Curie's Principle and Causal Graphs

David Kinney

Forthcoming in Studies in History and Philosophy of Science.[*]

February 22, 2021

Abstract

Curie's Principle says that any symmetry property of a cause must be found in its effect. In this article, I consider Curie's Principle from the point of view of graphical causal models, and demonstrate that, under one definition of a symmetry transformation, the causal modeling framework does not require anything like Curie's Principle to be true. On another definition of a symmetry transformation, the graphical causal modeling formalism does imply a version of Curie's Principle. These results yield a better understanding of the logical landscape with respect to the relationship between Curie's Principle and graphical causal modeling.

1 Introduction

There is currently a lively debate in philosophy of physics over whether or not Curie's Principle says anything meaningful about the role that symmetries play in the causal structure of the world. While competing precisifications of the principle abound, the core idea can be traced back to Pierre Curie's results regarding the piezoelectric properties of crystals (Curie 1894). On the basis of these results, Curie conjectured that nature does not contain spontaneous asymmetries; any symmetry property of a cause must also be a symmetry property of its effect, where the term 'symmetry' has a technical meaning that I will soon spell out in more detail. This prohibition on spontaneous asymmetry has come to be known as Curie's Principle. Ismael (1997), Brading and Castellani (2003), and Castellani and Ismael (2016) have argued that Curie's Principle is

[*]I am especially grateful to Bryan Roberts for asking the question that led me to write this paper, and for his comments on and discussions of several initial drafts. I am also grateful to Jonathan Birch, Luc Bovens, Paul Daniell, Katie Steele, Jeremy Strasser, Philippe van Basshuysen, and several anonymous referees for their comments on earlier drafts, and to audiences at the LSE PhD Student work-in-progress seminar, the 2018 Society for Exact Philosophy Conference at the University of Connecticut, and the 2018 Philosophy of Science Association Conference in Seattle.

a true generalization in physics, while Belot (2003), Earman (2004), Roberts (2013, 2016), and Norton (2016) have all claimed that the principle, precisely stated, is either false or a tautology that is at least sometimes vacuous.[1]

The debate over the status of Curie's Principle has centered around various cases in classical and quantum mechanics, and whether or not they validate the principle. On both sides of this debate, causal relations are mostly cashed out by appealing to an intuitive or "folk" understanding of cause and effect.[2] This recourse to intuition creates substantial room for maneuvering for the defender of Curie's Principle. In the face of counterexamples to the principle, it is always possible to claim that the putative counterexample does not correctly represent the causal metaphysics of a given scenario, and to reinterpret the example so that the principle is non-vacuously true. What is missing from this debate is an independent understanding of what counts as a causal relation, such that we can determine whether or not this understanding implies the truth of Curie's Principle.

In light of this observation, it is surprising that the discussion of Curie's Principle in philosophy of physics has been conducted without any consideration of the much larger body of literature on graphical causal modeling, an approach pioneered by Pearl (2000), Halpern (2000), and Spirtes et al. (2000). In this tradition, causal relationships are represented using the resources of graph theory. Causal relata are represented as random variables at the vertices of a graph, with one vertex being interpreted as the direct cause of another just in case there is a directed edge relating the causal variable to the effect variable. In the deterministic context, the graphical approach to causal modeling is usually articulated using the theory of *structural equation models*, or 'SEMs', and *directed acyclic graphs*, or 'DAGs'. To be clear, SEMs and DAGs are two formal approaches that complement each other in formalizing the causal structure of a target system; they are not competing. For the purpose of this paper, the crucial feature of a graphical causal model is that it satisfies the *Markov condition*, which states that variables in a causal model are screened off from their non-effects by their causes. Where the Curie's Principle literature is vague about what constitutes a causal relation in a model of a physical system, the graphical causal modeling framework is precise; one random variable is a direct cause of the other if there is a directed edge from the first to the second in a graphical model of the system that satisfies the Markov condition. The meanings of the technical terms

[1] For an older philosophical discussion of Curie's Principle, see Chalmers (1970).

[2] For instance, Roberts (2013) treats the initial state of a harmonic oscillator as a cause of subsequent states of the oscillator, without any substantive argument for this treatment.

used in the previous sentence will be made more precise in what follows.

This lack of a connection between the literature on Curie's Principle and the literature on causal modeling is surprising in light of the significant role that the graphical formalism for causal modeling plays in contemporary discussions of causation throughout philosophy of science, including philosophy of physics. Woodward's (2003) widely-cited account of causation and causal explanation in the sciences is derived in large part from the graphical causal modeling formalism developed in Pearl (2000), Halpern (2000), and Spirtes et al. (2000). In the specific context of physics, Frisch (2014) puts forward a book-length treatment of causation in physics that aims in large part to show that physical theories obey the axioms of graphical causal modeling. Further, in recent work, Ismael (2018) has herself defended an account of the physical reality of causation that she explicitly states is consonant with Pearl's account of graphical causal modeling. Finally, recent work in physics by Costa and Shrapnel (2016) attempts to use an extension of the graphical causal modeling framework to formalize causal relations in quantum mechanics. To be clear, I do not wish to claim that connections between Curie's Principle and causal modeling are the *only* interesting connections that one can draw between Curie's Principle and other accounts of causation. Indeed, an interesting avenue for additional work would be to consider what connections exist between Curie's Principle and Dowe's (2000) account of causation in physics. However, for the purposes of this paper I confine myself to exploring the connection between Curie's Principle and graphical causal modeling.

