Which Curie's Principle?

Elena Castellani[*]and Jenann Ismael[†]

June 2015

Since its first formulation in a famous 1894 article by Pierre Curie, the principle stating that "the symmetries of the causes are to be found in the effects" has been defended or questioned on diffferent grounds. In recent decades, it has become the object of renewed philosophical discussion in connection with the growing interest in the role of symmetry and symmetry breaking in physics (Ismael, 1997; Belot, 2003; Earman, 2004; Roberts 2013). In this literature, it has become current to understand (and question) the principle as following from the invariance properties of deterministic physical laws. The seminal paper for this "received view" is Chalmers (1970), introducing a formulation of Curie's principle in terms of the relationship between the symmetries of earlier and later states of a system and the dynamical law connecting these states. This re-formulation places the emphasis in a different place with respect to Curie. Curie's focus was clearly on the case of co-existing, functionally related features of a system's state, rather than temporally ordered cause and effect pairs. While Chalmers (1970) and Ismael (1997) still emphasize the generality of the principle by including in their formulations physical situations of the type considered by Curie, this is no more the case in what has become the received view.

Is there more that one "Curie's principle", then? How far are different formulations legitimate? Given the important and widely acknowledged methodological role of the principle in science, are there features to be highlighted and used for a modern formulation? What are the aspects that make it so scientifically fruitful, independently of how it is formulated?

---

[*]Department of Letters and Philosophy, University of Florence, via Bolognese 52, 50139, Firenze, Italy.

[†]Department Philosophy, University of Arizona and Center for Advanced Studies in the Behavioral Sciences, Stanford University, 75 Alta Rd., 94305.

The paper is devoted to exploring these questions. We start with illustrating Curie's original article and his focus. Then, we consider the way that the discussion of the principle took shape from early commentators to its modern form. We say why we think that the modern focus on the inter-state version of the principle loses sight of some of the most interesting significant applications of the principle. Finally, we address criticism of the principle put forward by Norton (2014) and purported counterexamples due to Roberts (2013, 2014).

1 Curie's "considérations": the origin of the principle

Curie devoted a series of works to examining the role of symmetry and asymmetry in physical phenomena. His analysis was centered on the question: which phenomena are allowed to occur in a given physical medium having specified symmetry properties? His studies of such properties as the pyro- and piezo-electricity of crystals persuaded him of the importance of the relationships between the symmetry of a physical medium and the symmetry of the phenomena occurring in it.

By applying methods and results of the crystallographic theory of symmetry groups to the study of a number of physical phenomena in his seminal 1894 paper, he arrived at the following conclusions:

(a_1) When certain causes produce certain effects, the symmetry elements of the causes must be found in their effects.

(a_2) When certain effects show a certain dissymmetry, this dissymmetry must be found in the causes which gave rise to them.[1]

(a_3) In practice, the converses of these two propositions are not true, i.e., the effects can be more symmetric than their causes.

(b) A phenomenon may exist in a medium having the same characteristic symmetry or the symmetry of a subgroup of its characteristic symmetry.[2] In other words, certain elements of symmetry can coexist with certain phenomena, but they are

[1] Curie uses the term "dissymetry" in his paper, where we would today use the term "asymmetry".

[2] Curie defines "the characteristic symmetry of a phenomenon" as "the maximum symmetry compatible with the existence of the phenomenon". For example, the characteristic symmetry of a force, a velocity, or the intensity of an electric field is that of an arrow

> not necessary. What is necessary, is that certain elements of symmetry do not exist. Dissymmetry is what creates the phenomenon.

Conclusion (a_1) is what has become known as *Curie's principle* (CP). Conclusion (a_2) is logically equivalent to (a_1); the claim is that symmetries of the cause are necessarily found in the effect, while dissymmetries of the cause need not be. Conclusion (a_3) clarifies this claim, emphasizing that since dissymmetries of the cause need not be found in the effect, the effect may be more symmetric than the cause.

