# The causal structure of natural kinds

**Olivier Lemeire**

## Abstract

One primary goal for metaphysical theories of natural kinds is to account for their epistemic fruitfulness. According to cluster theories of natural kinds, this epistemic fruitfulness is grounded in the regular and stable co-occurrence of a broad set of properties. In this paper, I defend the view that such cluster theories can only provide a partial account of the epistemic fruitfulness of kinds. I argue that they can indeed account for the projectibility of natural kinds but not for several other epistemic operations that natural kinds support. Natural kinds also play a role in scientific explanations and categorizations. A theory of natural kinds can only fully account for these additional kind-based epistemic practices if it includes an analysis of their causal structure.

## Introduction

Since Goodman's (1954) argument that inductive inferences require projectible categories, natural kind theorists have been interested in accounting for their projectibility, that is, for the aptness of natural kinds to support inductive inferences (Boyd, 1991).[1] When one observes that some members of a natural kind have a property, one can draw a legitimate inductive inference to other members of the kind having that same property. A natural kind predicate like 'metal,' for example, is projectible relative to the predicate 'good electrical conductivity,' because based on our observations of some metals we can legitimately infer that other members of that same kind will also be good electrical conductors.[2] In fact

members of a natural kind tend to have many such projectible properties in common. In Millikan's words: "a natural kind corresponds not just to a projectable predicate, but must figure as the subject of many empirical generalizations. No science consists of a single generalization, nor of a heap of generalizations about different kinds of things. A science begins only when, at minimum, a number of generalizations can be made over instances of a single kind" (Millikan, 1999, p. 48).

According to causal theories of natural kinds, this multiple projectibility of kinds is causally grounded. A metaphysical theory of natural kinds should therefore describe this causal ground. Essentialists would have it that natural kinds are associated with an essential and intrinsic property (or set of properties) the instantiation of which causes many other properties to be instantiated as well, thereby explaining their regular co-occurrence (Ellis, 2001; Wilkerson, 1988). Others have argued that the causal ground of the multiple projectability of kinds need not always be a shared essential property but can also be an extrinsic homeostatic mechanism (Boyd, 1999a; Kornblith, 1993). In the case of biological species, for instance, this mechanism could consist of a variety of factors that explain the similarities of the members of a species, like reliable reproductive copying mechanisms, stable selective forces, and so on. Though the details of these causal theories vary, what they have in common is the view that properties do not regularly co-occur by accident and hence that a causal theory is needed to account for the epistemic fruitfulness of kinds.

Cluster theorists, however, argue that it is unnecessary to resort to causal notions to account for the epistemic fruitfulness of kinds.[3] Most of them hold that it is not always the case that the clustering of properties is causally grounded. Anjan Chakravartty (2007), for example, has introduced the notion of the *sociability* of properties, explaining that some properties just 'like' each other's company and tend to be systematically instantiated together, like the basic properties of elementary particles. Similarly, Häggqvist (2005) has argued for a position called *bare projectibilism*, claiming that not all projectible kinds are causally grounded and that this should not deter us from considering them natural kinds.

The most promising cluster theory to recently challenge the causal view is that of Slater (2013, 2015). Like other cluster theorists, Slater holds that the regular co-occurrence of the properties associated with a kind is not always causally grounded. More importantly, however, he argues that analyzing the metaphysical nature of kinds in causal terms does not contribute to our understanding of their epistemic usefulness, even in those cases where they are *de facto* causally grounded. As he sees it, it is the fact that the properties associated with natural kinds cluster in a counterfactually stable manner that makes them epistemically useful. A metaphysical theory that aims to account for the multiple projectibility of natural kinds can therefore suffice with analyzing the precise nature of this nomic clustering. A similar view is voiced by Chakravarrty when he says that "it is the fact that members of kinds share properties, to whatever degree, that underwrites the inductive generalizations and predictions to which these categories lend themselves" (2007, p. 107).

In this paper, I respond to these cluster theories and argue that they only provide a partial account of the epistemic fruitfulness of kinds. After all, natural kinds are not only useful because they support many inductive generalizations, they are also referred to in successful kind-based explanations and they allow for the categorization of their members. Although the explanatory and taxonomic (categorizing) uses of natural kinds have not received the same level of attention in the literature as their projectibility, they are just as crucial to the ways in which natural kinds support scientific reasoning. Yet as I will argue, the success of kind-based explanations and categorizations is (in many cases) not grounded in the clustering of properties but rather in the structure of the causal relations between these properties. An account of these epistemic uses of natural kinds requires a theory about their causal structure.

Here is how I will proceed. In section 1, I introduce one prominent example of a causal theory of natural kinds that I will draw on later, namely Khalidi's *Simple Causal Theory*. Section 2 elaborates on the challenge to such causal theories posed by cluster theories, focusing on Slater's competing *Stable Cluster Theory.* I then argue that cluster theories like that of Slater cannot completely account for the explanatory

and categorizing uses of kinds in sections 3 and 4. I do this by explaining how the structure of the causal relations between the properties associated with a kind determines its explanatory and taxonomic usefulness.

