Dualities and intertheoretic relations

Elena Castellani[*]

1   Introduction

The idea of duality is at the core of the most relevant developments in recent fundamental physics. During the last forty years theoretical physics has used the notion of duality in different ways and frameworks: in the so-called dual resonance model of the late sixties, which gave birth to early string theory; in the context of quantum field theory, where a groundbreaking generalization of electromagnetic duality was conjectured by Claus Montonen and David Olive in 1977; in supersymmetric string theory, where various sorts of dualities are playing a key role in the theoretical elaboration.

This paper is concerned with the significance of physical dualities from the viewpoint of philosophy of science. The idea is that, for its peculiarity, this 'new' ingredient in theory construction can open unexpected perspectives for the current philosophical reflection on contemporary physics.[1] In particular, dualities represent an unusual type of intertheory relation, the meaning of which deserves to be investigated. It is the aim of the paper to show how discussing this point brings into play, at the same time, what is intended by a 'theory' and in which sense dualities are to be considered 'symmetries' (if they are).

---

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

[1]The philosophical literature on physical dualities is still very meagre. The philosophers of physics are just starting to turn their attention to string theory and its forms of dualities. One of the few contributions in this direction is Dawid (2007).

Considering the role and meaning of physical dualities in general poses immediately a problem. The dualities applied in recent fundamental physics are of different forms and status. While some of them seem to have a sound basis, others are just theoretical conjectures and a good part of the last developments grounded on dualities are still at a work-in-progress stage. Nonetheless, in most of the cases where dualities are applied in a quantum framework it is possible to individuate some common relevant characteristic features. A duality type that results particularly representative from this point of view is the so-called *electromagnetic duality* (EM duality). EM duality also represents the first form of duality explicitly applied in twentieth century physics: namely, in P. A.M. Dirac's famous two papers (published, respectively, in 1931 and 1948) on his 'theory of magnetic poles'. It therefore offers an appropriate, however specific, case study to begin with. Starting to investigate the significance of physical dualities by focussing on this case study is the object of the paper.

2 The case of electromagnetic duality

Electromagnetic duality as formulated by Dirac is, in a sense, the prototype of today's physical dualities. In this Section we present a brief survey of the development of this duality idea from the classical to the quantum context.

2.1 EM duality (1): classical electrodynamics

EM duality is grounded on the idea that there is a substantial symmetry between electricity and magnetism. This is an old idea, going back to Michael Faraday and first made more precise with the formulation by James Clerk Maxwell of his famous equations regulating the behaviour of electric and magnetic fields.

In current notation (using a unit system for which $c = 1$), *Maxwell's equations* read:

$$\vec{\nabla} \cdot \vec{E} = \rho_e, \tag{1}$$
$$\vec{\nabla} \cdot \vec{B} = 0,$$
$$\vec{\nabla} \wedge \vec{E} = -\frac{\partial \vec{B}}{\partial t},$$
$$\vec{\nabla} \wedge \vec{B} = \vec{j}_e + \frac{\partial \vec{E}}{\partial t}.$$

where $\vec{E}$ is the electric field, $\vec{B}$ the magnetic field, ρ_e the density of electric charge and $\vec{J}_e$ the density of electric current.

There is an evident similarity in the role of electric and magnetic fields in these equations, apart from the presence of the electric source terms. In the absence of such terms – that is, in the case of *free Maxwell's equations* – the similarity becomes complete.

2.1.1 EM duality in the absence of sources

In fact, when there are no charges and current ($\rho_e = \vec{J}_e = 0$), the equations read:

$$\vec{\nabla} \cdot \vec{E} = 0, \tag{2}$$
$$\vec{\nabla} \cdot \vec{B} = 0,$$
$$\vec{\nabla} \wedge \vec{E} = -\frac{\partial \vec{B}}{\partial t},$$
$$\vec{\nabla} \wedge \vec{B} = \frac{\partial \vec{E}}{\partial t}.$$

As is immediately apparent, the free Maxwell's equations are invariant under the following *duality transformation*:

$$D: \quad \vec{E} \to \vec{B}, \quad \vec{B} \to -\vec{E}. \tag{3}$$

In the sourceless case EM duality, expressed by the invariance of the equations under the duality transformation D, is thus an *exact symmetry*.