Conclusion (b) invokes a distinction, found in all of Curie's examples, between medium and phenomena. CP states that the symmetry of the medium cannot be higher than the symmetry of the phenomenon.[3] If the medium in which a phenomenon occurs starts out in a highly symmetric state, CP entails that the original symmetry group of the medium must be lowered to the symmetry group of the phenomenon (or to a subgroup thereof). In this sense symmetry breaking is what "creates the phenomenon".

1.1 Applications

In order to illustrate how his principle applies, Curie discusses a number of cases where symmetry considerations impose necessary conditions for the occurrence of some phenomenon. He typically considers situations where "several phenomena of different nature superpose in a single system" and asserts that in those cases, the symmetry group of the system includes only those symmetries common to all of the superposed phenomena (p. 127).

His first illustration is the phenomenon known as the *Wiedemann effect*. This phenomenon, discovered by Gustav Wiedemann in 1858, can be represented in three different ways, depending on the order in which an electric field, a magnetic field and a torque are combined in an iron cylindrical wire:

(a) The generation of an asymmetrical torque in an iron cylindrical wire when a longitudinal electric field and a longitudinal magnetization are applied.

(truncated cone), while the characteristic symmetry of the intensity of a magnetic field is that of a cylinder rotating about its axis.

[3] Thus, for a magnetic field (the effect) to exist, the medium (the cause) must have a symmetry lower or equal to that of a rotating cylinder.

(b) The generation of a longitudinal electric field when a longitudinal magnetization of an iron cylindrical wire and an asymmetrical torque are applied.

(c) The longitudinal magnetization of an iron cylindrical wire when a longitudinal electric field and an asymmetrical torque are applied.

What is being considered cause and effect for the purposes of applying CP depends on the description chosen.

Let us choose one description, say (c). In this case, the longitudinal magnetization is the effect, and all the rest – the cylindrical wire, the longitudinal electric field, the asymmetrical torque – the cause. The original symmetry group of the wire, before the application of the electric field and the torque, is the symmetry group of a cylinder at rest. For the magnetization of the wire to occur, the symmetry group must be lowered to that of a rotating cylinder. This is what happens under the combined action of the electric field and the torque (dissymmetrization of the cause).

1.2 Curie's focus

In the last section of his paper, reflecting on the heuristic power of his symmetry considerations, Curie emphasizes two kinds of conclusions.

The first are "firm but negative" conclusions:

> *There is no effect without causes.* Effects are the phenomena which always necessarily require a certain dissymmetry in order to arise. If this dissymmetry does not exist, the phenomena are impossible.

The utility of CP, in this case, is to save us the trouble of searching for phenomena that can't occur.

The second are "positive but lacking the same degree of certainty and precision":

> *There is no cause without effects.* Effects are the phenomena which can arise in a medium possessing a certain dissymmetry. One has here precious directions for the discovery of new phenomena [...] We have no idea of the order of magnitude of the predicted phenomena; we have only an imperfect idea of their precise nature.

The utility of CP, here, is the guidance it provides about where to search for new phenomena, although we are not assured of finding anything.

The original CP has thus an important methodological function. On the one hand, it furnishes a selection rule (given an initial situation with a specified symmetry, only certain phenomena are allowed to occur); on the other hand, it offers a falsification criterion for physical theories (a violation of CP may indicate that something is wrong in the physical description). Of course, in order for the principle to be applicable, various conditions need to be satisfied: the cause and effect must be well-defined, they must be related by a deterministic equation, and the symmetries of both the cause and the effect must be well-defined. This last condition applies to both the physical and the geometrical properties of the system under consideration. *Modulo* these conditions, CP provides necessary conditions for given phenomena to occur: only those phenomena can occur that are compatible with the symmetry conditions stated by the principle.

Curie was relatively unconcerned with providing a justification for his conclusions. At the beginning of the paper, he states that the nature of symmetry considerations is analogous to that of dimensional considerations. By this he meant that CP is to be intended as a heuristic tool, demanding no more justification than its successful application in guiding our search for new phenomena and theory construction.

