Laws, Symmetry, and Symmetry Breaking: Invariance, Conservation Principles, and Objectivity

John Earman†‡

Given its importance in modern physics, philosophers of science have paid surprisingly little attention to the subject of symmetries and invariances, and they have largely neglected the subtopic of symmetry breaking. I illustrate how the topic of laws and symmetries brings into fruitful interaction technical issues in physics and mathematics with both methodological issues in philosophy of science, such as the status of laws of physics, and metaphysical issues, such as the nature of objectivity.

1. Introduction. The focus of this address is on the web of connections that tie together laws, symmetries and invariances, and conservation principles. There are many ways to pursue this topic. My line of pursuit will be somewhat unorthodox, but it has the virtue of connecting a number of fundamental issues in the foundations of physics. Reflecting on these issues prompts a reevaluation of basic issues in metaphysics, such as the nature of objectivity and the nature of change. Perhaps because of the formidable technical challenges they pose, philosophers have tended to shy away from the foundations of physics problems that I will identify. And perhaps because they are embarrassed to be seen to be doing metaphysics, philosophers of science have been reluctant to take up the philosophical issues. My message to both groups is the same: Have courage! The nub of the issues in foundations of physics can be made accessible even to the non-specialist. And there is no shame in doing metaphysics

†To contact the author, please write to: Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260; e-mail: jearman@pitt.edu.

‡The ungainly title is due to my desire to do homage to my predecessor, Bas van Fraassen, and my teacher, Robert Nozick. The version printed here is a brief summary of some of the major themes of the longer paper of the same title posted on the PhilSci Archive http://philsci-archive.pitt.edu.

Philosophy of Science, 71 (December 2004) pp. 1227–1241. 0031-8248/2004/7105-0050$10.00

as long as the activity is informed by scientific practice. I will be begin with a brief look at the philosophical literature on laws.

2. Laws: A Scandal in the Philosophy of Science. It is hard to imagine how there could be more disagreement about the fundamentals of the concept of laws of nature—or any other concept so basic to the philosophy of science—than currently exists in philosophy. A cursory survey of the recent literature reveals the following oppositions (among others): there are no laws versus there are/must be laws; laws express relations among universals versus laws do not express such relations; laws are not/cannot be Humean supervenient versus laws are/must be Humean supervenient; laws do not/cannot contain *ceteris paribus* clauses versus laws do/must contain *ceteris paribus* clauses.

One might shrug off this situation with the remark that in philosophy disagreement is par for the course. But the correct characterization of this situation seems to me to be "disarray" rather than "disagreement." Moreover, much of the philosophical discussion of laws seems disconnected from the practice and substance of science: scientists overhearing typical philosophical debates about laws would take away the impression of scholasticism—and they would be right!

There is, however, one place where the philosophical investigation of the concept of laws can make solid contact with science, and that place is to be found in the topic of laws and symmetries. Philosophers of science have done some good work on this topic (e.g., van Fraassen 1989), but it is only a beginning. And the surface of many important subtopics, such as gauge symmetries and symmetry breaking, has barely been scratched. I will emphasize these neglected topics here.

3. Symmetries and Laws: Laws of Nature versus Laws of Science. The topic of symmetries and invariances in physics provides no comfort for those who hanker after *laws of nature* in the sense of critters that embody the goodies on the wish list drawn up by philosophers: laws of nature are supposed to be objective (independent of our interests and beliefs); they are supposed to express a strong form of non-logical necessity ("nomological necessity"); they are supposed to cut nature at the joints by expressing truths about natural kinds; they are supposed to have the power to explain the why of things; etc. Here I declare myself in sympathy with the milder form of the "no-laws" view: if there are such critters, I see no good reason to think that science can be counted on to corral them, or that we can tell the difference between cases where science has succeeded in corralling them and cases where it has failed to do so.[1]