Notice that the duality transformation D can be generalized to *duality rotations* parameterized by an arbitrary angle θ as follows:

$$\vec{E} \to \cos\theta \vec{E} + \sin\theta \vec{B}, \tag{4}$$
$$\vec{B} \to -\sin\theta \vec{E} + \cos\theta \vec{B}.$$

EM duality can then be expressed as the invariance of the sourceless Maxwell's equations under 'rotations' of the electric and magnetic fields. This can be better visualized by introducing the *complex vector field* $\vec{E} + i\vec{B}$, in terms of which Maxwell's equations can be written in the following concise form:

$$\vec{\nabla} \cdot (\vec{E} + i\vec{B}) = 0, \tag{5}$$
$$\vec{\nabla} \wedge (\vec{E} + i\vec{B}) = i\frac{\partial}{\partial t}(\vec{E} + i\vec{B}).$$

Maxwell's equations in the above form remain invariant under the *duality rotations*:

$$\vec{E} + i\vec{B} \rightarrow e^{i\theta}(\vec{E} + i\vec{B}). \tag{6}$$

In these terms, it is easy to see that the energy and momentum densities of the electromagnetic field, represented respectively by the following two expressions,

$$\mathcal{E} = \frac{1}{2}|\vec{E} + i\vec{B}|^2 = \frac{1}{2}(E^2 + B^2), \tag{7}$$
$$\mathcal{P} = \frac{1}{2i}(\vec{E} + i\vec{B})^* \wedge (\vec{E} + i\vec{B}) = \vec{E} \wedge \vec{B},$$

are invariant with respect to the EM duality transformations.

To sum up:

- When no source terms are present, the duality D exchanges the roles of the electric and magnetic fields while leaving the 'physics' –that is, the Maxwell's equations and physical relevant quantities such as the energy and momentum densities of the electromagnetic field – invariant.

- When electric source terms are present, the Maxwell equations are no longer invariant under the duality D and EM symmetry is broken.

2.1.2 Restoring EM duality in the presence of sources

There is a way to restore the symmetry between the electric and magnetic fields in the presence of sources: that is, by including *magnetic source terms*. Assuming the existence of a magnetic density of charge ρ_g and magnetic

current $\vec{j}_g$, in addition to the usual electric charge density ρ_e and electric current $\vec{j}_e$, the Maxwell's equations take the form

$$
\begin{aligned}
\vec{\nabla} \cdot \vec{E} &= \rho_e, \\
\vec{\nabla} \cdot \vec{B} &= \rho_g, \\
-\vec{\nabla} \wedge \vec{E} &= \vec{j}_g + \frac{\partial \vec{B}}{\partial t}, \\
\vec{\nabla} \wedge \vec{B} &= \vec{j}_e + \frac{\partial \vec{E}}{\partial t}.
\end{aligned}
\tag{8}
$$

These equations are invariant under the following duality transformation, interchanging the roles of the electric and magnetic fields and – at the same time – the roles of the electric and magnetic charges and currents:

$$
\begin{aligned}
\vec{E} &\to \vec{B}, & \vec{B} &\to -\vec{E}, \\
\rho_e, \vec{j}_e &\to \rho_g, \vec{j}_g, & \rho_g, \vec{j}_g &\to -\rho_e, -\vec{j}_e.
\end{aligned}
\tag{9}
$$

In terms of the complex vector field $\vec{E} + i\vec{B}$, the above equations can be written concisely as:

$$
\begin{aligned}
\vec{\nabla} \cdot (\vec{E} + i\vec{B}) &= \rho_e + i\,\rho_g, \\
\vec{\nabla} \wedge (\vec{E} + i\vec{B}) &= i\left[\left(\vec{j}_e + i\,\vec{j}_g\right) + \frac{\partial}{\partial t}\left(\vec{E} + i\,\vec{B}\right) \right].
\end{aligned}
\tag{10}
$$

These equations are invariant under the *duality rotations*:

$$
\begin{aligned}
\vec{E} + i\,\vec{B} &\to e^{i\theta}\left(\vec{E} + i\,\vec{B}\right), \\
\rho_e + i\,\rho_g &\to e^{i\theta}\left(\rho_e + i\,\rho_g\right), \\
\vec{j}_e + i\,\vec{j}_g &\to e^{i\theta}\left(\vec{j}_e + i\,\vec{j}_g\right).
\end{aligned}
\tag{11}
$$

