Essay Review: Symmetries in Physics*

Jordi Cat†

K. Brading and E. Castellani, eds., ***Symmetries in Physics: Philosophical Reflections***. Cambridge: Cambridge University Press (2003), xii+445 pp.

Editors, like generals, often have to make do with the articles they have and not the ones they wish they had. The competent and committed editors of this volume, Katherine Brading and Elena Castellani, have done their earnest to package it as a means—or rather with hope—to bring to "non-experts, including students and philosophers of science in general" (ix) a point of access into the question of the role and significance of the technical notions of symmetry in contemporary physics and into a spectrum of current philosophical debates about them. The centrality of the notion of symmetry in mathematics—in general a formal property of equivalence or invariance in equations and quantities—and its application to physics cannot escape the attention of any student of the evolution and many successes of modern physics. In fact, Klein and Lie's invariants theory is what Emmy Noether worked out in mathematics, and Invariance Theory is what Einstein had originally meant to call his infamously misleadingly renamed Relativity Theory in physics. (The actual scope of most of the articles suggests that a more adequate title might have been *Symmetries in Twentieth-Century Physics* [cursory historical remarks and references to classical physics notwithstanding]).

To some philosophers the scientific significance and philosophical puzzle and lessons involving symmetries are linked to their status as redundancy or "surplus structure" in physical descriptions. To others, they are linked to the resulting invariance property of quantities or laws of evolution under the corresponding mathematical transformations. But what exactly the different kinds of symmetries do, how they are linked to much trum-

*Received July 2006.

†To contact the author, please write to: Department of History and Philosophy of Science, Goodbody Hall, 1011 E. Third Street, Indiana University, Bloomington, IN 47405; e-mail: jcat@indiana.edu.

Philosophy of Science, 73 (October 2006): 459–468. 0031-8248/2006/7304-0007$10.00
Copyright 2006 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/516808 Published online by Cambridge University Press

peted successes and what follows from it all more broadly, connectively, abstractly, and deeply is far from clear, certain, or unique. The historical and philosophical challenges have not gone away.

The actual level of discussion in the volume might give the foreboding appearance of philosophers of physics writing for one another, enthusiastically engaging and exchanging contemporary mathematical and physical formalisms they have recently encountered. Moreover, since the volume stems exclusively from contributions to a workshop, it does not include significant previous contributions to the literature on this topic, even by some of their volume contributors themselves (whose subsequent contribution here amounts hardly to more than a summary or an addendum). But do not fret. Anyone who wants to pursue these issues further will benefit from the bibliographical references. The editors doubtless realized, in addition, that the only hope for those "non-experts," the readers who may lack technical and more philosophical interests in physics, was to read something close to the remarkable introductory and concluding short pieces.

These texts lay out a helpful framework with a survey of ontological, epistemological, and methodological issues. Some of the issues are more general and applied, ranging from descriptive issues of classification, (structural) realism, and empirical content and equivalence to methodological and normative ones of prediction, objectivity, and unity (although some of those important issues—classification, unification, and others—are not really addressed by the articles). Other issues such as determinism and nonlocality are more substantive and specific to philosophy of physics and raise a tangle of conceptual and methodological questions. Beyond the editors' contributions, those lay readers are advised to continue reading the rest of the articles at their own risk, and I suggest that they do so in reverse order. They—and anyone else—will also do well to read first the widely informative volume *Symmetries in Physics (1600–1980)*, the illustrious predecessor edited by M. Doncel, A. Hermann, L. Michel, and A. Pais (1983).

The main value of the new volume resides, in my view, in three overall contributions: its intended survey of many current issues and debates around the topic of symmetries in physics, with no signs of impending settlement; its attention to symmetry breaking; and, no less interesting, its accidental family portrait of different approaches in philosophy of physics and science with their respective preferred types of problems and projects, styles and standards. Next I provide a summary. It is, of course, necessarily brief, selective, and unjust.

The volume is divided into four parts; each deals with a specific physical or philosophical focus: continuous symmetries, discrete symmetries, symmetry breaking, and general interpretive issues. Each part also opens with

https://doi.org/10.1086/516808 Published online by Cambridge University Press

two of three concise excerpts of classical texts by such authors as Leibniz, Kant, Black, Curie, Weyl, and Wigner on some of the topics at hand (I must say, they come to the lay reader without sufficient information or context to serve as an enlightening historical or conceptual entry point to the respective part, even though they are subsequently referred to by some of the contributors).