My aim in this paper is to make progress on the debate over the status of Curie's Principle by formulating the principle in the language of causal modeling, in order to determine what logical connections exist between Curie's Principle on the one hand and the adequacy conditions for causal graphs on the other. If one can show that any adequate causal graph must be such that symmetries are preserved across causal relations, then this would show that defenders of Curie's principle and those who take causal modeling to provide a compelling account of causation in the sciences, including physics, share a common argumentative ground. If, on the other hand, there is no such connection, then the defender of Curie's Principle should be read as offering an account of causation that is logically independent from at least one of the other prominent accounts of causation on offer. In what follows, I will show that whether or not there is a logical connection between Curie's Principle and the axioms of graphical causal modeling depends on how we represent symmetry transformations formally. If we represent a symmetry of a random variable as the composition of some transformation function and the random variable in question that preserves the value of the random variable (call this the "composition definition" of a symmetry), then we can give examples of causal models that satisfy the Markov condition and yet are such that Curie's Principle is false. On the other hand, if we represent a symmetry transformation on a random variable as a transformation on the domain of a random variable that preserves the value of that random variable (call this the "domain transformation definition" of a symmetry), then an analogue of Curie's Principle is implied by the causal modeling formalism. In this way, my results here help to provide a clearer map of the argumentative territory when it comes to the formalization of causal relations and Curie's Principle: if either the defenders or the detractors of Curie's Principle are correct in their thesis about the symmetry-preserving nature of causal relations, then they will have to either defend a particular formalization of the concept of a symmetry transformation, or else say something about where the causal modeling framework goes wrong.

The plan for this paper is as follows. In Section 2, I provide background details on Curie's Principle and the recent debate over its status as a physical principle. In Section 3, I review the graphical modeling approach to the representation of causal structure, and propose a formulation of Curie's Principle using the formalism of causal modeling. In Section 4, I give a counterexample showing that, under the composition definition of a symmetry, a graphical causal model need not be such that the causal relations represented by the graph must satisfy Curie's Principle. I then prove a result showing that, under the domain transformation definition of a symmetry, a graphical causal model must obey Curie's Principle. In Section 5, I offer concluding remarks.

2 Curie's Principle

Pierre Curie's interest in the role that symmetries might play in the relationship between cause and effect began with his study of the piezoelectricity of crystals in 1894 (Castellani 2003, 323). Very briefly, Curie discovered that by compressing a crystal so as to eliminate the existence of certain geometric symmetries in its molecular structure, one could also induce asymmetries in the crystal's electric field, creating a displacement current. From this observation, Curie made the conjecture that symmetry-preservation is a necessary feature of causal relations in general. That is, he held that "effects are phenomena which always require a certain asymmetry [in their causes] in order to arise. If this asymmetry does not exist, the phenomena are impossible" (Curie

1894; translated in Brading and Castellani 2003, 312). Since all effects must share asymmetries with their causes, it follows from Curie's claim that any symmetry of a cause must also be a symmetry of its effect. Indeed, Curie writes that "when certain causes produce certain effects, the symmetry elements of the causes must be found in their effects" (Curie 1894; translated in Brading and Castellani 2003, 313).

Brading and Castellani take the physical significance of Curie's Principle seriously, writing that the conjecture "offers a falsification criterion for physical theories", and that "a violation of Curie's Principle may indicate that something is wrong in the physical description" (2003, 10). They claim that Curie's Principle can be motivated by appeal to the principle of sufficient reason; arguing that "the breaking of [an] initial symmetry cannot happen without a reason, or *an asymmetry cannot originate spontaneously*" (2003, 9-10, emphasis theirs). Similarly, Ismael (1997) argues that if the laws of nature are deterministic, then the entailments of those laws must preserve any symmetries of the laws themselves. These statements suggest that Curie's Principle is a powerful metaphysical constraint on acceptable representations of physical reality. On a plausible reading, these authors seem to be suggesting that whatever form our final physics takes, it must not allow for spontaneous asymmetries between a cause and its effects.

Norton (2016) formalizes Curie's principle as follows. He begins by representing possible causes and effects as the sets $\mathcal{C} = \{C_1, C_2, ..., C_n\}$ and $\mathcal{E} = \{E_1, E_2, ..., E_m\}$. He then stipulates that the *symmetry groups* $G_\mathcal{C}$ and $G_\mathcal{E}$ are sets satisfying the following conditions:

1. If $S_\mathcal{C} \in G_\mathcal{C}$, then for any C_j, $S_\mathcal{C} \cdot C_j = C_j$.

2. If $S_\mathcal{E} \in G_\mathcal{E}$, then for any E_i, $S_\mathcal{E} \cdot E_i = E_i$ (Norton 2016, 1016).

This formalization allows him to state Curie's Principle as follows (1017):

> **Curie's Principle:** If causes $\mathcal{C}$ admit symmetries $G_\mathcal{C}$ and are mapped to effects $\mathcal{E}$ via the function $E_i = f(C_j)$, then there is a symmetry group $G_\mathcal{E}$ that is isomorphic to $G_\mathcal{C}$.[3]