But by the same token the topic of symmetries and invariances does not support the strong version of the no-laws view, which intimates that both the history of science and its current practice can be understood without giving pride of place to the search for *laws of science*. In the case of physics—which will be my focus—what physicists mean by the *laws of physics* is, roughly, a set of true principles that form a strong but simple and unified system that can be used to predict and explain. As Steven Weinberg puts it

> Our job as physicists is to see things simply, to understand a great many complicated phenomena in a unified way, in terms of a few simple principles.[2] (1980, 515)

Rather than coming at the topic of laws of physics with preconceptions of what these laws must deliver if they are to support favored philosophical accounts of causation, counterfactuals, explanation, etc., historians and philosophers of science would do better to investigate how physicists use the concept of law. It would not be surprising to find that there is nothing neat and precise that corresponds to Weinberg's "few simple principles." But so what? To dismiss the notion of laws of physics on the grounds that it is messy and imprecise would be to miss important points not only about the motivation of physicists but about the methodology and content of physics as well. In particular, the relevant senses of symmetry and invariance in physics *presuppose* a distinction between what holds as a consequence of the laws of physics and what is compatible with but does not follow directly from these laws.

This is hardly a novel idea. It pervades Eugene Wigner's writings on symmetries and invariances. Consider, for instance:

> It would be very difficult to find a meaning for invariance principles if the two categories of our knowledge of the physical world [laws vs. initial/boundary conditions] could no longer be sharply drawn. (Houtappel, Van Dam, and Wigner 1965, 596)[3]

1. See van Fraassen (1989) and Giere (1999) for different versions of the "no-laws" view.

2. The reader acquainted with the philosophical literature on laws will notice a resemblance between Weinberg's notion and David Lewis's (1973) analysis of laws as the axioms or theorems that belong to the best deductive system, where "best" means achieves the optimal balance between strength and simplicity. On this account of laws, the distinction between accidental and lawful regularities does not presuppose or commit one to non-Humean necessity.

3. For additional expressions of this idea, see Wigner (1967).

The point is that the symmetries of most concern to physicists—Poincaré invariance, time reversal invariance, etc.—are typically broken by the actual phenomena (see Curie 1894). But, one might ask, if this is so why do physicists set such store by these symmetries? The obvious—and I think largely correct—answer is that research in physics is guided by Weinberg's injunction to understand complicated phenomena in terms of a few simple principles—that get dubbed the laws of physics—and that physicists have, or think they have, reasons to believe that the laws of physics do and, perhaps, must reflect the symmetry at issue. I will have more to say about the second part of the answer in the following section.

4. The Status of Symmetry Principles. The received wisdom about the status of symmetry principles has it that one must confront a choice between the a posteriori approach (a.k.a. the bottom up approach) versus the a priori approach (a.k.a. the top down approach). The former approach means that we subject candidate laws of physics to empirical checks and then derive the symmetries from the candidates that have passed muster. The latter approach means that symmetry principles are viewed as being more fundamental than the laws they constrain or as being second order laws that dictate symmetries to first order laws. The choice on offer is not an either/or one. That symmetry principles have a meta-character follows from the characterization of symmetries of laws given in section 2. But viewing symmetry principles as meta-laws doesn't commit one to treating them a priori in the sense of known to be true independently of experience. For instance, that a symmetry principle functions as a valid meta-law can be known a posteriori by a second level induction on the character of first-order law candidates that have passed empirical muster. From the other direction, the a priori of the top-down approach doesn't have to be understood either in the sense of necessarily true (or true for all times) or in the sense of knowable independently of experience; rather it can be understood in the sense of a revisable constitutive a priori.

The last remark applies to the symmetries of physical laws deriving from the symmetries of spacetime in the pre-general relativistic era. During that innocent era the favored spacetime—say, neo-Newtonian spacetime or Minkowski spacetime—was supposed to be constitutive of physical possibility in that is was supposed to serve as the fixed backdrop for any acceptable theory of physics.[4] Then if the laws of physics are formulated in terms of geometric object fields on spacetime and if laws are to be "general" or "universal" in the minimal sense that they cannot use names or designators for particular spacetime points or regions, it follows that

4. Here I am borrowing from—and distorting—Reichenbach (1965, ch. 5); see also Friedman (1999, ch. 3).

any acceptable candidate for a law of physics must share the symmetries of the spacetime.