Part I deals with continuous symmetries, mainly with gauge symmetries. Gauge symmetries are at the core of the most important predictive, explanatory, and unifying successes of modern fundamental theories of physical interactions. What used to be conceptualized as a force and what new kinds of interactions, such as weak "forces," have been discovered are now conceptualized as gauge fields. Gauge transformations and invariances appeared originally as properties of continuous electromagnetic descriptions: first in Maxwell's use and awareness of underdetermined potential functions associated with local force and energy quantities, and later in Weyl's notion of the connection between a local gauge or unity of length at different and separate spatial points. Unlike these and the more intuitive and discrete kinds of transformations, which involve standard elements of space, time, or matter, gauge transformations quickly became far more abstract and useful in the context of quantum physics. At the same time, like so many other kinds of descriptive excesses in twentieth-century culture, they also became formally and interpretively more controversial: are gauge fields real physical entities? Is the ontology of fundamental physics definitively nonlocal or nonseparable? Are gauge symmetries better understood in terms of fiber bundles, state-space orbits, or Hamiltonian constraints? Et cetera.

Part I opens, past the excerpts just mentioned, with two of the longest and most important and substantive articles in the volume. They tackle the question of the normative and a priori role of the gauge symmetry principle or argument, namely, the demand that equations of motion and physical quantities be invariant under local gauge transformations. To that effect they examine their philosophical and methodological roots, uses, and defenses. Chris Martin's article elaborates historically and philosophically on this issue on two related fronts, scientific and philosophical, of the gauge argument. The first front extends between Steven Weinberg's (also C. N. Yang's) two well-known methodological banners: the more historical is that twentieth-century physics is the century of symmetries and the more Platonic, with more fin de siècle hubris, is that symmetry principles "dictate the very existence of all known forces of nature" (29). The second front lays out a tension he sees from a normative and interpretive point of view, between the two attitudes I have mentioned above now for the case of gauge symmetries: the "redundancy view"—physically vacuous—and the "profundity view," Weinberg's—methodologically and

https://doi.org/10.1086/516808 Published online by Cambridge University Press

aesthetically supreme. Martin's position holds the historical banner but advocates, against Weinberg, a heuristic, inclusive and local attitude toward understanding the normative role of gauge principles.

In the second paper Thomas Ryckman examines the philosophical roots of the gauge principle in Weyl's thought. In this case, the guiding banner is Weyl's dictum that "all *a priori* statements in physics have their origin in symmetry" (61). In a highly original analysis he argues that Weyl's scientific strategy stemmed from his epistemological problem of understanding the sense and justification of the mathematical structures of classical field theory. Weyl's solution—anticipating Carnap, also under Husserl—was the construction or constitution of such structures out of geometrical relations evident to the subject within an infinitesimal manifold of intuition. The solution rests on Weyl's joint adherence to Reimann's and Leibniz's infinitesimal geometry and metaphysics, and to Husserl's phenomenology (or transcendental phenomenological idealism).

Brading and Harvey Brown elaborate on a historical paper by Hans Kasturp in Doncel et al.'s *Symmetries in Physics* to confront with a careful and "friendly" presentation of Emmy Noether's 1918 mathematical work on variational invariants their physical interpretation in standard texts of Lagrangian dynamics. Noether's work is based on the variational problem of deriving the general conditions satisfied by a set of variables so that the first-order functional variation of their action functional vanishes for an arbitrary region of integration. The solution to this problem led to the so-called Noether's theorems deriving from global and local symmetries conservation laws and constraints, respectively. However, decry Brading and Brown, physics textbooks derive these mathematical results as theorems of physics, committing an interpretive faux pas. The authors also remind the reader that the identity constraints derived in Noether's second theorem (third by their count, including a boundary condition derived from the first theorem) show that local symmetry in this context entails underdetermination of the unknown independent variables (potentially describing the evolution of a system) by the fewer independent equations of motion. The next four papers point precisely to this result as the worrying relation between local symmetry and a failure of determinism.

John Norton's paper picks up on Brading and Brown's and Martin's problem of physical interpretation of symmetry and its consequences in the context of Einstein's General Theory of Relativity. In this context one speaks of covariance rather than invariance. In one way of drawing this infamously confusing distinction, invariance is said to describe the geometrical symmetry of the underlying space-time of the Special Theory of Relativity, whereas covariance describes the form invariance of the (gravitational) laws of the General Theory under arbitrary smooth coordinate transformations. The general covariance of the physical equations of the

https://doi.org/10.1086/516808 Published online by Cambridge University Press

General Theory have been taken, then, as an example of a gauge argument, in the "profundity view." By contrast, in a famous exchange with Einstein, Erich Kretchmann argued in 1917 that *all* space-time theories can be trivially given generally covariant formulations. In other words, the symmetry at hand is but a vacuous formal trick. General covariance, active or passive? Which is it? Norton's strategy is that of divide and conquer. For a given formulation and interpretation, the formal property will carry some physical meaning.