2 Re-formulations of CP

Since Curie's paper, the significance of CP has been questioned at various times. In the last three decades especially, the focus has shifted from the heuristic reading of the principle to questions about the status and validity of CP. A point of special concern, for example, has been the relation of the principle to the phenomenon of "spontaneous symmetry breaking".[4]

The central question has become whether the validity of CP can be demonstrated. The common strategy has been a) to start with making Curie's original statement more precise, by offering a definition of cause and effect, and, then, b) to appeal to other principles (determinism, invariance) from which CP could follow analytically. The seminal work, in this direction, is Chalmers (1970): the paper is entirely devoted to examining the status and uses of CP, offering a proof for it where the principle is derived from the

[4]This issue has been dealt with in Ismael, 1997, and Castellani, 2003.

invariance properties of deterministic physical laws. This view has become the dominant one in the philosophical discussion on CP (see Earman, 2004; Roberts, 2013).

2.1 From Chalmers (1970) on

In his 1970 paper Chalmers' strategy was, first, "to present as precise and as general a formulation of Curie's principle as possible" (*reformulation* of CP); then, on the basis of this reformulation, to show that "CP follows, for deterministic laws, from their invariance properties" (*proof* of CP by reduction to other principles).

Chalmers' main novelty is the extension of the principle from the synchronic cases considered by Curie, where the cause and effect are simultaneous, to diachronic cases, where the cause precedes the effect and they are related by a deterministic dynamical law. Thus, for example, "the positions and velocities of a system of Newtonian particles can be considered the cause of their positions and velocities at some later time" (p. 137).

With this extension, Chalmers proceeds to prove the validity of CP. Schematically, his argument goes as follows.

Claim: *CP follows from the invariance properties of physical laws if these are deterministic.*

Premises (P1-P4)

P1: *Definition of cause (effect)*: $C(t_1)$ is the cause of the effect $E(t_2)$ ($t_2 \geq t_1$), if $C(t_1)$ is sufficient to ensure $E(t_2)$.

P2: *Determinism*: $E = f(C)$, where f is a function equivalent to a set of ordered pairs (C, E), known if the laws of nature are known.

P3: *T-symmetry of the laws*: invariance of the laws under the transformation T, that is $T[f(C)] = f[T(C)]$ (the symmetry transformation commutes with the function f, that is the pair (C, E) is transformed into the pair (TC, TE)).

P4: *T-symmetry of the cause*: invariance of C under the transformation T, that is $T(C) = C$.

Conclusion (C)

C: *T-symmetry of the effect*: invariance of E under the transformation T, that is $T(E) = E$.

Chalmers' proof that **C** (*the symmetry of the cause is to be found in the effect*) obtains from **P1**, **P2**, **P3** and **P4** as follows:

From **P2**: $E = f(C)$

Apply T: $T(E) = T[f(C)]$

For **P3**: $T(E) = f[T(C)]$

For **P4**: $T(E) = f(C)$

For **P2** again: $T(E) = E$

Summing up in a diagram:

$$\begin{tikzcd} C \arrow[r, "f"] \arrow[d, "T"'] & E \arrow[d, "T"] \\ T(C) \arrow[r, "f"'] & T(E) \end{tikzcd}$$

Chalmers' proof is taken for granted in recent literature, albeit reformulated in slightly different terms and from a completely diachronic point of view (Belot, 2003; Earman, 2004; Roberts, 2013). Accordingly, the main issue has become whether symmetry properties of states are preserved through dynamical evolution (which was not at all Curie's original concern). In terms of this "received view", the natural reading of CP is that *a system cannot evolve from a symmetric to an asymmetric state*. In deterministic cases, this version of CP (let's call it CP_{RV}) follows straightforwardly from the symmetry of the dynamical law.[5]

[5] Proof proceeds in the same way as Chalmers' one, just reformulated in terms of the symmetry properties of earlier and later states of a system.

CP$_{RV}$ differs from the original principle in its exclusive focus on dynamics. We have been concerned here to exhibit the synchronic applications of the principle that Curie himself emphasized and restore the general reading of the principle given, for example in Ismael (1997). In what follows, we respond to some recent discussions that we believe treat Curie unfairly.