The force of these considerations is illustrated in two ways. First, they help to explain why Huygens and Leibniz were morally certain that Descartes' candidates for laws of elastic impact cannot be laws of physics, despite the fact that neither Huygens nor Leibniz had done the relevant experiments: Descartes' 'laws' lack a symmetry property they must have if laws are to be set in the spacetime structures preferred by Huygens and Leibniz. Second, the considerations show how various strands of the debate over absolute versus relational accounts of space, time, and motion are tied together; in particular, if one wants to allow for the possibility of determinism, then holding a relational account of the motion of bodies forces one to be a relationist about space (and spacetime): a relational account of motion requires that the spacetime setting has such a large symmetry group that determinism cannot possibly hold on a "container" view of space.

There are two limitations to the above thesis about the a priori status of symmetry principles. The first is that it does not apply to "internal" or non-spacetime symmetries. Of course, one could try to tell a parallel story about how the structure of internal spaces serves as the grounds of a constitutive a priori for internal symmetries, but such a story does not have the ring of plausibility. The second is that even for spacetime symmetries it does not survive far into the twentieth century. Both the notion of spacetime as the grounds for a constitutive a priori and the above way of connecting the symmetries of spacetime and the symmetries of laws disintegrate with the advent of Einstein's general theory of relativity (GTR). In GTR no one spacetime serves as the fixed backdrop for physics since different spacetime structures belong to different solutions to Einstein's field equations. Moreover, most of these spacetimes lack any nontrivial symmetries on either the global or local level. There is still a sense in which Einstein's gravitational field equations satisfy a strong symmetry principle; but, contrary to what Einstein originally thought, this symmetry principle is not a relativity principle that generalizes to arbitrary reference frames the special principle of relativity that is implemented in Newtonian theories by Galilean invariance and in special relativistic theories by Lorentz invariance. Rather the symmetry principle satisfied by GTR is a gauge principle, about which I will have more to say below in sections 6 and 7. In general, the radical nature of the change GTR necessitates in our conception of symmetry principles is an underappreciated moral.

5. Symmetries and Conservation Principles: Noether's First Theorem. Do symmetries of laws of motion entail conservation principles and conversely? The answer is yes as shown by Emmy Noether first theorem, *if* the laws are in the form of differential equations that are derivable from an action principle (and thus can be written in Euler-Lagrange (EL) form), *if* the symmetries are continuous (technically, they form a finite parameter Lie group), and *if* the symmetries are variational symmetries (and thus carry solutions of the EL equations to solutions).

The vast majority of candidates for fundamental laws of motion in physics satisfy the crucial first condition of the antecedent. But this fact may represent an artifact of scientific theorizing rather than a fundamental feature of nature; for physicists choose equations of motion with an eye to quantization, and the standard route to quantization is via a Hamiltonian formulation, which can be produced once the a Lagrangian formulation is in hand. Necessary and sufficient conditions for a Lagrangian formulation are known for some classes of equations of motion, but in general we are in a state of ignorance about how hard, or easy, it is to satisfy the first antecedent condition.

Noether's first theorem shows that, under the stated conditions, a variational symmetry gives rise to a conserved current, and vice versa. When the independent variables of the action are those of space and time, the time component of a Noether current can be integrated over space to give a Noether charge, and under appropriate boundary conditions this charge can be shown to be constant over time. By way of illustration, the standard action principle for Newtonian particle mechanics admits as variational symmetries the elements of the inhomogeneous Galilean group, and the application of Noether's first theorem to this case yields the conservation of energy, angular and linear momentum, and the uniform motion of the center of mass.