Maxwell's equations can thus be modified to accomodate the inclusion of magnetic charges and currents. The problem is that isolated magnetic charges, the so-called *magnetic monopoles* (or, in Dirac's terminology, *magnetic poles*), have never been observed. If we break a magnet bar in two parts, we always obtain two smaller magnets and never an isolated North pole and an isolated South pole. Quoting Dirac (1948, p. 817): "The field equations of electrodynamics are symmetrical between electric and magnetic forces. The symmetry

5

between electricity and magnetism is, however, disturbed by the fact that a single electric charge may occur on a particle, while a single magnetic pole has not been observed to occur on a particle."

If, in order to save the EM symmetry, we nevertheless assume the existence of isolated magnetic poles, the question is: why are isolated magnetic poles not observed? As we shall see in the next Section, Dirac investigated the problem in the context of quantum electrodynamics, arriving at the following answer: because an enormous energy is needed to produce a particle with a single magnetic pole.

2.2   EM duality (2): quantum electrodynamics

Dirac's solution to the problem posed by EM symmetry is contained in his theory of magnetic poles. The theory was first proposed in his seminal 1931 paper *Quantised Singularities in the Electromagnetic Field* (Dirac, 1931). In his second paper on the subject, appeared in 1948 with the title *The Theory of Magnetic Poles* (Dirac, 1948), Dirac completed the theory by providing "all the equations of motion for magnetic poles and charged particles interacting with each other through the medium of the electromagnetic field in accordance to quantum mechanics" (Dirac, 1948, p. 817-18).

In his 1931 paper Dirac put forward the idea of magnetic pole as "quantised singularities of the EM field", working out the consequences of this idea in the formalism of quantum mechanics. Declared object of his paper was "to show that quantum mechanics does not really preclude the existence of isolated magnetic poles" (Dirac, 1931, p. 71). Why did quantum mechanics present a specific problem for the existence of isolated magnetic poles? The issue at stake was the following: turning from the classical to the quantum formulation of electromagnetic theory with magnetic sources posed a *consistency problem*. On the one hand, the electromagnetic vector potential $\vec{A}$ plays a central role in coupling electromagnetism to quantum mechanics.[2]

---

[2]In the canonical quantization procedure followed by Dirac, the electromagnetic potentials are required for putting the equations of motions into the form of an action principle. In general, the standard way of describing the electromagnetic couplings of the matter wave functions is in terms of the so-called minimal coupling prescription (requiring to replace the momentum operator $\vec{p} = -i\vec{\nabla}$ by its 'covariant' generalization $-i(\vec{\nabla} - ie \cdot \vec{A})$, where e is the electric charge). In other words, the vector potentials $\vec{A}$ explicitly enter the covariant derivative of the wave function of the electrically charged particle and therefore are needed to determine its evolution.

On the other hand, the vector potential $\vec{A}$ is introduced in standard electromagnetism by taking advantage of the absence of magnetic source terms:

$$\vec{\nabla} \cdot \vec{B} = 0 \rightarrow \vec{B} = \vec{\nabla} \wedge \vec{A} \tag{12}$$

(for all $\vec{A}$, $\vec{\nabla} \cdot (\vec{\nabla} \wedge \vec{A}) = 0$).

This seems to imply that quantum mechanics is inconsistent with the presence of magnetic charge. Dirac had thus to address the following consistency issue: whether it was possible to include particles carrying a magnetic charge without disturbing the consistency of the coupling of electromagnetism to quantum mechanics.

The argument he proposed in his 1931 paper for solving this apparent inconsistency is remarkable under many aspects. In particular, it represents one of the first example of an explicit use of *topological considerations* in the early twentieth century physics. In developing his argument, centered on the relation between the phase change of the wave functions round closed curves and the flux of the magnetic field $\vec{B}$ through closed surfaces, Dirac in fact applied ideas involving the structure of the space in the large (what is now known as global topology).[3] The result he obtained was the following: the introduction of magnetic charge can be consistent with the quantum theory provided its values are 'quantized'. In his own words (Dirac 1931, p. 68): "Our theory thus allows isolated magnetic poles, but the strength of such poles must be quantised, the quantum μ_0 being connected with the electronic charge e by $\hbar c / e \mu_0 = 2$."