3 Defending Curie

3.1 On Norton's criticism

Norton (2014) portrays Curies Principle as a contribution to a long history of causal metaphysics. He argues that CP fails to provide a "general, factual causal principle to which all science must conform" or a "factual principle of causality that usefully restricts our science". All of this seems to us orthogonal to Curie's intent. The causal vocabulary was perhaps unfortunate, but Curie doesn't seem to have meant anything more by 'causal relation' than a functional relationship. He writes: "Whenever a physical phenomenon is expressed as an equation, there is a causal relation between the quantities appearing in both terms".

Early commentators on the paper explicitly adopted the deflationary reading of the causal language. For example, Ismael (1997, p. 169) writes: "What the principle says depends crucially on how the terms 'cause' and 'effect' are understood, ... Let A and B be families $\{A_1, A_2, ...\}$ and $\{B_1, B_2, ...\}$ respectively, of mutually exclusive and jointly exhaustive event types, and let the statement that A is a Curie-cause and B its Curie-effect mean that the physical laws provide a many-one mapping of A into B". So construed, the principle is quite general. There is no restriction on the terms on what the terms of the relation can be. As we have seen, they can be different aspects of the total state of one system at a single time, or total states of a system at two different times. Someone who wants to apply the principle simply has to find a functional relationship.

Norton seems doubtful that the principle in this form – i.e., as he says, "a simple tautology" – can play any useful role in science. The only response to this sort of doubt is provide examples. In the highly mathematical setting of modern physics, CP is remarkable for being a purely qualitative principle that gives one a very abstract sense of how the states that fall in one class relate to states that fall in another class, when there is a deterministic relation between them. The most straightforward cases are those in which we have a pair of states of a physical system related by a dynamical law. Molecular biology is replete with particularly impressive instances of apparently T-symmetric conditions giving rise, by evidently deterministic processes, to apparently T-asymmetric effects. Frog zygotes, for example, start out as spherical cells suspended in a homogenous seeming fluid and develop into highly structured organisms; almost every stage in their development introduces asymmetries not apparently present in the preceding stage. So long as the evolution is deterministic we can conclude that either (i) the symmetric facade of the initial state concealed all of the asymmetries revealed in the final state, (ii) dynamically relevant T-asymmetric factors were introduced by the environment, or (iii) T is not a symmetry of the law of evolution. CP just focuses ones physical understanding by setting the task of finding out which it is.

The frog zygote example is very easy to visualize because cause and effect are physical states and the law equation is a dynamical equation, but CP also applies to cases that aren't so easy to visualize. Here is one, for example, that might not seem obvious. Consider all of the confusion about the separation of ontic and epistemic states in quantum mechanics. The psi-empiricists think that a lot of what look like ontic effects in quantum mechanics are disguised epistemic affects. They derive various kinds of oddities in the evolution of the quantum state (from collapse to teleportation to remote steering) by interpreting the quantum state as a representation of partial knowledge about an underlying ontic state. CP can act in that setting as a desideratum on how much you have to put into the ontic state. It can tell us, without knowing anything more specific, that if we have an ontic asymmetry in the final state after any quantum process, there has to be an ontic asymmetry in the initial state or the law of evolution.

CP also suggests a purely qualitative way of understanding the physical significance of automorphisms of the solution set. Instead of thinking of them as automorphisms of the set of solutions, we can think of them as transformations among the values of dynamically relevant parameters that preserve the relations described by the laws. The symmetries of a set of equations determining one among a family B of alternatives, then, correspond physically to either

1. Permutations of the values of B-irrelevant parameters, or

2. Irrelevant permutations of B-relevant parameters, i.e. transformations

> which either map them onto themselves or are accompanied by compensating transformations in the values of other parameters in such a way as to preserve the relation described by the law.

The contrapositive of this is that transformations which *aren't* symmetries correspond physically to relevant permutations of the values of relevant parameters. That is a very simple idea that captures the physical content of the symmetries of a set of laws, and provides a kind of physical insight into what the symmetries of a set of laws are, in a way that you can explain to someone who has never taken a math class in his life.