6. Invariance and Objectivity: Noether's Second Theorem. The overarching theme of Nozick's *Invariances: The Structure of the Objective World* (2001) is that invariance is the root of objectivity: the familiar marks of objectivity—accessibility from different angles, intersubjectivity, and independence from people's beliefs and desires—are all to be explained in terms of invariances. This is a potentially powerful theme. But actual as opposed to potential power can only come from specificity: if objectivity is to be construed as invariance, we need to know what it is that can be, or fail to be invariant, and under what transformations the said things must be invariant if they are to capture objective features of nature. Naturally, Nozick has a good deal to say about these questions, not all of which is successful.

For instance, I think Nozick reaches too hard in trying to make a connection between objectivity and conservation principles:

> Emmy Noether showed that for each symmetry/invariance that sat-

> isfies a Lie group, there is some quantity that is conserved . . . . So [!] it is not surprising that laws that are invariant under various transformations are held to be more objective. Such laws correspond to a quantity that is conserved, and something whose amount in this universe cannot be altered, diminished, or augmented should count as (at least tied for being) the most objective thing there actually is. (81)

Here I must take exception with my teacher. For I do not see that it makes much sense to speak of gradations of objectivity of laws. And even if it did, I don't see how the invariances of laws could provide a means of assigning the gradations. And while conserved quantities may play a special role in a theory of motion, I do not see why a quantity that is conserved is any more objective or real than one whose value changes with time.

Nevertheless, I think that, when properly interpreted, Nozick's theme that *objectivity* = *invariance* has great merit. Where Nozick went wrong was in focusing on the symmetries of Noether's first theorem rather than those of her second theorem.

The second Noether theorem concerns the case where the action is invariant under a Lie group of transformations whose parameters are arbitrary functions of all of the independent variables. The second theorem then tells us that the Euler-Lagrange equations are not independent and, hence, the solutions are underdetermined. When the independent variables are those of space and time the underdeterminism amounts to an apparent breakdown of Laplacian determinism. This can be seen more explicitly by noting that in the cases at issue arbitrary functions of time appear in solutions of the Euler-Lagrange equations so that a unique solution is not picked out by initial data. I say there is an *apparent* breakdown in determinism because the option remains open to blame the appearance of a breakdown on a redundancy in the descriptive apparatus of the theory in the sense that the correspondence between the state descriptions given in the theory and the "real" or "objective" state of affairs is many-one. In particular, one can take the elements of the symmetry group to be gauge transformations in that they relate different descriptions of the same physical state rather than different states. The objective facts about a possible world (as described by the theory) are precisely those that can be stated in gauge-invariant terms.

There is an apparatus—called the constrained Hamiltonian formalism—that applies to any theory whose equations of motion can be derived from a variational principle and that gives a principled way of identifying the gauge freedom of the theory and of characterizing the "observables" or gauge independent quantities of the theory.[5]

7. Gauge, Objectivity, and the General Theory of Relativity. The implementation of part of Nozick's formula *objectivity = invariance* by means of the constrained Hamiltonian formalism goes swimmingly: in case after case it yields intuitively satisfying results. But the application to Einstein's GTR yields some surprising and seemingly unpalatable consequences. One is that the apparatus says that motion according to Einstein's gravitational field equations is pure gauge and, thus, that the observables are constants of the motion. Some philosophers and physicists have found this "frozen dynamics" so bizarre that they think it shows that the constraint apparatus, which is otherwise so fruitful and successful in other domains, has gone haywire when applied to GTR. But one lesson of twentieth century physics is that results that initially shock our intuitions often have to be accommodated in the scientific image. In keeping with this lesson I want to suggest that much is to learned from trying to accommodate rather than dismissing the result in question (see Earman 2002a).