In current notation (denoting magnetic charge by g and using the unit system $\hbar = c = 1$), Dirac's result was that a magnetic charge g can occur in the presence of an electric charge e if the following condition, known as *Dirac quantization condition*, is satisfied:

$$e\, g = 2\pi n \qquad n = 0, \pm 1, \pm 2, \ldots . \tag{13}$$

This condition has an immediate striking consequence: the mere existence of a magnetic charge g somewhere in the universe implies the *quantization of electric charge*, since any electric charge must then occur in integer multiples of the unit $2\pi/g$. In Dirac's words (*ibid.*), "The theory also requires a quantisation of electric charge, since any charged particle moving in the field of a

---

[3]On Dirac's anticipation of topological ideas in physics see, for example, Olive (2003).

pole of strength μ_0 must have for its charge some integral multiple (positive or negative) of e, in order that wave functions describing the motion may exist."

The quantization of electric charge was a fact of observation, but theoretically unexplained. For Dirac, it was indeed the possibility of obtaining an explanation of this fact to constitute one of the main reason of interest in his theory of magnetic poles. As he wrote in his 1948 paper (Dirac 1948, p. 817), "The interest of the theory of magnetic poles is that it forms a natural generalization of the usual electrodynamics and *it leads to the quantization of electricity.* [...] The quantization of electricity is one of the most fundamental and striking features of atomic physics, and there seems to be no explanation for it apart from the theory of poles. This provides some grounds for believing in the existence of these poles."

In substance, according to Dirac, even if magnetic charges are not observed the theory provides a good reason for believing in their existence. In fact, the theory also provides an explanation of why isolated magnetic poles are not observed. The explanation is based on the great difference between the numerical values for the quantum of electric charge e_0 and the quantum of magnetic pole g_0. In the notation used by Dirac in his 1948 paper, if we take the experimental value for the *fine structure constant*, i.e. $\alpha = e_0^2/(\hbar c) = 1/137$, and we use the quantization condition (in its original form: $e_0 g_0 = (1/2)\hbar c$), we can infer that the value of the quantum of magnetic pole is $g_0^2 = (137/4)\,\hbar c$, that is much greater than the numerical value for the quantum of electric charge, $e_0^2 = (1/137)\,\hbar c$.[4]

Thus, Dirac notes, "although there is symmetry between charges and poles from the point of view of general theory, there is a difference in practice" (Dirac 1948, p. 830). For example, two one-quantum poles of opposite sign attract one another with a force $(137/2)^2$ times as great as that between two one-quantum charges at the same distance. "It must therefore be very difficult to separate poles of opposite sign", Dirac continues, and his conclusion is that "this explains why electric charges are easily produced and not magnetic poles" (*ibid.*).

---

[4]With respect to the quantization condition (13), formulated according to the notation using the unit system $\hbar = c = 1$ and the Heaviside-Lorentz conventions for electromagnetism, the quantization condition in the form originally given by Dirac uses definitions of the electric charge and the magnetic charge differing by a factor of 4π.

3 The meaning of EM duality

In *classical electrodynamics* (with magnetic source terms included), we have seen that the EM duality transformation

$$\begin{aligned}
\vec{E} &\rightarrow \vec{B}, & \vec{B} &\rightarrow -\vec{E}, \\
\rho_e, \vec{j}_e &\rightarrow \rho_g, \vec{j}_g, & \rho_g, \vec{j}_g &\rightarrow -\rho_e, -\vec{j}_e,
\end{aligned} \tag{14}$$

exchanges, at the same time, the roles of the electric and magnetic fields and the roles of the electric and magnetic charges and currents, while leaving the physics invariant. 'The physics' means the Maxwell's equations and the relevant physical quantities (such as the energy and momentum densities of the electromagnetic field). EM duality is thus a symmetry of the theory, expressing the equivalence of the following *dual ways* of describing the same physics:

1) *Description$_1$*. The physics is described in terms of:

- the electric field $\vec{E}_1$ and the magnetic field $\vec{B}_1$;

- the electric charge and current densities ρ_{e_1} and $\vec{j}_{e_1}$, and the magnetic charge and current densities ρ_{g_1} and $\vec{j}_{g_1}$.