**1. Causal Theories of Kinds**

According to causal theorists about natural kinds, the projectibility of kinds is causally grounded. They propose that "projectibility is the epistemic marker for the metaphysical relation of causality" (Khalidi, 2013, p. 80) and that "[t]he causal structure of the world as exhibited in natural kinds […] provides the natural ground of inductive inference" (Kornblith, 1993, p. 7).[4] Boyd explains it thus:

> [W]e are able to identify true generalizations in science – and in everyday life – because we are able to accommodate our inductive practices to the causal factors which sustain those generalizations. In order to do this – in order to frame such projectible scientific generalizations at all – we require a vocabulary […] which is itself accommodated to relevant causal structures. (Boyd, 2000, p. 60)

The challenge for causal theorists is to provide an account of the causal ground of kinds that is both informative enough to explain their projectibility and at the same time not too restrictive so that it would only apply to some epistemically fruitful kinds. While an essentialist account might work for chemical kinds, for instance, it does not apply to biological species the members of which do not share an intrinsic essence (Boyd, 1999b). In an effort to provide a more generalizable causal theory, Khalidi has advocated for a *Simple Causal Theory* (2013, 2015).[5] He proposes that "[o]n a simple causal theory of natural kinds, a natural kind is associated with a set of properties whose co-instantiation causes the instantiation of other properties" (2013, p. 80). Since in what follows I will build on this metaphysical theory to argue against the sufficiency of cluster theories, I will explain the essentials of it here.

Khalidi (2015) distinguishes between the causally prior properties and the causally secondary properties associated with a natural kind. Typically, kind predicates denote a (possibly loose) cluster of causally primary 'core' properties. When this primary set of properties is co-instantiated, it causes a whole set of causally secondary 'derivative' properties to be instantiated as well.[6] The co-instantiation of the primary core properties can be the result of causal relations between these properties but this need not be the case. In some instances, it may be that these core properties co-occur simply based on non-causal laws of co-existence, like the basic properties of elementary physical particles. Crucially, however, natural kinds support so many inductive generalizations because when co-instantiated, this small set of core properties tends to cause the instantiation of other properties, which in turn cause other properties, until these causal relations reach a dead-end. Natural kinds are in that sense like highly connected networks of causal relations. As an example, Khalidi (2013, p. 173) provides the following model depicting some of the causal relations between the properties associated with the kind *polymer*. The single core property of this kind ($P_1$) is to consist of a long molecular chain of many repeating monomers.

Figure 1 :
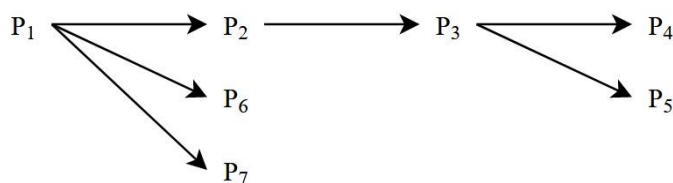


[T]he properties associated with polymers are not all on a par, since some of them are causally involved in the production of other properties. Instead of a picture whereby kind *K* is simply associated with a set of properties, {$P_1$, $P_2$, $P_3$, $P_4$, $P_5$, $P_6$, $P_7$}, we have a model according to which one property, $P_1$, is primary, in the sense that it is causally efficacious in instantiating $P_2$ (long molecular chains tend not to pack together in a uniform pattern), which is, in turn, causally efficacious in producing $P_3$, (nonuniformly packed molecules tend not to have a crystalline phase but enter into a semicrystalline phase), which then leads to $P_4$ (semicrystalline macromolecules show increased rigidity) and $P_5$ (increased tensile strength). Additionally, the primary property ($P_1$)

tends to lead directly to other properties – for instance, to high viscosity ($P_6$) and to the lack of a gaseous phase ($P_7$). (Khalidi, 2013, p. 173-174)

For Khalidi's proposal to be applicable to kinds of all scientific domains, it is perhaps best to think of it as providing a template. On this view, the crucial point is that kinds support many projectible generalizations when a broad set of properties are causally related in such a way that when some of them are (co-)instantiated, others will tend be instantiated as well. In many cases, this means these causal relations will have a common-cause structure, as in the example of *polymers* above (Khalidi, 2013). To account for the projectibility of a particular kind, one should aim to provide a causal model of the properties associated with that kind. This causal model will depict the metaphysical ground of the kind's projectibility.

Surprisingly, perhaps, Khalidi does not advocate a particular view about the nature of these causal relations (2013, p. 174). He approvingly cites interventionist theories like that of Woodward (2003) but also says that the arrows in his model represent (types of) causal processes (Khalidi 2015, p. 1387). This too is so that the account would be as widely applicable as possible.