The problem of time and change in GTR is one aspect of a more general interpretation problem of a kind that philosophers of science claim to take to heart but have shied away from in the case of GTR. Whatever else it means to interpret a scientific theory, it means saying what the world would have to be like if the theory is true, and this in turn means specifying which quantities the theory takes to be "observables" in the sense of genuine physical magnitudes and under what circumstances these quantities take on values. In ordinary QM it is assumed we know more or less how to characterize the observables, and most of the interpretational angst is vested in the problem of how to assign values to these quantities in such a way as not to be so profligate as to run into impossibility results of the Kochen-Specker type or so parsimonious as to be unable to account for the outcomes of measurements. In GTR the situation is, so to speak, just the reverse. There is no problem about a value assignment rule: all observables always take on definite values. The problem is rather to construct the observables. Those who know a bit about GTR might guess that we can make a beginning on the construction by starting with "scalar invariants," e.g. things like curvature scalars. But on the line I am pursuing these quantities cannot count as observables since they are not gauge invariant (diffeomorphic invariant) quantities. The gauge invariants constructible from the basic dynamical variables of

5. This approach to gauge was developed independently by P. A. M. Dirac and Peter Bergmann. For an authoritative overview, see Henneaux and Teitelboim (1989). For a more user friendly introduction, see Earman (2002b, 2003a).

GTR include highly non-local quantities such as the four-volume integral of the Ricci scalar curvature over all spacetime (assuming such an integral converges), but obviously such quantities are not very useful in describing the outcomes of typical measurements and observations. Another class of diffeomorphic invariants is comprised by what can be called coincidence quantities, a name chosen to reflect the fact such quantities are the counterparts for fields of Einstein's "point coincidences" for material particles. To illustrate, consider a solution to Einstein's field equations with the generic property that the spacetime metric does not have non-trivial symmetries. In that case the spacetime manifold can be coordinatized by the values of four scalar fields constructed from the metric and its derivatives. Then, for example, the taking on of the electromagnetic field of such-and-such a value coincident with the four scalar fields having values such-and-so is a diffeomorphic invariant. Note that an ontology comprised of such coincidence events is rather strange. We are used to thinking of an event as the taking on (or losing) of a property by a subject, whether that subject is a concrete object or an immaterial spacetime point or region. But the coincidence events in question are apparently subjectless. Note also that one doesn't verify the occurrence of a coincidence event by first measuring the values of the electromagmetic and the scalar fields in question, and then verifying that the required coincidence of the value of the former with the latter does indeed hold; for by themselves none of these fields are gauge invariant quantities and so cannot be measured. The verifying measurement has to respond directly to the coincidence. What this implication means for measurement and observation obviously requires spelling out, a task I cannot undertake here.

These strange features may be an indication that the interpretational stance I have suggested is on the wrong track. But it is surprising (and disappointing!) to me that philosophers of science think they can know this a priori. I propose that one way of testing an interpretational stance for classical GTR is to see how well the stance lends itself to promoting a marriage of GTR and quantum physics that issues in a successful quantum theory of gravity. And here I would like to correct the impression conveyed by the popular media that string story (or M-theory, or whatever it is now calling itself) is the only viable route to a quantum theory of gravity. In fact, the loop formulation of quantum gravity–which uses the interpretational stance I have been pushing–is a viable program. In particular, in contrast to M(ystery)-theory it is a genuine theory rather than a wannabe theory, and it has enjoyed theoretical success (e.g., explanation of black hole entropy).[6] Furthermore, it may be technically feasible to

6. For a review of loop quantum gravity, see Rovelli (1998).

test its predictions in the near future. If this approach to quantum gravity falters for reasons connected with the suggested interpretational stance, then that stance is disconfirmed. But, to repeat, if philosophers think that they can prove a priori that this will happen, they have an obligation to submit their results to the *Physical Review* so as to kill off a non-viable program.

8. The Cosmological Constant, the Fate of the Universe, and Change. The cosmological constant Λ has had a long and checkered history since its introduction into GTR in 1917, with periods where Λ plays an important role in cosmology alternating with periods where it pushed off stage (see Earman 2001). We are presently in a period where Λ—or some surrogate for Λ—is holding center stage. Recent observations of Type Ia supernovae indicate that the rate of expansion of the universe is increasing and, thus, that either $\Lambda > 0$ or else the universe is dominated by a strange form of matter-energy (commonly dubbed "quintessence") which exerts a sufficiently negative pressure as to mimic the action of a positive Λ.