[3]Norton actually defines Curies Principle as a tautology of the following form: If (1) causes $\mathcal{C}$ admit symmetries $G_\mathcal{C}$ and are mapped to effects $\mathcal{E}$ via the function $E_i = f(C_j)$, and (2) if causes $\mathcal{C}$ admit symmetries $G_\mathcal{C}$ and are mapped to effects $\mathcal{E}$ via the function $E_i = f(C_j)$, then there is a symmetry group $G_\mathcal{E}$ that is isomorphic to $G_\mathcal{C}$, then there is a symmetry group $G_\mathcal{E}$ that is isomorphic to $G_\mathcal{C}$. Norton represents Curie's Principle as a tautology for dialectical reasons: since proponents of Curie's Principle as a falsification criterion for physics generally argue that the principle is tautological, opponents such as Norton want to grant them this claim, but then show that the tautology is vacuous. In practice, however, such debate always concerns the truth value of claim (2) in the antecedent of the tautological formulation of Curie's Principle. It is easier, therefore, to take 'Curie's Principle' to refer to claim (2) in the tautological formulation.

Several aspects of Norton's formalization are left under-specified. This is deliberate, as Norton intends to give a "sparse" characterization of the principle. Nevertheless, it is worth being explicit about which components of the principle are under-specified. First, it is not specified what, formally speaking, the elements of $\mathcal{C}$ and $\mathcal{E}$ are; in Norton's examples, they are interchangeably functions or numbers. Second, the nature of the operations $S_{\mathcal{C}} \cdot C_j$ and $S_{\mathcal{E}} \cdot E_i$ is not specified. If, for example, causes and effects are real numbers, then it is natural to read $S_{\mathcal{C}} \cdot C_j$ as the value of a function $\varphi(C_j)$, where $\varphi : \mathbb{R} \to \mathbb{R}$. On the other hand, if C_j is a function, then one might read $S_{\mathcal{C}} \cdot C_j$ as the composition $S_{\mathcal{C}} \circ C_j$. As I will demonstrate in what follows, considering Curie's Principle in the context of graphical causal models allows for decidedly less ambiguity in the formulation of the principle and the assessment of its status.

Roberts (2013) argues that Curie's Principle is falsified by a simple case in classical mechanics. Consider a harmonic oscillator in the form of a bob on a spring. At $t = 0$, the spring is compressed so that the bob is at a certain position. Let the state of the bob be a pair (q, p) representing its position and momentum along a single spatial dimension. The momentum term p is a one-dimensional momentum vector, such that its sign changes depending on the direction of the momentum. The bob has zero velocity at $t = 0$, and therefore has zero momentum. When an object has zero momentum, its momentum is time-reversal invariant; regardless of whether time is running backwards or forwards, the momentum remains zero. However, the position of the bob at $t = 0$, i.e. how far down we compress the spring, causes the spring to have a certain non-zero momentum once the stress on the spring is released. This non-zero momentum changes through time, and is therefore not time-reversal invariant. Thus, even though the initial state of the spring is a cause of the subsequent states of the spring, the initial state of the spring is symmetric with respect to time reversals, while subsequent states of the spring are not symmetric with respect to time reversals. Thus, the system violates Curie's Principle. In Roberts (2016), other cases in physics are described in which causes are invariant under symmetries like rotation but their effects are not, thereby putatively falsifying or rendering vacuous Curie's principle.

More precisely, let $(q(0), p(0))$ be the state of the bob on the spring at $t = 0$, and let this state be the sole element of the set of causes $\mathcal{C}$. Let it also be the case that for all times $t \in (0, \infty)$, $(q(t), p(t)) = (\cos(2t), -\sin(2t))$. Thus, at $t = 0$, the state of the bob is $(1, 0)$, and at some fixed future time t the state of the bob is $(\cos(2t), -\sin(2t))$. Let this future state of the bob be the sole member of the set of effects $\mathcal{E}$. Let $S_{\mathcal{C}} : (q, p) \mapsto (q, -p)$ be a transformation on states of the bob, representing the change in the direction of the momentum of the bob under time reversal. As $(1,0) = (1,-0)$, $S_{\mathcal{C}}$ is clearly a symmetry of the initial state of the spring. Further, for a fixed t we can define a function such that $(1,0) \mapsto (q(t), p(t))$. Thus, the two antecedent conditions for Curie's Principle are satisfied. However, for any $t \in (0, 2\pi)$, $(q(t), p(t)) \neq (q(t), -p(t))$. Thus, the symmetry group $G_{\mathcal{C}}$ is not isomorphic to the symmetry group $G_{\mathcal{E}}$, and so Curie's Principle is putatively violated (Roberts 2013, 580-3).