However, by focusing on the *projectibility* of kinds as the explanandum of one's metaphysical theory of kind, the necessity of providing a causal theory becomes questionable.[7] In the following section, I lay out the challenge posed by cluster theorists, who argue that a causal theory is unnecessary to account for a kind's projectibility. In what follows afterward, I will then argue that this is true for a kind's projectibility but not for the explanatory and taxonomic uses of kinds. Khalidi's idea that kinds are causally structured can be developed further, allowing us to determine the ways in which the structure of the causal relations between the properties associated with a kind determines the kind's epistemic fruitfulness.

**2. Cluster Theories**

Several philosophers have defended cluster theories of natural kinds (Chakravartty; 2007; Häggqvist, 2005; Slater, 2015). They challenge the causal ground hypothesis by maintaining that a metaphysical theory of natural kinds can and should account for their epistemic fruitfulness in non-causal terms. Slater's recent version of this challenge has been particularly well-received (Slater, 2015).[8] The first step in his argument, according to my reconstruction of it, is to divide the hypothesis defended by causal theorists into two metaphysical sub-hypotheses.

META-1: For a kind to be epistemically fruitful, it is necessary that its members share non-accidental similarities.

META-2: For the members of a kind to share non-accidental similarities, it is necessary that the kind is causally grounded.

Although Slater also aims to cast doubt on META-2, the essence of his argument is that by fleshing out the particular type of non-accidental similarity that is referred to in META-1, a theory of natural kinds is able to fully account for their epistemic fruitfulness.[9] He believes "an account of natural kinds would do better to focus on the special sort of stability a cluster of properties might possess in virtue of which it is apt for induction and explanation rather than focusing on the *something* causing that stability" (Slater, 2015, p. 396).

On Slater's view, the aptness of kinds to support our epistemic projects is grounded in the fact that kinds are cliquish clusters of properties that would remain clustered even given some relevant counterfactual circumstances. Two ideas shape this theory; cliquish clustering and counterfactual stability. A cluster of properties forms a *clique* when we are likely to find all the properties of the cluster when we find some of them. To get this point across, Slater asks us to imagine a clique of friends at the mall, consisting of the individuals Peg, Quinn, Ralph, Sarah and Tim [P, Q, R, S, T].

Spotting Peg, Quinn, and Ralph at the mall means that Sarah and Tim are probably there as well. Nothing is implied about how long they'll stay. Perhaps they flit from place to place, but when a few of them are around, you can bet that the others will be as well. Call this conception of stability cliquish stability. (Slater, 2015, p. 397)

More formally, Slater uses the predicate [Φ] as the equivalent of a cluster of properties [P, Q, R, S, T]. He introduces the notion of sub-clusters, like [Φ$_1$] and [Φ$_2$], that include some but not all of the properties in [Φ], and a probabilistic entailment relation represented as '$\Rightarrow$'. Given this, Slater (2015, p. 398) maintains that "a property cluster [Φ] is cliquishly-stable when for all x and for many sub-clusters [Φ$_1$], [Φ$_2$], [Φ$_3$]…" it is true that:

■ (([Φ$_1$]x$\Rightarrow$[Φ]x) $\land$ ([Φ$_2$]x$\Rightarrow$[Φ]x) $\land$ ([Φ$_3$]x$\Rightarrow$[Φ]x) $\land$ …)

The robustness box indicates that this clustering is to be counterfactually stable under the relevant set of counterfactual suppositions. The details of this notion of stability will not concern us here, but generally, which counterfactuals are relevant is determined by the inductive and explanatory projects of a particular scientific discipline.

Given this cluster theory of natural kinds, Slater contends that an additional causal theory would not contribute to our understanding of the ground of a kind's epistemic fruitfulness. That is not to say it might not be interesting in its own right to explain why it is that certain properties tend to cluster but this additional explanation would not contribute to our understanding of *what it is that makes kinds epistemically useful*. Slater summarizes his view, writing that:

[W]hat makes natural kinds so epistemically useful is the fact that the properties that characterize a kind are stable in the sense that they would remain clustered together no matter what was (or could be) the case. In some cases, essences may secure this stability; in other cases, perhaps homeostatic mechanisms of a certain sort are up to the challenge. Perhaps stability can be multiply-realized or simply a brute fact about the world. The SPC account places no requirements

on the causes of stability: just that kinds are, in fact, stable in the cliquish sense. (Slater, 2013, p. 148)

Hence, on Slater's proposal, natural kinds are associated with cliquish clusters of properties such that there is a high probability that when a sub-cluster of properties is instantiated, the whole cluster will be instantiated. It is important to note, however, that these "sub-cluster-to-whole-cluster" probabilities are not to be understood causally. In his formalization of the clustering constitutive of kinds, the probabilistic entailment relation ($\Rightarrow$) does not refer to a causally directed relation, such that the instantiation of a sub-cluster would *increase* the probability that the whole cluster is instantiated.[10] Instead, it rather refers to the probability of observing all the other properties associated with the kind upon observing some of them, whatever the possible causal relations between these properties. Hence there is no place in Slater's cluster theory for the notion that the relations between the properties associated with a kind are causally structured.