For present purposes I will assume that Λ rather than quintessence is at work and will note two implications of this assumption. The first is that if a positive cosmological constant is indeed responsible for the speeding up of the expansion of the universe, then the universe will not end in a big crunch but will expand forever. The second implication concerns the issues discussed in the preceding section. The connection is made by asking the seemingly naive question: In what sense is the cosmological constant a constant? It must be a constant in the sense that it has the same value throughout spacetime, for this is necessary in order that Einstein's field equations with cosmological constant imply the local energy conservation law in the form of the vanishing of the covariant divergence of the stress-energy tensor. But there is a further sense in which Λ must be a constant, at least if the standard derivation of Einstein's field equations from a variational principle is followed. For that derivation implies that, on pain of setting the volume of spacetime to 0, Λ is not a dynamical variable in the sense that it does not vary from solution to solution.

However, there is nothing to prevent the cosmological constant from being treated as a spacetime constant within each solution but having a value that varies from solution to solution—in effect, the cosmological constant is treated as a constant of integration rather than a new fundamental constant of nature. I will use the lower case λ to denote this sense. But recall that we are demanding that a candidate for a fundamental law of motion must be derivable from an action principle. Applying this demand to the λ version of Einstein's field equations leads to some interesting consequences. In particular, it is found that the derivation requires that spacetime of standard GTR be enriched by the addition of new object fields, and when the resulting action is run through the constrained Hamiltonian formalism it is found that the class of observables of λ-GTR is richer than for Λ-GTR (see Earman 2003b for details). In fact, in λ-GTR the dynamics is "unfrozen" in that there are gauge independent quantities that are not constants of the motion. This finding caused a flurry of excitement in the late 1980's in the quantum gravity community because it was thought that λ-GTR would overcome some of the obstacles in the path of the canonical quantization program. But when these hopes were dashed because of technical difficulties, physicists bent on finding a quantum theory of gravity quickly lost interest. Nevertheless Λ vs. λ remains for philosophers of science an interesting illustration of the interconnections among action principles, constraints, gauge principles, observables, etc., and it illustrates the power of the analytical apparatus I have been touting to reveal these interconnections.

9. Spontaneous Symmetry Breaking. My final topic brings together several of the themes discussed above–the breaking of a lawlike symmetry by particular states, conservation laws, the Noether's theorems, and gauge freedom.

Consider a Lagrangian for a classical field admitting symmetries that form, say, a one-parameter Lie group. We know by Noether's first theorem that there is an associated conserved current. Now suppose that the field is quantized by giving a Fock space representation where there is a distinguished state (the "vacuum state") which gives the ground state of the quantum field and from which excited states are built up by applying creation operators. One can ask whether the action of the one-parameter symmetry group of the Lagrangian[7] can be represented by a one-parameter group of unitary operators on the Fock space. Under very mild and reasonable assumptions the answer can be shown to be in the negative. If there were such a unitary group its generator would be a self-adjoint operator $\hat{Q}$ corresponding to the global Noether charge Q obtained by integrating the time component of the conserved Noether current over all space. But a simple reductio argument shows that if the vacuum is translationally invariant and if $\hat{Q}$ commutes with translations, then the existence of $\hat{Q}$ leads to contradiction (see Earman 2004). This result is puzzling to intuitions trained in ordinary QM where "symmetry transformation" and "unitary transformation"are virtually synonymous.[8] The

7. From here on by "symmetry" I mean internal symmetry. For example, if the Lagrangian $\mathcal{L}(\varphi)$ for a real-valued scalar field φ is $\partial^{\mu}\varphi\partial_{\mu}\varphi$, then it admits the one-parameter group of internal symmetries $\varphi \rightarrow \varphi' = \varphi + \beta$, $\beta =$ const.