Michael Redhead begins his paper with a brief presentation of his well-known surplus-structure account of symmetries. He then, in a dramatic conclusion, touts the gauge principle as "the most pressing problem in contemporary philosophy of physics" (138). Well, he would. But here I beg to differ. As Martin and the author of this review have argued, much hinges on this issue on philosophical implicit (or explicit) agendas. Overarching and more fundamental questions of explanation, mathematical representation, and unification are, for instance, a factor. (The Wittgensteinian philosopher Friedrich Waissman more modestly argued that redundancy was key to linguistic communication). In the bulk of the paper, and in a valuable conclusion—which would have been a more helpful contribution as part of the introduction—Redhead surveys three approaches to interpreting gauge potentials, presents their respective philosophical advantages, and exposes their costs.

If one chooses to interpret gauge potentials as describing some previously unknown physical reality, then one can have a local theory, but the cost one must be prepared to pay is indeterminism. To prevent it, one would then be faced with the equally disagreeable choice of ad hoc hidden variables that fix the gauge. If one chooses to reformulate the theory in terms of gauge-invariant quantities, the price is nonlocality. If one allows a formulation in terms of non-gauge-invariant potentials, now presumably without physical interpretation, additional surplus structure will be needed to develop the theory (as the case of ghost fields in quantum field theory illustrates). The price of this move, he concludes, is "the mysterious, even mystical, Platonist-Pythagorean role for purely mathematical physics in theoretical physics" (138). Clearly, this is a non sequitur or else it is based on a distinctive philosophical perspective. This can be seen by contrast with nonmystical approaches such as the structural realism defended by French and Rickles (in the volume under review), with Van Fraassen's deflationary empiricist attitude toward the mathematics of symmetries (also in this volume), or with Martin's own discussion. On the whole, Redhead builds up the mystery in the form of a triadic version of a catch-22, in which we are faced with three allegedly impossible choices or philosophical dead ends. But that is simply a manifestation of his own philosophical approach and its cost. In science as in philosophy or politics

https://doi.org/10.1086/516808 Published online by Cambridge University Press

and much real life, there exists always an element whose name is legion: an Archimedian point, a blind spot, a residual price, an auxiliary motive, an extralogical factor, a decision without an argument, a presupposition, a constraint, a bias, a value, an ideology, an attitude, a taste, a virtue, a commitment, a skill, or lack of it, a preference, you name it. The best we can often do is to simply acknowledge it and its many forms (not just names). And that's precisely why science, even physics, has a developing history (*pace* Stephen Hawking). A commitment to determinism or locality as parts of the picture of the world is not the only or necessarily most desirable Archimedian points in philosophy of physics.

The last three papers of Part I are more technical and narrowly focused. John Earman defends in his paper the value of Dirac's constrained Hamiltonian formalism for making sense of gauge symmetries. This paper, as well as another in this volume, is an example of a Carnap-style philosophy (also embraced with notorious virtuosity by another philosopher of physics, David Malament). By that I mean an approach whose core lies in the naturalistic defense and application of a technical formalism or apparatus from physics itself which is to play the role of ideal language that, by reformulating problems or their theoretical setting, solves some and exposes and eliminates others as pseudo-problems.

Earman shows how the application of Dirac's formalism to gauge symmetry issues helps philosophers to economically and clearly bring out the question of the form of observable quantities (he links the two by stipulating that determinism can be restored by adopting gauge equivalence as a condition of physical identity and uniqueness), relationalism about space-time, indeterminism, the conditions for the applicability of fiber bundle formalism, and the problems with four approaches to the quantization of gauge theories.

In a concise and conceptually subtle, clear, and controlled paper, David Wallace takes up the question of the failure of determinism associated with theories manifesting local symmetry. He argues that by drawing a distinction between spatial and temporal notions of locality one can draw interesting distinctions between different kinds of violation of determinism by local symmetries. He shows that this result reveals different meanings of symmetry in local symmetries when introducing criteria to restore determinism in different kinds of theories—symmetry in electromagnetism as physical identity of many different configurations and symmetry in general relativity as physical identity of many different sequences of configurations.