It can also be useful for the physicist and philosopher of physics, because it can be so easy for physical insight to get compromised by mathematics. The interpretation of transformations that aren't symmetries of the laws is easy to see in the case of non-geometric transformations, when the transformations permute the values of parameters in the equation expressing the laws. In the case of geometric symmetries, it can play a particularly useful role. If dynamical theories are formulated in their traditional coordinate-dependent manner and geometric transformations are represented as transformations between coordinate-systems, T may be an asymmetry of the laws determining B, even though no T-asymmetric parameter appears in the B-determining equations. This, combined with the historical confusion about the precise nature of the coordinate-dependence, obscured the physical significance of geometric transformations for generations. Its only when we insist on a generally covariant formulation that we have to put something in the equations that itself is not invariant under T. Consider a universe which exists for exactly a minute and consists of a sphere which gradually deforms into an ellipse. You cant write down a deterministic equation describing the evolution of the sphere in a generally covariant form without including a parameter whose value is not invariant under spatial rotations, because you will need to distinguish the direction along which the sphere elongates.

3.2 On Roberts' criticism

Roberts (2013) describes a simple classical situation that presents a straightforward apparent failure of CP. He generalizes it to a wide class that includes both classical and quantum cases. Roberts (2014) describes related apparent failures of the principle. Since we saw that the principle that we are calling Curie's Principle is a tautology, we know that the examples can't be failures of that. The mathematical analysis of the examples shows what is really going on. All of the examples concern time reversal. The physical interpretation of the formal operation known as time reversal symmetry has always been obscure and it becomes clear that time reversal as it is standardly understood both classically and in quantum mechanics is not a symmetry transformation in a sense that is needed for the truth of CP. That conclusion reveals something of importance about this complex and contested transformation. The fact that the transformation as standardly defined doesn't satisfy CP gives us good reasons for being suspicious of reasoning with it as we do with other symmetry transformations.

A set of dynamical laws is invariant under time reversal, just in case, if A is a solution to the laws, where A is an initial state and B is a final state, then $B \rightarrow A$ is as well. Given a set of laws that have this property, CP says that if the initial state has some symmetry, the final state has it as well. CP will hold, however, only if the states on which time-reversal operates do not themselves contain dynamical information. This is just what one would expect, for if the states contain dynamical information i.e., if the states contain information not just about where the system is at a given moment (in physical space, or state-space) but where it will be the next - time reversing the sequence doesnt even make sense. If we include velocities in state descriptions, and time reverse a sequence of states by simply reversing the order of the states in the sequence, we get nonsense. The standard fix is to also reverse the velocities, and any other quantities that go into the specification of state that also contain dynamical information. This raises the question of which quantities contain dynamical information.

In physical terms, time reversal should leave the states intrinsically untouched and just change their order. If we cleave to that understanding of time reversal, none of the counterexamples Roberts offers constitutes a failure of CP. The central example in Roberts (2013) is that of a classical harmonic oscillator initially compressed out of equilibrium with zero momentum. The final state has positive momentum in the direction away from the way to which it is affixed. Roberts' claim is that the initial state is invariant under time reversal, and the final state is not. Since the laws are invariant under time reversal, this is a failure of CP. We claim, contra Roberts, that the final state is invariant under time reversal if we consider the intrinsic, instantaneous state of the spring.

In his contribution to this symposium, Roberts considers a collection of examples drawn from electromagnetism. In those cases, the problem is rather different. There we say that the theory is not time reversal invariant in the sense relevant to the application of CP. Roberts briefly considers ways of restricting what counts as a "symmetry transformation" that would preserve the truth of CP and rejects it on the grounds that time reversal will no longer count as a symmetry of the laws, remarking that "The price of this response is that one must give up the standard meaning of 'time reversal invariance', in favor of a property that is almost never satisfied" (p. 12). It is not obvious that this isnt the right response. Indeed, it was on grounds very like this that Albert (2000) argued that no theory since Newtonian Mechanics has been genuinely time reversal invariant. The discussion of time reversal invariance is ongoing and heavily contested. There is no broad agreement on the proper definition. The operation is quite singular in various ways.