2) *Description$_2$*. The physics is described in terms of:

- the electric field $\vec{E}_2 = \vec{B}_1$ and the magnetic field $\vec{B}_2 = -\vec{E}_1$;

- the electric charge and current densities $\rho_{e_2} = \rho_{g_1}$ and $j_{e_2} = j_{g_1}$, and the magnetic charge and current densities $\rho_{g_2} = -\rho_{e_1}$ and $\vec{j}_{g_2} = -\vec{j}_{e_1}$.

This means, in concrete, that a calculation of a physical quantity in the framework of *description$_1$* can be obtained by means of another calculation in the dual framework of *description$_2$*. For example, calculating the force of the electric field $\vec{E}_1$ on a particle with electric charge e_1 in the framework of *description$_1$* is the same as calculating the force of the magnetic field $\vec{B}_2$ on a particle with magnetic charge $g_2 = -e_1$ in the framework of *description$_2$*.

For the duality issue of concern here, this does not say much. The idea of a symmetry between electricity and magnetism is, of course, more profound then what the above consideration can show. In particular, it has played a very important heuristic role in the history of pre-quantum electrodynamics –

think about its influence on Faraday's discovery of electromagnetic induction or Einstein's 1905 work on special relativity. But it is only in the quantum context that the full theoretical significance of physical dualities does actually emerge.

In order to have a complete grasp on the real meaning of EM duality in quantum physics, we should follow the development of this idea in quantum field theory and string theory. In this paper we pursue a much more modest scope. We remain in the conceptual range of the preceding Section, and consider what can be extracted from Dirac's seminal work for the issue at stake. In fact Dirac anticipated so much that, on the basis of his results, it is possible to get an idea of some general features of today's physical dualities. Here we focus on the most striking of these features: that is, the fact that dualities typically interrelate weak and strong coupling. This is known, in the physics literature, as *weak-strong duality*.

In the framework of Dirac's theory of magnetic monopoles, it is easy to see how the weak-strong interchange naturally follows from assuming EM duality and the quantization condition. As we have seen, EM duality implies interchanging electric and magnetic charges:

- EM duality: $\qquad\qquad e \to g, \qquad\quad g \to -e,$

while Dirac's quantization condition implies that the electric and magnetic charges (that is, the electric and magnetic coupling constants) are so related:

- Quantization condition: : $\quad e\,g = 2\pi n.$

Putting the two together, we obtain:

$$e \to g = \tfrac{2\pi n}{e}, \quad g \to -e = -\tfrac{2\pi n}{g}.$$

This means that if the charge e is small, the charge g into which it is transformed is strong and *vice versa*. That is: in quantum physics, EM duality relates weak and strong coupling.

In general, turning to the more appropriate context of quantum field theory and string theory, what happens is that dualities typically relate a theoretical description concerning a strong-coupling regime to another description concerning a weak-coupling regime (while leaving the 'physics' invariant). That is, dualities exchange physical regimes that are very different, with the remarkable consequence that calculations involving strong forces in one theoretical description can be obtained from calculations involving weak

forces in the dual theoretical description.[5] This is not all: at the same time, dualities also typically exchange elementary quanta ('electric charges') with collective excitations ('solitons' or 'magnetic charges'), with the consequence that what was viewed as *fundamental* in one theoretical description becomes *composite* in the dual description.[6]

4 Concluding remarks: dualities and physical theories

From a philosophical point of view, the above illustrated features are rather unusual, especially if dualities are to be considered as intertheoretic relations. Physical theories are generally intended to describe a given range of phenomena: they have specific domains of application, defined in correspondence to some range or level of the adopted physical scale (for example, the energy scale). In the cases of intertheoretic relations usually discussed in the philosophy of science – in connection, for example, with such issues as reductionism and continuity across theory change – , the theories considered are either on the same level or on successive levels. In this latter case, the two theories are typically so related that one can be seen as 'emerging' from the other. But dualities show that another type of situation is possible: the two interrelated theoretical descriptions can be on very different scale levels. Moreover, by means of dualities the 'same physics' is described by two theoretical formulations presenting apparently different ontologies: the fundamental objects in one formulation become composite objects in the dual formulation, and viceversa.