This is a good place to warn the readers about the potential confusion raised by the fact that many articles in the volume offer what will appear as different definitions or technical characterizations of what symmetries are.

https://doi.org/10.1086/516808 Published online by Cambridge University Press

Antigoni Nounou takes up Redhead's question of the interpretation of gauge potentials in the context of the project of explaining the Aharonov-Bohm effect (in its semiclassical formulation). The effect is the shift in the interference pattern on a screen formed by electrons coming from two slits when a magnetic field is switched on inside a solenoid (only) between the two electron paths from the slits to the screen where the pattern forms. The shift is proportional to the magnetic flux between the trajectories. She reviews three current explanatory options, causal explanations, for this effect and defends a fourth, a "topological explanation" in terms of fiber bundles. The first of the three causal explanations takes the magnetic field inside the solenoid to be the causal agent of the distant shift on the screen, which she dismisses on grounds of nonlocality and nonseparability. The same problems accrues, she argues following Richard Healey, to the second, which claims that the gauge potential, the vector magnetic potential, is the local cause. Also the third explanation, which takes the phase factor, or holonomy, to be the real cause, violates nonseparability. The lesson she draws from these failures is that the adequate explanation cannot be both holistic (nonlocal or nonseparable) and causal. Instead she introduces the notion of topological explanation and suggests that this is the adequate explanation, which is provided by the fiber bundle formulation of the Aharonov-Bohm effect. Her defense of the fundamentality of this formalism stands at odds with Earman's defense of the constrained Hamiltonian one. Unfortunately, they do not engage each other's views. But this should provide the reader with a stimulating challenge. A difficulty with Nounou's argument is that she doesn't consider the possibility that the three explanations she rejects might not be necessarily causal, and she doesn't make a convincing case for the notion of topological explanation—not even looking for support from Graham Nerlich's notion of geometrical explanation.

My summary of the much shorter three other parts will also be shorter. Part II is devoted to discrete symmetries such as parity and permutation. This will be more intuitive territory for the lay readers, despite the less intuitive direction some of the discussions take. Three articles plumb the metaphysical depths on the Leibnizian question of individuality and permutation invariance or indistinguishability. What difference does it make if physical systems are redistributed over the possible properties or states? This has been one of the areas in which metaphysical lines have been drawn in the recent past between classical and quantum physics. Quantum systems—sorted by their statistical group behavior—have been thought to violate Leibniz's Principle of the Identity of Indiscernibles (PII) and be less than bona fide metaphysical individuals (to use the quaint Latin terminology that this scholastic line of metaphysics invites).

Steven French and Dean Rickles survey different views of the meta-

https://doi.org/10.1086/516808 Published online by Cambridge University Press

physical lessons from the statistical behavior of classical and quantum systems (objects/particles/individuals?) and their violation of PII, and defend a relational, structuralist reading of permutation symmetry: it is not a metaphysically derivative property reflecting the metaphysical nature of physical particles; rather, it is a reflection of a deeply structural feature of the world-as-web-of-relations introducing broad kind-distinctions. Nick Huggett argues that the scope of the violation of PII in the quantum world (with respect of observable quantities represented by Hermitian operators) can be extended to "quarticles," systems characterized by the more exotic, parastatistical behavior. Simon Saunders introduces and defends an updated logical notion of object and individuality and a corresponding reading of the principles of Leibniz's rationalist philosophy such as PII and the Principle of Sufficient Reason. While French and Rickle's mathematically formal, structural, framework focuses on relations, Saunders's logically formal one focuses on relata. Within this new logical framework he can then defend a weaker form of PII in terms of "weak discernability" based on irreflexive relations. Aside from the characteristic elusive intricateness of Saunders's essays, a lurking tension in this article appears between his neo-rationalist framework and his claim to a descriptive metaphysics, in Strawson's sense, applied only to established scientific practice.

Oliver Pooley contributes a long, informative, and insightful article about Kant's favorite symmetry, handedness or left-right symmetry. He argues for a relational account of handedness between incongruent parts and defends its compatibility with parity violation in Nature—as many philosophers call fundamental physics—distinguishing between left and right and also its lesson for the relationalism-substantivalism debate about the nature of space-time. In a brief and very original paper Nick Huggett picks up on Pooley's account of handedness and suggests an intriguing relationalist account of orientability and of space more generally by analogy with the Mill-Ramsey-Lewis account of nomological modality of natural laws without primitive necessities: as the most convenient—the simplest—way to describe the regular evolution of relations.

Part III, by far the shortest, deals with the philosophically much neglected topic of symmetry breaking, the key to the true unifying, predictive, and explanatory value of much fundamental and nonfundamental physics. The main idea, called spontaneous symmetry breaking, is that the systems we encounter are characterized by a degree of symmetry inferior to that of the laws or equations of evolution they obey or the action function that characterizes them. The system and its equations of motion can also be explicitly made asymmetric by the introduction of terms representing external factors.