On Slater's view, then, a causal model of the properties associated with a kind cannot contribute anything to our understanding of a kind's epistemic fruitfulness, beyond what it would say about the probability for these properties to co-occur in a cliquish way and the circumstances in which they will do so. Even though it might also be interesting to know *why* these properties tend to co-occur and why this is counterfactually stable, this additional causal information would not improve our understanding of what it is that makes the kind epistemically fruitful.

The challenge posed by cluster theories like that of Slater to causal theories like that of Khalidi is on point. In fact, *provided that we restrict their scope to the projectibility of natural kinds*, clusters theories are indeed sufficient. The aptness of natural kinds to support many inductive inferences is due to the counterfactually stable clustering of properties, however that is brought about.

Nevertheless, a cluster theory cannot fully account for the epistemic usefulness of kinds. Before proceeding to my argument in favor of this point, recall that cluster theorists like Slater (2015) and

Häggqvist (2005) do not object to the claim that most natural kinds are *de facto* causally grounded in some way or another. Their claim, however, is that one can safely ignore this causal ground when analyzing the metaphysical ground of a kind's epistemic usefulness. In what follows, I will argue that this claim is incorrect. When a kind is causally grounded, there will be many cases where one cannot fully understand that kind's epistemic fruitfulness without analyzing the structure of the causal relations that exist between the properties associated with the kind. Hence the point of this paper is not to argue that all natural kinds must be causally grounded, which would require a different line of argumentation. The scope of the argument that follows is restricted to those kinds that everyone would agree are causally grounded in some way, simply calling these 'causal kinds.' The claim that I defend is that a metaphysical theory of natural kinds cannot fully explain the epistemic usefulness of causal kinds without also taking their causal structure into account. In many cases, the nature of this causal structure says something about the epistemic usefulness of the kind that cannot be accounted for in terms of the cliquish clustering of properties.

## 3. Kind-based Explanation

If cluster theories are correct, we should not only be able to account for the projectibility of causal kinds but also for their explanatory usefulness in terms of the clustering of their properties. On Slater's cluster view, for instance, it ought to be the case that natural kinds are apt to support kind-based explanations due to the cliquish and stable clustering of the properties associated with the kind. In this section, I argue that the explanatory use of kinds is instead dependent on the structure of the causal relations between the properties associated with the kind.

To do so, I will assume that in many cases, scientific explanations are causal explanations. On such a view, a successful scientific explanation cites the causal difference-makers of the explanandum (Strevens, 2008). Since it is widely accepted that some version of this causal view on scientific explanation is correct,

it will not be further supported here, nor will it be described in any more detail. Instead, I proceed by describing the role of kinds in scientific – causal – explanations, after which I will argue that cluster theories cannot account for this explanatory role of kinds.

**3.1 The Explanatory Role of Kinds**

Scientific explanations for the instantiation of a property by a particular object often cite what *kind* of object it is. Consider the following examples: [11]

(1) Q: Why can this material object not be transformed into a gas?
    A: Because it is a polymer.

(2) Q: How come this animal has stripes?
    A: Because it is a tiger.

(3) Q: Why does he have suicidal thoughts?
    A: Because he suffers from Major Depressive Disorder (MDD).

Needless to say, these answers do not provide full scientific explanations of the explananda. They require additional information about the causal difference-makers, processes, or mechanisms involved in bringing about the explananda, depending on which version of a causal account of explanation one subscribes to. Nevertheless, even when they do not constitute complete explanations, these kind-predications do provide explanatorily relevant information. As Khalidi has it, kind predicates typically denote a set of causal properties that are responsible for many of the other properties associated with the kind. By predicating kind-membership of an individual, one predicates their instantiation of that set of causal properties. This is explanatorily relevant when the explanandum is an effect of one of those properties. A complete explanans pairs this kind-predication with a causal model of the kind that allows one to specify which of the denoted causal properties is the actual difference-maker (and perhaps further information about the process or mechanism involved.)

Take the first example of a kind-based explanation. The role of the kind predicate 'polymer' here is to denote the defining core property of polymers. The core property of polymers is to consist of a long molecular chain of repeating monomers, which also causes polymers not to have a gaseous phase. Since the phenomenon to be explained is the lack of a gaseous phase by a particular object, determining that the object is a polymer in conjunction with the causal model depicted in *figure 1* above, suffices to specify the cause of the explanandum. Further understanding is perhaps gained by knowing the mechanism by which consisting of a long molecular chain of monomers causes polymers to lack a gaseous phase, namely that it makes for such a high molecular weight that the temperature required for them to evaporate is higher than the temperature at which they disintegrate. But even without this further information, the kind-predication and causal model of the kind suffice to specify the cause of the explanandum.