8. I say "virtually" because some discrete symmetries, such as time reversal, are implemented by anti-unitary operators.

https://doi.org/10.1086/428016 Published online by Cambridge University Press

puzzle deepens when one realizes that the non-unitary implementability of the symmetry leads to the "degeneracy" of the vacuum state, for the vacuum state is supposed to be the unique Poincaré invariant state. Part of the puzzle is resolved by noting that this uniqueness assertion is not contradicted by the relevant sense of degeneracy, which means that there are many unitarily inequivalent representations of the canonical commutation relations of the field algebra, each with its own unique vacuum state. But again, this information is unhelpful to someone operating on intuitions trained on ordinary QM where no such phenomenon can arise.

In the longer version of this paper I indicate how the algebraic formulation of QFT can be used to take the puzzlement out of spontaneous symmetry breaking. In this formulation symmetries are taken to be automorphisms of algebras of observables, the relevant algebras being Weyl algebras that code up the familiar canonical commutation relations. For the finite dimensional Weyl algebras encountered in ordinary QM, any automorphism—and in particular an automorphism induced by a symmetry of the Lagrangian—is always implementable by a unitary transformation on a Hilbert space representation of the algebra. But for the infinite dimensional Weyl algebras relevant to QFT an automorphism can fail to be unitarily implementable, and such an automorphism will lead to unitarily inequivalent representations of the canonical commutation relations (see Earman 2004).

If this were all there was to the story of spontaneous symmetry breaking it would already hold interesting morals for the foundations of QFT. But there was much more to come. Before various ideas in elementary particle physics could coalesce to form what became known as the Standard Model it was necessary to find a mechanism by which the particles could acquire their mass. It turned out that the answer was suggested by a means of avoiding an embarrassing consequence of spontaneous symmetry breaking. It was discovered that the spontaneous breaking of a continuous symmetry subject to Noether's first theorem, together with some standard assumptions of QFT—such as Poincaré invariance, local commutativity, and the spectrum condition—implies the existence of "Goldstone bosons" (massless scalar bosons). Since there was very good evidence that such particles do not exist, it seemed that either spontaneous symmetry breaking has to go or else there has to be some radical modification in the way the business of QFT was conducted. A way out of this uncomfortable situation was found by Peter Higgs, who suggested, in effect, that the problem be changed. He showed that by introducing additional fields the symmetry group of the Lagrangian could be enlarged to an infinite dimensional Lie group whose parameters are arbitrary functions of the spacetime variables. One now is in the domain of Noether's second theorem and gauge transformations. Higgs further showed that the gauge could be chosen so that the Goldstone bosons are suppressed and that in this "unitary gauge" the new field had acquired a mass. As the semi-popular presentations put it, "Particles get their masses by eating the Higgs field."

Readers of *Scientific American* can be satisfied with these just-so stories. But philosophers of science should not be. For a genuine property like mass cannot be gained by eating descriptive fluff, which is just what gauge is. Philosophers of science should be asking the Nozick question: What is the objective (i.e., gauge invariant) structure of the world corresponding to the gauge theory presented in the Higgs mechanism? From the above discussion we know that there is in principle a way to answer this question; namely, apply the constraint formalism. When the shift is made from the Lagrangian to the Hamiltonian formulation, constraints will appear; find all of the constraints and single out the first class constraints; quotient out the gauge orbits generated by the first class constraints to get the reduced Hamiltonian phase space whose phase functions are gauge-invariant magnitudes; finally, quantize the unconstrained system to get a quantum field theoretic description stripped of surplus gauge structure. To my knowledge this program has not been carried out. To indicate why it is important to carry it out, consider the following three-tiered dilemma. *First tier*: Either the gauge invariant content of the Higgs mechanism is described by local quantum fields satisfying the standard assumptions of Poincaré invariance, local commutativity, spectrum condition, etc. or not. If not, then the implementation of the Higgs mechanism requires a major overhaul of conventional QFT. If so, go to the second tier. *Second tier*: Either the gauge invariant system admits a finite dimensional Lie group as an internal symmetry group or not. If so, Noether's first theorem applies again. But (since the other standard assumptions are in place) Goldstone's theorem also applies and, hence, Goldstone bosons have not been suppressed after all. If not, go to the third tier. *Third tier*: Either the gauge invariant system admits no non-trivial symmetries at all or else it admits only discrete symmetries. In either case Goldstone bosons are quashed. In the former case spontaneous symmetry breaking is not an issue since there is no symmetry to break. In the latter case it is possible that the discrete symmetry is spontaneously broken. But the usual argument for symmetry breaking using the conserved Noether current does not apply. And while it is possible that some completely different sort of construction will demonstrate the spontaneous breakdown of the hypothesized discrete symmetry there are no extant demonstrations that have more than a hand waving force.