Part III opens with a classic heterodox text by the physicist Pierre Curie

https://doi.org/10.1086/516808 Published online by Cambridge University Press

and a misplaced and unfortunate choice of another brief text by Weyl—the co-patron saint of symmetry physics with Wigner. This is followed by a brief personal account by Giovanni Jona-Lisinio of his role in the episode of cross-fertilization in physics in which the application of the idea of symmetry breaking in elementary particle physics was inspired by the application of this notion in condensed-matter physics of superconductivity. It is a short physicist-style history for physicists and is not nearly as informative and insightful as, for instance, the paper on the topic by L. Brown and T. Cao of 1991 in *Historical Studies in the Physical Sciences*.

Not a page too early, next is a thoughtful introduction to the concept of symmetry breaking in classical and quantum field physics by one of the editors, Elena Castellani. She concludes the piece with several philosophical questions, duly acknowledging inspiration by Curie's ideas: for instance, Curie's principle, that the symmetry of the effects must be found in the causes and his dictum that asymmetry creates the phenomenon—that is, a distinctively structured system. Her characterization of broken symmetries in terms of hidden symmetries and other conceptual and epistemological remarks cannot, however, escape the dominating tone of the symmetry-centric paradigm. John Earman contributes another characteristic Carnap-style paper aiming to clarify fundamental ideas and foundational issues: in this case defending as ideal language the algebraic formulation of quantum field theory to represent key ideas of symmetry and symmetry breaking and to solve and dissolve a number of technical puzzles he raises about degeneracy of the vacuum states and unitarity of equivalence of representations of algebras of observables. In a less technical article Margaret Morrison begins with a historical narrative about the introduction of the notion of symmetry breaking in particle physics and addresses the epistemological issues of realism and the problematic theoretical status and empirical support of the symmetry-breaking mechanisms in the Standard Model of unified interactions. The argument dovetails with Martin's criticism of some defenses of gauge symmetry principles as fundamental and normative (also in this volume).

Part IV deals with general epistemological and methodological issues of symmetry thinking. Surprisingly it is also philosophically the least original and compelling. Jennan Ismael and Bas van Fraassen sketch out a qualitative account of symmetries as a model of theory construction. Their role consists in guiding the elimination of what they call superfluous structure. By contrast with Redhead's notion, here this is implicitly and normatively taken for granted as an episode in a narrow empiricist methodology only tangential to actual scientific practice in the spirit of van Fraassen's well-known brand of empiricism. In a technically more formal paper Gordon Belot models notions of symmetry in order to provide a methodological account of problem solving. He illustrates with sketches

https://doi.org/10.1086/516808 Published online by Cambridge University Press

of episodes from the history of physics that, to the lay reader's surprise and relief, involve none of the technicalities. He concludes with a more technical account modeled after the constrained Hamiltonian formalism in terms of what he calls quotienting the space of solutions. A suggestive, fast, and casually toned (and argued) article by Peter Kosso discusses the relation between symmetry and the notions of objectivity and design. Kosso loosely identifies, after the manner of the late Robert Nozick, the idea of objectivity with perspective-independence, in a loose Einsteinian analogy, and with invariance, hence with symmetry. A principle of symmetry is a principle of objectivity, and this, he argues, applies only to laws. Then the relativistic analogy takes a fatal linguistic turn and Kosso identifies subjectivity with language choice and objectivity with the alleged fact that "once a particular language is chosen, nature itself constrains the possible combinations of words" (516)! He tries next with the analogy to plurality of languages—that is, subjectivity—and invariance of meaning translation—objectivity. He then proceeds to characterize objectivity as lack of intentional control and opposes it to design, or the application of detailed, creative intentional control. Design is also associated with structure, and . . . what else? Symmetry breaking. The reader will be surprised—or alarmed—by how this identification works effectively as an argument against natural theology. At the very least it is supposed to shed some insight into debates about realism, although clearly not about structures and not outside physics. Part IV ends with one of the editor's, Castellani's, mercifully sensible and comprehensive conceptual introduction to a general notion of symmetry in the intuitive terms of equality, equivalence, arbitrariness, irrelevance, and redundancy. The piece attempts to map out some of the ideas in the preceding articles in the volume. But some of its insights prove less effectively illuminating in the conclusion than they might have in the introduction. Somewhere between misleading and redundant is her claim that "symmetry identifies and preserves structure" (529). The more interesting suggestions are the association of symmetries with scales of representation and the contextual nature of arbitrariness and irrelevance.

All in all, this important volume should both stimulate and challenge the reader with its issues and debates; it should also provide some of the much needed help with some of the required background and the many and up-to-date bibliographical references.

https://doi.org/10.1086/516808 Published online by Cambridge University Press