The second example is more complex since the kind predicate 'tiger' cannot be sufficient to denote the property that is (proximately) causally responsible for the stripes of a particular tiger. Whatever property causes most tigers to have stripes, it is not a necessary property of tigers. There are after all also albino tigers. In a case like this, the kind predicate can only play its role of denoting the causally relevant property if in addition, one also provides what Strevens has called a *basing generalization* (Strevens, 2008, 2012). A basing generalization states a lawful association between a kind and the property that is causally responsible for the explanandum under consideration. In addition to identifying a particular animal as a tiger, for example, one would have to provide a basing generalization according to which '*Tigers tend to have property P'* (*P* being whatever property is causally responsible for the development of their fur pattern). In combination with the information that *P* causes the development of a striped fur pattern, this kind-predication and basing generalization allow one to specify the causal difference-maker of the explanandum.

In both examples (1) and (2), citing what kind of thing something is allows one to specify the property that is causally responsible for the explanandum, provided one also has further theoretical information

about the causal relations associated with the kind. In the following section, I will argue that not all kinds are equally apt to support this type of explanatory practice and that nothing in a cluster theory allows one to account for this fact.

**3.2 The Causal Homogeneity Requirement for Kind-based Explanation**

Not all kinds support kind-based causal explanations equally well, independently of how well the properties associated with the kind cluster. The role of a kind predicate in a causal explanation is to denote a set of causal properties. This supports a successful causal explanation only when the structure of the causal relations associated with the kind is such that only one property (or combination of properties) in that set could be causally responsible for the explanandum. In such a case, a causal model of the kind specifies the cause of the explanandum. When, on the other hand, the structure of the causal relations associated with a kind is such that several properties (or combinations of properties) in the denoted set could be causally responsible for the explanandum, kind-membership as such is not explanatory. A causal model of the kind would in that case not specify the cause of the explanandum. This is the *causal homogeneity requirement* for kind-based explanation; kind-membership is explanatorily relevant to the instantiation of one of the properties associated with the kind, only if there is no more than one cause of that property associated with the kind. This is a requirement that a cluster view of kinds cannot account for.

Recall Khalidi's depiction of the causal structure associated with the kind *polymer* (Figure 1). According to that causal model, each derivative property associated with polymers only has one possible cause associated with that kind, which is ultimately the instantiation of the defining property of the kind itself (although there will be many contributing background conditions). The kind *polymer* therefore satisfies the causal homogeneity requirement for each of its derivative properties. If one of these derivative properties is instantiated by an object, it suffices to determine that the object is a polymer to specify the

cause of the instantiation of that property. When, for example, an object cannot be transformed into a gas ($P_7$ according to *figure 1*) the causal structure of the kind is such that only one of its associated properties could be causally responsible ($P_1$). Even if one were to hold that the predicate 'polymer' denotes not only the defining property of the kind but rather all the properties associated with the kind, there would still only be one property in the denoted set that could be causally responsible for the explanandum. It is because the structure of the causal relations associated with the kind *polymer* satisfies the causal homogeneity requirement, that predicating kind-membership is explanatory.

The same holds true in the case of tigers. In conjunction with a basing generalization, identifying an animal as being a tiger is sufficient to specify the property (or more likely combination of properties) that causes the striped fur pattern of that animal. That is because there is only one property (or combination of properties) associated with tigers that is responsible for their striped fur. If, on the other hand, it would have been the case that the stripes of male tigers and those of female tigers are the result of a distinct causal mechanism, then determining that a particular animal is a tiger would not have sufficed to specify the mechanism responsible for its striped fur pattern. In that case, one would rather have to identify a particular animal as being either a 'male tiger' or a 'female tiger,' and hence the predicate 'tiger' would not support kind-based explanations of fur pattern.

Importantly, whether a kind satisfies this causal homogeneity requirement for any or all of its properties is independent of the extent to which the properties associated with the kind cluster. Consider the third example of a kind-based explanation provided above. In order to explain why a patient has suicidal thoughts, one might want to diagnose the mental disorder that patient is suffering from. But many philosophers of psychiatry would object to such an explanatory use of a kind diagnosis in psychiatry (Maung, 2016). That is because most of the disorders currently listed by the Diagnostic and Statistical Manual of Mental Disorders (APA, 2013) do not satisfy the causal homogeneity requirement. For many of the symptoms associated with a kind of mental disorder, there are several different (sufficient) causes

associated with that same disorder. Mental disorders should not be thought of as common-cause networks – such as in the case of polymers – but rather as complex causal structures that are partly disjunctive in nature. Consider the following example. When reflecting on the causal structure associated with some of the observable symptoms of Major Depressive Disorder (MDD), Kendler, Zacher and Craver (2011) provide the following causal model:

Figure 2:
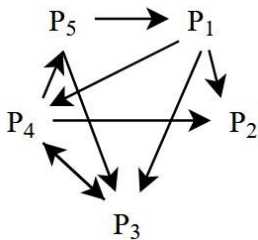


According to this model, the symptoms associated with MDD are causally structured in such a way that suicidal ideation ($P_2$) can be caused both by having a depressed mood ($P_1$) or by feelings of guilt ($P_4$). Both are causes that – together with several background conditions in place – lead to the development of suicidal thoughts. If this is a correct causal model of some of the properties associated with MDD, one cannot provide an explanation of a patient's suicidal thoughts by diagnosing their disorder as MDD, even if one knows everything there is to know about the causal relations associated with this disorder. Identifying the kind of disorder the patient is suffering from would be compatible with two different possible causes of their suicidal thoughts. To explain someone's suicidal thoughts ($P_2$) by diagnosing them with MDD is to say that these thoughts are caused "either by their depressed mood ($P_1$) or by their feelings of guilt ($P_4$)." This disjunction does not specify the causal difference-maker of the explanandum.