A first question is then: what do dualities indeed relate? Two different theories or just two different formulations of the same theory? The answer surely depends on the sort of duality we are considering. But also on what we intend by a 'theory', and this is also closely connected to the question of what

---

[5]This is what makes dualities particularly interesting and useful in the context of quantum field theory and string theory, as we usually know only the perturbative part of a theory, that is its 'weak coupling' regime. Dualities thus relate what is still unknown to what can be calculated.

[6]To be honest, this cannot be seen in the context of Dirac's theory of magnetic poles. It is important to underline that this feature could emerge only with the extension of dualities in the framework of quantum field theory. An excellent review of these developments of electromagnetic duality is Olive (1995).

sort of symmetry is represented by dualities, if these are indeed symmetries (as is commonly assumed).

It is usually said that dual theories, or dual theoretical descriptions, are connected with one another by transformations 'leaving the physics invariant': dualities are in this sense 'symmetries'. This can be made more precise by specifying the meaning of the expression 'leaving the physics invariant'. If by this we intend that the dynamical equations of the theory remain invariant, as in the EM duality case discussed in Section 2 (where the Maxwell's equations are invariant under the duality transformation D), then the duality is a symmetry of the theory in the precise sense normally used in contemporary physics. That is, the sense according to which G is a symmetry group of a theory if the dynamical equations (or the 'action') of the theory are invariant under the transformations (that are the elements) of the group G. The symmetries postulated through the so-called invariance principles of physics, such as the space-time symmetries and the gauge symmetries of the Standard Model of particle physics, are properties of physical theories in this sense.

But, in general, the dualities used in today physics relate two different theoretical descriptions that concern different scale levels and present apparently different ontological scenarios. These descriptions can even involve different actions (or Hamiltonians) and different fields.[7] In which sense, then, they are just two different formulations of the same underlying theory, as is commonly maintained? This clearly depends on the meaning attributed to the notion of *theory*. The clue is given by the extended sense in which duality is considered a symmetry. That is: the 'theory' is identified on the basis of what remains invariant under the duality transformations, the 'same physics' that is differently described by means of the dual formulations. And this 'same physics', according to the physicists working on the subject, is given by the spectra and the transition amplitudes.[8]

We thus arrive at an apparently 'phenomenological' understanding of the notion of a theory that may seem paradoxical in such a highly mathematized and far away from common (and, for now, possible) experience as is string theory. Note that such a notion is not new in the history of quantum physics: think about the ideology behind Heisenberg's matrix mechanics in the 1920s or the S-matrix approach dominating in the 1960s (which was, it is worth

---

[7]A companion paper in preparation is devoted to examining in some detail this point, by focussing on the developments of electromagnetic duality in quantum field theory and string theory

[8]See, for example, Polchinski (1998), Section 4.

noting, at the basis of the so-called 'dual resonance model' from which early string theory was born in the late 60s).[9]

Summing up, physical dualities pose a dilemma to the philosophers of science: either the physicists's 'received view' that dualities relate different formulations of the same theory is accepted, but this implies a notion of what is a 'physical theory' which is quite different from the common idea that a theory is identified on the basis of its fundamental dynamical equations and ontology; or, on the contrary, dualities are understood as relations between different physical theories, but then it is difficult to understand the real meaning of such inter-theory relations and to see in which sense they can be considered 'symmetries'.

5   References

A. Cappelli, E. Castellani, F. Colomo, and P. Di Vecchia (eds.) (2010), *The Birth of String Theory*, Cambridge, Cambridge University Press.

R. Dawid (2007), "Scientific Realism in the Age of String Theory", *Physics and Philosophy* **11**.

P. A. M. Dirac (1931), "Quantised Singularities in the Electromagnetic Field", *Proc. Roy. Soc. Lond. A* **133**: 60-72.

P. A. M. Dirac (1948), "The Theory of Magnetic Poles", *Phys. Rev.* **74**: 817-830.

D. Olive (1995), "Exact electromagnetic duality", arXiv: hep-th/9508089.

D. Olive (2003), "Paul Dirac and the pervasiveness of his thinking", arXiv: hep-th/03041133.

---

[9]See, on this point, the introduction and part I of Cappelli, Castellani, Colomo and Di Vecchia (2010).

J. Polchinski (1998), "Quantum gravity at the Planck length", arXiv: hep-th/9812104.