In response to the example described above, Castellani and Ismael (2016) argue that the final state of the bob on the spring *is* symmetric across time-reversal, under a non-standard definition of time reversal due to Albert (2000) and Callender (2000). They argue that we have to consider "the intrinsic, instantaneous state of the spring", which is specified by the spring's position and not its momentum. Their idea is that we should treat any state of the bob on the spring after $t = 0$ as being frozen in time, such that we can only consider the symmetries of its position, and not its momentum. Since time-reversal is a symmetry of any instantaneous state of the oscillator in this non-standard sense, Ismael and Castellani conclude that Curie's Principle is still true and Curie's Principle is not vacuous.[4]

The exchange between Roberts on the one hand and Ismael and Castellani on the other is characteristic of the recent debate over the status of Curie's Principle. Roberts starts with an intuitive understanding of cause and effect wherein the state of the spring when it is compressed is treated as a cause, and the state of the spring after it is released is treated as an effect, where the state of the spring includes its momentum. On this understanding of cause and effect, Curie's Principle is rendered false. Ismael and Castellani respond that if we take the instantaneous position of the spring to be a cause, then the effect of that cause must be some subsequent instantaneous description of the spring, thereby salvaging the truth of Curie's Principle. Thus, the debate over whether or not Curie's Principle is violated by any dynamical system boils down to whether or not we are required to represent causes and effects in a way that ensures symmetry preservation.

As mentioned in the introduction, what is striking about this debate is the way in which a commonsense or folk understanding of causation is taken for granted when discussing putative counterexamples to Curie's Principle. Roberts, Ismael and Castellani all assume that we can treat the initial state of the spring as the cause of some subsequent state of the spring, and then debate over whether Curie's Principle provides a constraint on how we can represent this subsequent state of the spring, in light of this causal relationship; at no point do they appeal to

[4]Roberts (2017) argues against this interpretation of time reversal.

any independent understanding of causation. The closest thing to a general account of causation is provided by Roberts, who says that "an 'effect' is a state and a 'cause' is another state related to the first by a deterministic law"; this minimal definition would indeed render the verdict that the initial state of the spring is a cause, and a subsequent state of the spring is its effect (2013, 579). Nowhere in Ismael and Castellani's response to Roberts (2016, 1011-2) do they take issue with Roberts' characterization of the causal relata in this scenario, focusing instead on the issues with time reversal discussed above. As such, this debate appears to be a clash of competing intuitions over how to correctly describe the states of some dynamical system.

In what follows, I pursue a novel method of adjudicating between the two views. As discussed in the introduction, I will consider causation from the point of view of graphical causal models, to see if there is any logical connection between this notion of causation and Curie's Principle. As I will make clear in what follows, the graphical causal modeling formalism lends precision to this debate by specifying that causal relata must be random variables that stand in a specific relation to one another within the formal model; namely, an effect variable must be a function defined on the range of possible values that can be taken by each of its causes, which are themselves random variables. I will then show that whether or not Curie's Principle holds in this context depends upon one's formalization of the concept of a symmetry transformation.

3 Curie's Principle and Causal Graphs

A causal model is a triple $\mathcal{M} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$ where where $\mathcal{V}$ is a set of random variables, $\mathcal{E}$ is a set of ordered pairs of those variables, or edges, and $\mathcal{F}$ is a set of structural equations. Note that any random variable V_i is a function from some underlying space of possibilities Ω into some range R_{V_i}. We assume that all random variables in a causal model share a common domain of possibilities Ω, and are all measurable with respect to some probability space (Ω, Σ, P), where Σ is an algebra on Ω (i.e. a set of subsets of Ω that is closed under union, intersection, and complement). We also introduce some graphical terminology. If there is a pair $(X, Y) \in \mathcal{E}$, then X is a *parent* of Y. If there is a chain of edges from X to Y, then X is an *ancestor* of Y and Y is a *descendant* of X. We stipulate that $\mathcal{E}$ defines a directed acyclic graph, or DAG, meaning that no variable $X \in \mathcal{V}$ is a descendant of itself; i.e. there are no cycles in the graph. So, in a simple causal graph $\mathcal{E} = X \to Y \to Z$, X is a parent and ancestor of Y, Y is a parent and ancestor of Z, X is an ancestor but not a parent of Z, Z and Y are descendants of X, and Z

is a descendant of Y.

The set of structural equations $\mathcal{F}$ is a set of equations of the form $v_i = f_i(\mathbf{pa}_i, \epsilon_i)$, where v_i is the value of a variable $V_i \in \mathcal{V}$, $\mathbf{pa}_i$ is the vector of values taken by V_i's parents in $\mathcal{E}$, and ϵ_i is an error term reflecting an exogenous source of indeterminacy in the value of the variable V_i. In other words, each equation represents how the value of each variable in the model is given by a function of the values taken by some subset of the other variables in the model, plus exogenous error. $\mathcal{F}$ contains one and only one such equation for each variable in $\mathcal{V}$. Let two variables V_i and V_j be independent, conditional on V_i's parents $\mathcal{PA}_i$, if for all values v_i and v_j, and all sets of values $\mathbf{pa}_i$:

$$P(V_i = v_i, V_j = v_j | \mathcal{PA}_i = \mathbf{pa}_i) = P(V_i = v_i | \mathcal{PA}_i = \mathbf{pa}_i) P(V_j = v_j | \mathcal{PA}_i = \mathbf{pa}_i) \tag{1}$$

Pearl (2000, 30) proves that if the error terms for each variable are not correlated, then the graph $(\mathcal{V}, \mathcal{E})$ will satisfy the following condition:

> **Markov Condition:** Any variable in $\mathcal{V}$ is independent of its non-descendants in $\mathcal{E}$ conditional on its parents in $\mathcal{E}$.