Note, however, that this causal structure modeled in figure 2 would not support kind-based explanations *even if it would result in a very stable cluster of properties*. For a kind to be explanatorily fruitful (when considering causal explanations) requires the relations between its properties to have a particular causal structure; there can only be one sufficient (set of) cause(s) associated with the kind for

each derivative property associated with it. Even if all patients who suffer from MDD developed the same symptoms, one would still not be able to causally explain the instantiation of a particular patient's symptoms by diagnosing that patient as suffering from MDD. The fact that psychiatric categories like MDD do not satisfy the causal homogeneity requirement is one reason why these kind categories are not as epistemically useful as other typical natural kinds, a fact that cannot be accounted for by a cluster theory. Kind predicates that support kind-based explanations do so by virtue of satisfying the causal homogeneity requirement, thereby allowing one to specify the causal property responsible for the explanandum. Kind predicates that do not support kind-based explanations fail to satisfy the causal homogeneity requirement. In both cases, however, the properties could be equally clustered and support equally many projectible generalizations.

This argument against cluster theories crucially relies on the view that a kind's projectibility and a kind's causal-explanatory usefulness are grounded in two different features of the world. The point is not just to say, as has been done before, that without a causal-explanatory theory of natural kinds one cannot explain why something like *jade* does not constitute a natural kind given that its members do have some properties in common. It is now common knowledge that 'jade' refers to two different causal-explanatory kinds; *jadeite* and *nephrite*.[12] Members of both kinds have a different microstructural composition. Due to this different composition, however, they also have many different chemical properties. This means that a cluster theorist could very well argue that one does not need a causal-explanatory theory to explain why there really are two distinct kinds. 'Jade' simply does not refer to a chemically interesting cluster of properties. Importantly, a similar response is not possible in the case of kinds with (partly) disjunctive causal structures, like the example of MDD before. The point is that *however many* properties are projectible of such a kind, if there are different causes of those properties associated with that kind, it is not useful to refer to it in a scientific explanation of those properties.

This constitutes a first argument against the sufficiency of cluster theories of natural kinds. One cannot understand the usefulness of natural kinds in causal explanations without taking their causal structure into account. Like Khalidi, I have not committed to a view about the nature of the causal relations that exist between the properties of kinds. These relations can be whatever one's causal theory about scientific explanation requires. For kinds to support kind-based explanations on any such theory of causal explanation, however, it is crucial that these causal relations have a specific structure. Only when the causal homogeneity requirement is satisfied can kind-predications contribute to scientific explanations by specifying the property (or set of properties) causally responsible for the explanandum.

## 4. Kind-based Categorization

Most philosophical theories about natural kinds that aim to account for their epistemic fruitfulness focus on their projectibility, sometimes also on their explanatory usefulness. Much less has been said about the *categorizing* use of kinds, namely the practice of assigning kind-membership to individuals. Nevertheless, this is perhaps the most fundamental operation that one performs based on one's knowledge about a kind. Before kinds can be used for inductive and explanatory practices, one has to correctly determine which individuals belong to it.

In the previous sections I have discussed the projectibility and explanatory usefulness of natural kinds from a perspective according to which these are two independent dimensions of a kind's epistemic usefulness and according to which some natural kinds are more useful in these regards than others. A third dimension of a kind's epistemic usefulness that I will add here is their *categorizability*, namely the extent to which kinds allow for the correct identification of their members. Some kinds are associated with many properties and many combinations of properties that allow one to correctly identify members of the kind, whereas other kinds have fewer diagnostic (combinations of) properties. This is one dimension of the epistemic usefulness of natural kinds.

As we have done in the previous sections for the projectibility and explanatory usefulness of kinds, we can also ask what it is about natural kinds that makes them apt to allow for accurate categorization decisions. It is clear that what allows us to correctly assign particulars to their respective kinds is not identical to what allows us to project properties from observed to unobserved members of these kinds. Very roughly, projectibility is rooted in the similarities of kind-members whereas categorizability also depends on the extent to which kind-members differ from non-members.

How exactly one should understand the metaphysical ground of a kind's categorizability is the subject of this section. If cluster theories are correct and complete, the clustering of properties would not only ground the projectibility of natural kinds but also their categorizability. In what follows, I provide both a metaphysical (4.1) and an epistemic (4.2) argument for the view that a cluster theory cannot suffice to account for a causal kind's categorizability.