10. Conclusion. Philosophers of science have barely scratched the surface of the topic of laws, symmetries, and symmetry breaking. What I find most attractive about this topic is that it brings into fruitful interaction issues from metaphysics, from mathematics and physics, from the philosophy of scientific methodology, and from foundations of physics. By the same token, the fact that all these issues are put into play means that the discussion is very difficult to control and that it is always in danger of getting lost in thickets of technicalia or degenerating into mush. Successfully confronting these dangers requires someone who understands and cares about the philosophy and who not only has a command of the mathematics and physics but can use it to illuminate and advance philosophical concerns. There are young people with these abilities. To them I say: The road ahead will be filled with tribulations and obstacles (not the least of which will be some of your colleagues), and it is uncertain what professional reward, if any, you will earn from traveling this road. But unless some of you have the courage to make the journey, the discipline will be immeasurably poorer.

REFERENCES

Curie, Pierre (1894), "Sur la symmétrie dans les phénomène physiques: Symmétrie d'un champ électrique et d'un champ magnétique", *Journal de Physique* 3: 393–415.

Earman, John (2001), "Lambda: The Constant That Refuses to Die", *Archive for History of the Exact Sciences* 55: 189–220.

——— (2002a), "Thoroughly Modern McTaggart: Or, What McTaggart Would Have Said if He Had Read the General Theory of Relativity", *Philosophers' Imprint*, 2. http://www.philosophersimprint.org.

——— (2002b), "Gauge Matters", *Philosophy of Science* 69 (Proceedings): S209–S220.

——— (2003a), "Tracking Down Gauge: An Ode to the Constrained Hamiltonian Formalism", in Katherine Brading and Elena Castellani (eds.), *Symmetries in Physics*. Cambridge: Cambridge University Press, 140–162.

——— (2003b), "The Cosmological Constant, the Fate of the Universe, Unimodular Gravity, and All That", *Studies in the History and Philosophy of Modern Physics* 34 B: 557–577.

——— (2004), "Curie's Principle and Spontaneous Symmetry Breaking", *International Studies in the Philosophy of Science* 18: 173–198.

Friedman, Michael (1999), *Reconsidering Logical Positivism*. Cambridge: Cambridge University Press.

Giere, Ronald (1999), *Science without Laws*. Chicago: University of Chicago Press.

Henneaux, Marc, and Claudio Teitelboim (1989), *Quantization of Gauge Theories*. Princeton, NJ: Princeton University Press.

Houtappel, R. M. F., H. Van Dam, and E. P. Wigner (1965), "The Conceptual Basis and Use of the Geometric Invariance Principles", *Reviews of Modern Physics* 37: 595–632.

Lewis, David K. (1973), *Counterfactuals*. Cambridge, MA: Harvard University Press.

Nozick, Robert (2001), *Invariances: The Structure of the Objective World*. Cambridge, MA: Belnap Press.

Reichenbach, Hans (1965), *The Theory of Relativity and A Priori Knowledge*. Los Angeles: University of California Press.

Rovelli, Carlo (1998), "Loop Quantum Gravity", *Living Reviews in Relativity*. http://www.livingreviews.org/.

van Fraassen, Bas C. (1989), *Laws and Symmetry*. Oxford: Clarendon Press.

Weinberg, Steven (1980), "Conceptual Foundations of the Unified Theory of Weak and Electromagnetic Interactions", *Reviews of Modern Physics* 52: 515–523.

Wigner, Eugene P. (1967), *Symmetries and Reflections*. Bloomington: Indiana University Press.