Within the causal modeling literature, this condition is taken to provide the primary justification for the causal interpretation of the edges in $\mathcal{E}$. In more detail, let us interpret the parenthood relation between two variables as a relation of *direct causation*, such that if V_i is a parent of V_j, then V_i is a direct cause of V_j. Under this interpretation, Pearl's result shows that if the value of a variable is solely a function of its direct causes, and if all direct causes of two or more variables are included in the model, then the Markov condition holds. There is some intuitive appeal to the idea that an effect is a function of all and only its direct causes, and indeed, Spirtes et al. (2000) take the acceptance of such a condition to be a primitive commitment of causal modeling, along with the accompanying commitment that causation is the ancestral of direct causation, so that if V_i is an ancestor of V_j, then V_i is a cause of V_j. Note that the same functional definition of causation is accepted by proponents of Curie's Principle; Ismael (1997) holds that causes and effects are "mutually exclusive and jointly exhaustive event types" related to one another by functions (169).

I confine myself in this paper to discussion of Curie's Principle under determinism, in large part because Ismael takes the principle to only be necessarily true in deterministic contexts (1997, 168). Thankfully, the special case in which variables are completely deterministic functions of their parents, with no exogenous error, is entirely consistent with the causal modeling framework. Let an *endogenous* variable in a model $\mathcal{M}$ be a variable with at least one parent, according to $\mathcal{E}$. Suppose that for each endogenous variable V_i, the function $f_i(\mathbf{pa}_i, \epsilon_i)$ that determines its value is such that the error term is redundant; for all ϵ_i, $f_i(\mathbf{pa}_i, \epsilon_i) = f_i(\mathbf{pa}_i)$. Under these conditions, the Markov condition is satisfied as long the error terms for all exogenous variables are not correlated. To see this, note that if the value of any endogenous V_i is determined by an error-free function of V_i's parents, then for any v_i, V_j, v_j, and $\mathbf{pa}_i$, $P(V_i = v_i, V_j = v_j | \mathcal{PA}_i = \mathbf{pa}_i) = P(V_j = v_j | \mathcal{PA}_i = \mathbf{pa}_i)$ if $P(V_i = v_i | \mathcal{PA}_i = \mathbf{pa}_i) = 1$ or $P(V_i = v_i, V_j = v_j | \mathcal{PA}_i = \mathbf{pa}_i) = 0$ if $P(V_i = v_i | \mathcal{PA}_i = \mathbf{pa}_i) = 0$.[5] Thus, where error terms of endogenous variables in a causal model are effectively non-existent, such that endogenous variables are entirely determined by the parents, endogenous variables are independent of all variables in the model, conditional on their parents, which in turn implies that the Markov condition holds.

Using the formalism of causal models, we can now re-define Curie's Principle graphically:

> **Graphical Curie's Principle**: For any variable X in a causal model $\mathcal{M} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$, if $G_{\mathcal{PA}_X}$ is the symmetry group for the parents of X in $\mathcal{E}$ and G_X is the symmetry group for X, then there is a symmetry group G_X to which $G_{\mathcal{PA}_X}$ is isomorphic.

The definition clarifies one aspect of Curie's Principle that is left under-specified in Norton's formulation of the principle, namely, the formal character of causes and effects. In the graphical framework, both causes and effects are represented as random variables, related by an edge in the graph. Since the range of a random variable partitions the domain of possibilities on which that random variable is defined, any random variable corresponds to a particular set of types of ways the world can be. Thus, the representation of causes and effects as random variables is in keeping with Ismael's definition of causes and effects as sets of mutually exclusive and jointly exhaustive event types, albeit with the added precision of defining these sets of event types as the range of a random variable.

For the sake of clarity, I note here that the graphical account of causation allows for the existence of *many* adequate causal models of a given target system. The same target system can be adequately modeled in multiple different ways, depending on one's choice of variables; see Spirtes (2007) and Eberhardt (2016) for rigorous demonstrations of this point. For this

[5] Arntzenius (1992, 236) proves that the Markov condition, which he calls the "restricted common cause principle", applies to such deterministic systems.

reason, it is a category mistake to ask whether *the* adequate causal model of a given target system must satisfy Curie's Principle. Instead, I ask here whether *any* adequate causal model of a given target system satisfies Curie's Principle.

It remains to formalize the nature of a symmetry transformation on a random variable. As mentioned in the introduction, there are two ways to formalize this notion, and our choice of formalization has consequences for the status of the graphical version of Curie's Principle. The first such definition of a symmetry transformation is as follows:

> **Composition Definition:** A function $S : R_X \rightarrow R_S$ is a symmetry of a random variable X if and only if, for all ω in the domain of X, $(S \circ X)(\omega) = X(\omega)$.

In other words, on the composition definition, S is a symmetry of a random variable X if and only if the composition of S and X is identical to the function X. On this definition, the domain of S must be the range of X. I take this definition of symmetry to be closest to the definition originally used by Ismael (1997, 171). Alternatively, one can define a symmetry of a random variable as follows:

> **Domain Transformation Definition:** A function $S : \Omega \rightarrow \Omega$ is a symmetry of a random variable X if and only if, for all ω in the domain of X, $X(S(\omega)) = X(\omega)$.