**4.1 Categorizing by Inferring Causes**

According to cluster theorists like Slater, the epistemic fruitfulness of kinds is grounded in the cliquish clustering of properties: if one finds some properties associated with the kind, one should also be likely to find *all the other properties* of the kind. This view, however, is too broad to account for a causal kind's categorizability. Membership of a causal kind is typically determined based on whether an individual instantiates the core causal properties of that kind, not by whether it instantiates the whole cluster of properties.[13] Hence the objective when categorizing individuals into causal kinds is to correctly estimate whether an individual instantiates the defining core properties of a particular kind. One can therefore say that a causal kind has good categorizability if for many of its associated properties (and combinations of those properties) their instantiation allows one to infer that the defining *core properties* of that kind will then be instantiated as well.

Think, for instance, of the diagnosis (categorization) of diseases. Often when the causes of the symptoms associated with a disease are known, the disease is defined based on these core causal properties. Diagnosing an illness is then a matter of observing some symptoms of a patient and estimating the probability that they were caused by the core properties associated with a particular illness. Hence a disease kind has good categorizability if many of its symptoms and combinations of symptoms are such that when they are observed, they allow a diagnostician to infer that the core causal properties of that particular disease must be responsible. What does *not* determine the categorizability of a disease kind is whether many symptoms and combinations of symptoms (i.e. sub-clusters of properties) allow a diagnostician to infer that *all the other properties* (i.e. the whole cluster) associated with that disease will be present as well, including all the other possible symptoms.[14] The categorizability of causal kinds is determined by the probability of finding the core causal properties of the kind upon observing some other properties associated with the kind, not the probability of then also finding all the other properties associated with the kind.

Let me use some of Slater's vocabulary to make this point more precise and to draw out the difference between a causal and a cluster theory. Take the predicate [$\Psi$] as the equivalent of the complete set of properties associated with a kind. For Khalidi's example of the kind *polymer*, [$\Psi$] consists of [$P_1$, $P_2$, $P_3$, $P_4$, $P_5$, $P_6$, $P_7$]. Take [$\Psi_1$], [$\Psi_2$], [$\Psi_3$] and so on, to consist of a (possible singleton) subset of the properties in [$\Psi$]. Furthermore, take [$\Psi_{CC}$] to consist of the core causal properties that ultimately determine whether an individual is a member of the kind or not. In Khalidi's example, [$\Psi_{CC}$] simply consists of [$P_1$]. Given this, the extent to which a causal kind supports the success of our categorization decisions is the extent to which for many x and many [$\Psi_1$], [$\Psi_2$], [$\Psi_3$], … it is the case that:

■ (([$\Psi_1$]x$\Rightarrow$[$\Psi_{CC}$]x) $\wedge$ ([$\Psi_2$]x$\Rightarrow$[$\Psi_{CC}$]x) $\wedge$ ([$\Psi_3$]x$\Rightarrow$[$\Psi_{CC}$]x) $\wedge$ …)

What matters when categorizing causally grounded kinds is the probability that one will find the set of core causal properties of the kind when observing some set of causally secondary ones. Note the

difference with the previous argument about the explanatory usefulness of kinds. In that case, the point was that kinds need to have a particular causal structure, – one that satisfies the causal homogeneity requirement –, in order for them to be relevant for causal explanations. In this case, the point is that it depends on the causal structure of the kind *which* probabilities ought to be high for a kind to support accurate categorization decisions. A causal kind is apt to support categorization decisions when there is a high probability that the kind's core causal properties are instantiated when some combination of its derivative properties is instantiated. The upshot of this argument is that a cluster theory of natural kinds, without a notion of their associated causal structure, cannot explain what it is about kinds that allows for the accurate categorization of their members.

**4.2 Categorizing Based on Causal Models**

In this section, I add an epistemological argument to the metaphysical one provided above. When categorizing a causal kind, a reasoner often relies on information provided by a causal model of that kind to accurately categorize its members. The causal information that is relied on, however, cannot be accounted for in terms of the cliquish clustering of properties. Hence a cluster theory about a kind does not represent the type of theory a reasoner relies on when categorizing members of the kind.

Consider a kind with four associated properties that are linearly structured such that $P_1$ causes $P_2$, which causes $P_3$, which in turn causes $P_4$. There is one core property ($P_1$) that ultimately determines the kind-membership of individuals, the other three are derivative ones. Now imagine observing that an individual instantiates $P_3$ and $P_4$, but also observing that it does *not* instantiate $P_2$. The case is represented in figure 3, both from the perspective of a causal theory and from the perspective of a cluster theory.