Is this a fair response to the examples? In what sense is CP a necessary truth if we get to dis-count counterexamples on the ground that they don't use "symmetry transformation" in the way demanded for the truth of CP? We believe that it is a legitimate response. The principle is not a logical truth. Its truth is not independent of the meanings of its terms. The physical intuition at the heart of CP depends on symmetry transformations being understood as operating on the intrinsic states of systems related by a deterministic equation. The fact that it fails for time reversal as standardly understood shows that the formal operation known as time-reversal in the physical literature has drifted rather far from that meaning. There is no point in fighting about whether to hang onto a formulation of CP that uses "the symmetries of the laws" in standard way, or to preserve the principle at the cost of restricting what counts as a symmetry of the laws. We can all agree that in trying to suss out the physical content of symmetry operations that we can define mathematically, it can be very helpful to force oneself back to the physical insight at the heart of CP and to understand what is happening in these apparent failures.

3.3 Summing up

The general physical insight at the heart of CP is: if A and B are the domain and range of a deterministic law, D, wherever D is not symmetric under a transformation T, T must permute some physically significant feature of the situation implicated in the production of B. Another way to put it is that T must take us across boundaries of physical equivalence classes.

Figuring out how to apply that basic insight in the mathematically complex setting of contemporary physics can help to clarify the physical content of mathematical operations like time reversal. Norton (2014) points out that there is a lot of leeway in how we apply CP to a situation, and we can formally put the asymmetry into the domain or into the law. Of course hes right about that. But then he goes on to say: "There is no higher principle that dictates which mapping [of causal vocabulary onto the physics] is correct. What decides the mapping used is familiarity, comfort and, ultimately, our whim" (p. 6).

This last remark is just wrong, both as a descriptive claim and as a normative one. Perhaps sometimes what decides the mapping is familiarity, comfort, or whim. But it doesnt have to be. It is common in physics to have a lot of leeway in how we represent a single physical situation. Different ways of modeling a situation will give us different kinds of insight, displaying different patterns or relationships in a perspicuous way and making others difficult to discern. We do well to keep them all at our disposal. In general, the better we understand all of the different ways of modeling a situation, the better we understand the situation we are modeling. One might say that there are heuristic reasons for aiming for a representation that 'puts the asymmetry into the cause rather than in the dependency relation', because it perspicuously displays, or makes explicit, the physically significant features of the situation implicated in the production of the range B. But one might find different reasons to leave it in the law of dependency. Either way we are going to have to recognize that there is some physically significant difference between T-related situations that manifests itself in the production of B. That is the objective physical insight at the heart of CP.

Acknowledgements— Many thanks to our co-symposiasts John Norton and Bryan Roberts, as well as to Katherine Brading, for stimulating and helpful discussion.

References

Albert, David. 2000. *Time and Chance.* Cambridge (MA): Harvard University Press.

Belot, Gordon. 2003. "Notes on symmetries." In *Symmetries in Physics: Philosophical Reflections*, ed. K. Brading and E. Castellani, 393-412. Cambridge: Cambridge University Press.

Castellani, Elena. 2003. "On the meaning of symmetry breaking." In *Symmetries in Physics: Philosophical Reflections*, ed. K. Brading and E. Castellani, 321-334. Cambridge: Cambridge University Press.

Chalmers, Alan F. 1970. "Curie's Principle." *The British Journal for the Philosophy of Science* 21: 133-148.

Curie, Pierre. 1894. "Sur la symétrie dans les phènomènes physiques. Symétrie d'un champ électrique et d'un champ magnétique." *Journal de Physique* 3: 393-415.

Earman, John. 2002. "Curie's Principle and Spontaneous Symmetry Breaking." *International Studies in the Philosophy of Science* 18: 173-198.

Ismael, Jenann. 1997. "Curie's Principle." *Synthese* 110: 167-190.

Norton, John D. 2014. "Curie's Truism." In: [2014] Philosophy of Science Assoc. 24th Biennial Mtg (Chicago, IL).

Roberts, Bryan W. 2013. "The Simple Failure of Curie's Principle." *Philosophy of Science* 80(4): 579-592.

Roberts, Bryan W. 2014. "Curie's Hazard: From electromagnetism to symmetry violation." In: [2014] Philosophy of Science Assoc. 24th Biennial Mtg (Chicago, IL).