In other words, S is a domain transformation symmetry of X if and only if applying the transformation S to the domain of X never changes X's value. In a different context, Jantzen (2019) defines a symmetry of a random variable in this way. Both definitions of a symmetry can represent physically meaningful symmetries, such that the graphical Curie's Principle tracks the relevant physical principle defended by proponents of Curie's Principle. For instance, if we let a random variable $X : \Omega \rightarrow R_X$ be such that the elements of R_X are various possible sets of points in three-dimensional space comprising a sphere, then one can represent a rotation or reflection of the sphere via the composition transformation definition of a symmetry. Alternatively, we could represent a sphere using a random variable $Y : \Omega \rightarrow \{0, 1\}$, where Ω is a set of coordinates in three-dimensional space, $Y(\omega) = 1$ if ω is an element of the sphere, and $Y(\omega) = 0$ otherwise. Here, we could rotate the sphere by transforming the underlying space Ω. In what follows, I show that, under the composition definition of a symmetry of a random variable, the graphical version of Curie's Principle does not hold. However, under the domain transformation definition of a symmetry, graphical causal models necessarily satisfy Curie's Principle.

4 Clarifying the Status of the Graphical Curie's Principle

Let us use an example from electromagnetism to demonstrate that the Markov condition can be satisfied by a causal model in which Curie's Principle is violated.[6] Let Ω be a sample space of possible worlds, each of which maximally specifies the physical state of a given material. Let $B^\rho : \Omega \to R_{B^\rho}$ be a random variable whose values represent the possible residual flux density fields of the material. That is, each $b_i^\rho \in R_{B^\rho}$ is a function $b_i^\rho : \mathbb{R}^3 \to \mathbb{R}^3$, or $b_i^\rho : \mathbf{u} \mapsto b_i^\rho(\mathbf{u})$, where the domain represents possible spatial locations within the material and the range represents the set of possible residual flux vectors at any point on the material. Let $M : \Omega \to R_M$ be a random variable whose values represent the possible magnetization fields of the material. That is, each $m_k \in R_M$ is a function $m_k : \mathbb{R}^3 \to \mathbb{R}^3$, or $m_k : \mathbf{u} \mapsto m_k(\mathbf{u})$, where the domain represents possible spatial locations within the material and the range represents the set of possible magnetization vectors at any point on the material. Let $J^M : \Omega \to R_{J^M}$ be a random variable whose values represent the possible magnetization current fields of the material. That is, each $j_l^M \in R_M$ is a function $j_l^M : \mathbb{R}^3 \to \mathbb{R}^3$, or $j_l^M : \mathbf{u} \mapsto j_l^M(\mathbf{u})$, where the domain represents possible spatial locations within the material and the range represents the set of possible magnetization current vectors at any point on the material.

Together, $\{B^\rho, M, J^M\}$ make up the variable set $\mathcal{V}$ of a causal model $(\mathcal{V}, \mathcal{E}, \mathcal{F})$. Let the edges $\mathcal{E}$ be defined so that $\mathcal{E} = B^\rho \to M \to J^M$; thus, the residual flux density field of the material causes its magnetization field, which in turn causes its magnetization current field. Finally, let $\mathcal{F}$ contain the following two structural equations:

$$m_k = \frac{1}{\mu_0} b_i^\rho \text{ where } \mu_0 \text{ is the permeability of free space.} \tag{2}$$

$$j_l^M = \nabla \times m_k \tag{3}$$

Thus, the values of M and J^M are both determined by functions on the range of their sole parent, in keeping with the graphical causal modeling formalism. It is worth noting that these structural equations are *functionals*, since they are mappings from one set of functions into another. While somewhat non-standard, there is no formal obstacle to deploying the graphical causal modeling formalism in this way. Note further that $\nabla \times m_k$ denotes the *curl* of the vector field m_k. The causal model defined here satisfies the Markov condition; since the value

[6] Both Chalmers (1970) and Roberts (2016) also use electrodynamic examples to demonstrate putative violations of Curie's Principle.

of each variable is wholly dependent on its sole parent, each variable is independent of its non-descendants, and indeed all other variables in the model, conditional on its parents.

Under the composition definition of a symmetry of a random variable, this causal model violates the graphical Curie's Principle. To see this, note first that any $m_k : \mathbf{u} \mapsto m_k(\mathbf{u})$ can be represented as a set of pairs:

$$m_k = \{(\mathbf{u}_1, m_k(\mathbf{u}_1)), \ldots, (\mathbf{u}_\infty, m_k(\mathbf{u}_\infty))\} \tag{4}$$

