifeartaigh (1997), *The Dawning of Gauge Theory*, Princeton University Press.
- J. B. Pitts (2006), ‘Absolute objects and counterexamples: Jones-Geroch Dust, Torretti Constant Curvature, Tetrad-Spinor, and Scalar Density’, *Studies in History and Philosophy of Modern Physics*, forthcoming. Manuscript available at Los Alamos arXive: <http://arxiv.org/abs/gr-qc/0506102/>.
- O. Pooley (2003), ‘Handedness, parity violation, and the reality of space’ , in K. Brading and E. Castellani (eds.) (2003), pp. 250-280.
- L. A. Radicati (1987), ‘Remarks on the early developments of the notion of symmetry breaking’, in M.G. Doncel, A. Hermann, L. Michel, and A. Pais (eds.), *Symmetries in physics (1600-1980)*, Barcelona: Servei de Publicacions, pp. 195-206.
- C. Rovelli (this volume), ‘Quantum Gravity’.
- T. Ryckman (2005), *The Reign of Relativity*, Oxford University Press.
- T. Sauer (1999), ‘The relativity of discovery: Hilbert’s first note on the foundations of physics’, *Arch. Hist. Exact Sci.* **53**, pp. 529-575.
- A. V. Shubnikow and V. A. Koptsik (1974), *Symmetry in Science and Art*, London: Plenum Press.

- J. Stachel (1993), 'The meaning of general covariance: the hole story', in *Philosophical Problems of the Internal and External World: Essays on the Philosophy of Adolf Grünbaum.*, ed. J. Earman et al., University of Pittsburgh Press, pp. 129-160.
- R. Torretti (1983), *Relativity and Geometry*, Oxford: Pergamon Press.
- A. J. Trautman (1962), 'Conservation laws in general relativity', in *Gravitation: an introduction to current research*, ed. L. Witten, New York: John Wiley and Sons, pp. 169-98.
- R. Utiyama (1956), 'Invariant theoretical interpretation of interaction', *Physical Review* **101**, pp. 1597-607.
- R. Utiyama (1959), 'Theory of invariant variation in the generalized canonical dynamics', *Prog. Theor. Phys. Suppl.* **9**, pp. 19-44.
- R. Wald (1984), *General Relativity*, University of Chicago Press.
- H. Weyl (1922), *Space, Time, Matter*, reprinted by Dover (1952).
- H. Weyl (1952), *Symmetry*, Princeton University Press.
- E. T. Whittaker ([1904] 1989), *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press.
- E. P. Wigner (1967), *Symmetries and Reflections*, Bloomington: Indiana University Press.
- I. M. Yaglom (1988), *Felix Klein and Sophus Lie. Evolution of the Idea of Symmetry in the Nineteenth Century*, Boston-Basel: Birkhäuser.