Let $S : R_M \to R_S$ be defined as follows:

$$S : \{(\mathbf{u}_1, m_k(\mathbf{u}_1)), \ldots, (\mathbf{u}_\infty, m_k(\mathbf{u}_\infty))\} \mapsto \{(\mathbf{u}_1, m_k(\mathbf{\Lambda u}_1)), \ldots, (\mathbf{u}_\infty, m_k(\mathbf{\Lambda u}_\infty))\} \tag{5}$$

where $\mathbf{\Lambda}$ is the transformation matrix for reflection about the origin:

$$\mathbf{\Lambda} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{6}$$

That is, S transforms a function m_k into a function that first reflects its input vector $\mathbf{u}$ about the origin, and then returns the value of m_k for that reflected vector. It is well known that any magnetization field is a *pseudovector* field, meaning that it returns the same value for an input vector and the reflection of that vector about the origin; for any m_k and any $\mathbf{u}$, $m_k(\mathbf{u}) = m_k(\mathbf{\Lambda r})$. Thus, for any m_k,

$$S(m_k) = \{(\mathbf{u}_1, m_k(\mathbf{\Lambda u}_1)), \ldots, (\mathbf{u}_\infty, m_k(\mathbf{\Lambda u}_\infty))\} = \{(\mathbf{u}_1, m_k(\mathbf{u}_1)), \ldots, (\mathbf{u}_\infty, m_k(\mathbf{u}_\infty))\} = m_k \tag{7}$$

and so $(S \circ m_k)(\omega) = m_k(\omega)$ for all $\omega \in \Omega$. Thus, S is a symmetry of M, according to the composition definition of a symmetry of a random variable.

However, S is not a symmetry of J^M under the composition definition of a symmetry. To see this, note that the curl of a pseudovector field is a *polar* vector field, and so each $j_l^M \in R_{J^M}$ is a polar vector field. Unlike pseudovector fields, polar vector fields do not return the same value for an input vector $\mathbf{u}$ and the reflection of $\mathbf{u}$ about the origin. So for any j_l^M, there exists a $\mathbf{u}$ such that $j_l^M(\mathbf{u}) \neq j_l^M(\mathbf{\Lambda v})$. It follows from this that there exists a j_l^M such that

$$S(j_l^M) = \{(\mathbf{u}_1, j_l^M(\mathbf{\Lambda u}_1)), \ldots, (\mathbf{u}_\infty, j_l^M(\mathbf{\Lambda u}_\infty))\} \neq \{(\mathbf{u}_1, j_l^M(\mathbf{u}_1)), \ldots, (\mathbf{u}_\infty, j_l^M(\mathbf{u}_\infty))\} = j_l^M \tag{8}$$

and so S is not a symmetry of J^M, despite M being the sole parent of J^M. This lack of isometry between the symmetry groups of M and J^M amounts to a violation of the graphical Curie's Principle under the composition definition of a symmetry.[7] One might object to this line of reasoning by arguing that the graphical model used here does not accurately represent the causal structure of its target system, on the grounds that one should not think of the residual flux density field of a material as causing its magnetization field, or its magnetization field as causing its magnetization current field. In response, I note that on Woodward's (2003) analysis of causation in terms of interventional counterfactuals, one can argue that the magnetization field of the material does cause its magnetization current field; a hypothetical, "surgical" intervention on the material leading to changes in its magnetization field, where those changes are independent of any other possible causes of the magnetization field, will lead to changes in the magnetization current field.

Alternatively, one might try to resist the conclusion that this amounts to a genuine counterexample to the graphical Curie's Principle by following Ismael in arguing that in real systems, apparent violations of Curie's Principle only indicate that there is a hidden variable that has been left out of the causal representation of the system. To illustrate, Ismael gives the example of frog zygotes that "start out as spherical cells suspended in an homogeneous seeming fluid and develop into highly structured organisms; almost every stage in their development introduces asymmetries not apparently present in the preceding stage" (1997, 181). For her, this state of affairs constrains physical theorizing about the development of frog zygotes; any more advanced theory of zygote development must represent the initial state of the frog in a way that accounts for the asymmetries present in the later stages. In response, I argue that while this may be a plausible constraint on the class of models that we should consider when searching for a new theory of frog zygote development, this constraint serves an entirely epistemological and heuristic purpose, in that it makes the problem of model search more tractable. However, to

[7] This counterexample to the graphical Curie's Principle makes no mention of the variable B^ρ, and indeed the variable can be omitted without obviating any implications of the example. However, as two-variable, one-edge causal models trivially satisfy the Markov condition, I have chosen to include a third variable so as to make this feature of the causal model in question non-trivial.

impose this constraint as a falsification criterion for physical theories is an exercise in *a priori* science. If we have empirically well-motivated reasons for proposing a causal model of zygote development that allows for such spontaneous symmetry breaking, Curie's Principle is not a compelling reason to see such a model as necessarily incomplete or defective, at least when symmetries are defined using the composition definition.

On the other hand, if we adopt the domain transformation definition of a symmetry, then the graphical Curie's Principle is implied by the causal modeling formalism. More precisely, the following proposition is true:

Proposition 1. *If S is a symmetry (in the domain transformation sense) of each of the parents $\mathcal{PA}_X$ of a variable X in a causal model $\mathcal{M}$, then S is a symmetry of X.*

Proof. Let $\mathcal{PA}_X = \{V_1, \ldots, V_n\}$. Let $\varphi : \Omega \to \mathbb{R}^n$ be such that $\varphi : \omega \mapsto [V_1(\omega), \ldots, V_n(\omega)] = \mathbf{pa_X}$. If S is a symmetry of each variable in $\mathcal{PA}_X$, then for all ω, $\varphi(\omega) = \varphi(S(\omega))$. Recall that $X : \Omega \to R_X$, and that for all x_i, $x_i = f_X(\mathbf{pa_X})$. This means that, for all ω, $X(\omega) = (f_X \circ \varphi)(\omega)$. Thus, if, for all ω, $\varphi(\omega) = \varphi(S(\omega))$, then for all ω, $X(\omega) = X(S(\omega))$, and so S is a symmetry of X, under the domain transformation definition. □

If any symmetry of $\mathcal{PA}_X$ is also a symmetry of X, then there is a symmetry group G_X that is isomorphic to the symmetry group $G_{\mathcal{PA}_X}$, so that the causal model in question satisfies the graphical formulation of Curie's Principle under the domain transformation definition of a symmetry.

In addition to demonstrating the soundness of a version of Curie's Principle in causal graphs, Proposition 1 also speaks to the epistemological significance of Curie's Principle. As Jantzen (2019) argues, simulations or actual experiments that perform transformations on a given sample space play an important role in the *validation* of graphical causal models defined on that sample space, where the validation of a causal model is a process of checking that the causal structure stipulated in the model accurately represents the causal structure of the system. Proposition 1 lends further support to this view, as it implies that one can check whether a hypothesized causal relation holds between two variables by checking for an isometry between the symmetry groups of the two variables, provided that one uses the domain transformation definition of a symmetry. As Jantzen does not comment on Curie's Principle in his work on symmetry transformations and graphical causal model validation, the result proved here establishes a previously unacknowledged connection between these two areas of the literature.

5 Conclusion

This paper seeks an answer to a well-defined question: does the graphical causal modeling framework, which is widely deployed throughout philosophy of science and the sciences themselves to represent the causal structure of systems, imply Curie's Principle, which is controversially argued by some to be a necessary condition on a causal representation of any system? I then show that the answer to this question depends on the formal definition of a symmetry transformation. If one takes the graphical causal modeling formalism to be well-motivated, then Curie's Principle is only true if one adopts the domain transformation definition of a symmetry transformation. On the other hand, if one wishes to maintain the physical importance of Curie's Principle, then one must either adopt the domain transformation definition of a symmetry transformation, or else reject the applicability of the graphical causal modeling formalism in at least some scenarios.

References

D. Z. Albert. *Time and Chance*. Harvard University Press, 2000.

F. Arntzenius. The common cause principle. In *PSA: Proceedings of the Biennial meeting of the philosophy of science association*, volume 1992, pages 227–237. Philosophy of Science Association, 1992.

G. Belot. Notes on symmetries. *Symmetries in physics: Philosophical reflections*, pages 393–412, 2003.

K. Brading and E. Castellani. *Symmetries in physics: philosophical reflections*. Cambridge University Press, 2003.

C. Callender. Is time 'handed' in a quantum world? *Proceedings of the Aristotelian Society*, 100:247–269, 2000.

E. Castellani. On the meaning of symmetry breaking. *Symmetries in physics: Philosophical reflections*, pages 321–334, 2003.

E. Castellani and J. Ismael. Which curie's principle? *Philosophy of Science*, 83(5):1002–1013, 2016.

A. F. Chalmers. Curie's principle. *British Journal for the Philosophy of Science*, 21(2):133–148, 1970.

F. Costa and S. Shrapnel. Quantum causal modelling. *New Journal of Physics*, 18(6):063032, 2016.

P. Curie. On symmetry in physical phenomena, symmetry of an electric field and of a magnetic field. *Journal de physique*, 3:401, 1894.

P. Dowe. *Physical causation*. Cambridge University Press, 2000.

J. Earman. Curie's principle and spontaneous symmetry breaking. *International Studies in the Philosophy of Science*, 18(2 & 3):173–198, 2004.

F. Eberhardt. Green and grue causal variables. *Synthese*, 193(4):1029–1046, 2016.

M. Frisch. *Causal reasoning in physics*. Cambridge University Press, 2014.

J. Y. Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12: 317–337, 2000.

J. Ismael. Curie's principle. *Synthese*, 110(2):167–190, 1997.

J. Ismael. Against globalism about laws. In I. Peschard and B. van Fraassen, editors, *The Experimental Side of Modeling*. University of Minnesota Press, 2018.

B. C. Jantzen. Dynamical symmetries and model validation. In *Algorithms and Complexity in Mathematics, Epistemology, and Science*, pages 153–176. Springer, 2019.

J. D. Norton. Curie's truism. *Philosophy of Science*, 83(5):1014–1026, 2016.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

B. W. Roberts. The simple failure of curie's principle. *Philosophy of Science*, 80(4):579–592, 2013.

B. W. Roberts. Curie's hazard: From electromagnetism to symmetry violation. *Erkenntnis*, 81 (5):1011–1029, 2016.

B. W. Roberts. Three myths about time reversal in quantum theory. *Philosophy of Science*, 84 (2):315–334, 2017.

P. Spirtes. Variable definition and causal inference. *Proceedings of the International Congress for Logic, Methodology and Philosophy of Science*, 2007.

P. Spirtes, C. Glymour, S. N., and Richard. *Causation, Prediction, and Search.* Mit Press: Cambridge, 2000.

J. Woodward. *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press, 2003.