Figure 3:

*Causal theory*                                                    *Cluster theory*
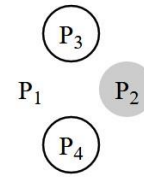


Since P$_1$ is the defining property of the kind, the goal of categorization is to determine the probability that P$_1$ will be instantiated, given the observed set of present and absent properties. How does one estimate this probability? One crucial piece of information is an understanding of the causal structure of the kind. In the example above, the precise impact that the absence of P$_2$ should have on one's confidence that P$_1$ is also instantiated depends on several probabilities that are not provided in these rudimentary models. What one does know based on the causal model, however, is that one should weigh the observation *of P$_3$ and P$_4$* less heavily than if one would have also observed P$_2$ to be present, given that this property is the cause of P$_3$ and P$_4$ according to the kind's causal model.

In fact, this is how all of us categorize individuals into kinds when we have causal information about them. In cognitive psychology this is known as the *coherence effect*. People categorize individuals by judging the coherence of the causal structure of a kind with the observed set of present *and absent* properties of that individual (Rehder, 2003).[15] That is, people categorize by asking themselves the question: 'Given what I know about the causal structure associated with this kind, how likely is it that the instantiation of that causal structure would have generated the combination of present and absent properties that I observe?' The information provided by the causal model of a kind is therefore crucial in order to accurately estimate the probability of kind-membership.

One might wonder, however, to what extent this information provided by the causal model of a kind cannot be described in terms of the probability of properties to cluster just as well. Could we not just take the instantiation of P$_3$ and P$_4$ combined with the *absence* of P$_2$ as one possible sub-cluster, such that the

probability of finding $P_1$ given this sub-cluster should be high? After all, cluster theorists would not object to the claim that knowing the causal ground of a kind can be a useful way of coming to know the probabilities of properties to cluster and the counterfactual circumstances in which this clustering would hold.

In this case such a response will not do. Knowing the causal structure of a kind is particularly helpful for one's categorization decision if one observes that some properties associated with the kind are *not* instantiated, as in the example above. There is, however, no place for a notion of sub-clusters that include the *absence* of properties in a cluster theory. If one were to adopt such a notion as part of Slater's theory, for example, it would mean that for any combination of present and absent properties, one should be likely to find the other properties associated with that kind. But that would be anathema to the notion of a cliquish cluster, since then the whole cluster of properties would not tend to co-occur anymore.

The fundamental point is that a causal model of a kind provides information that is not just about the probability of properties to co-occur. As a result, a cluster theory of kinds does not suffice to account for the success of our kind-based categorization decisions. When individuals are categorized as members of causal kinds, the accuracy of these categorization decisions depends partly on an understanding of the causal structure of these kinds. This is a third reason to adopt a causal theory of kinds. Only by including information about their causal structure does a theory of kinds also represent the knowledge that allows for accurate categorization decisions.


**Conclusion**

According to causal theories of natural kinds, the epistemic fruitfulness of natural kinds can only be accounted for in causal terms. Cluster theorists like Slater, however, have successfully argued that a causal theory is unnecessary to account for the projectibility of kinds, since a kind's projectibility is ultimately grounded in the stable and cliquish clustering of properties.

However, I have argued that cluster theories only provide a partial analysis of the metaphysical ground of a natural kind's epistemic usefulness. In the case of causal kinds, their explanatory and categorizing usefulness cannot be completely accounted for in terms of the clustering of properties. For a causal kind to support a kind-based explanation requires that the homogeneity requirement is satisfied for the explanandum under consideration. Furthermore, correctly categorizing the members of a causal kind does not require the clustering of all the properties associated with the kind, but rather a high probability that the kind's core causal properties are instantiated when some combination of derivative properties is instantiated. And when one categorizes members of a causal kind, it is often one's knowledge of their causal structure that allows one to make the correct decision about kind-membership.

In arguing that one should analyze the causal structure of a kind to understand its epistemic usefulness, I have aimed to show that there are different independent metaphysical dimensions to a kind's usefulness. We will move forward in the metaphysical debate about natural kinds when we no longer aim to determine a single set of necessary and sufficient metaphysical criteria of kinds to explain their fruitfulness, but rather analyze the various ways in which kinds can be used and the metaphysical success conditions of those uses.

---

[1] The goal of a metaphysical theory of natural kinds is not to *justify* kind-based inductive inferences nor to describe the rules of such inferences, but rather to say what it is about the metaphysical nature of kinds that makes them apt to support successful ones. See Beebee (2011) for why a *justification* of induction based on natural kinds would fail. I subscribe to this methodological view that a metaphysical theory of natural kinds ought to be able to explain their epistemic usefulness. This excludes from consideration here those theories of natural kinds that are explicit about not wanting to provide a metaphysical account of their epistemic usefulness (e.g. Franklin-Hall, 2015)

[2] Throughout this text, I will speak of natural kinds being projectible and supporting kind-based projections. This is a shorthand. More precise would be to say that it is the kind terms or predicates that are projectible. The general notion of projectibility I use here is based on Khalidi's characterization of it, when he states that: "predicate *P* is projectible relative to predicate *Q* if and only if we can draw a legitimate inductive inference from *x* is *P* to *x* is *Q*, where the predicates stand for properties or natural kinds and *x* denotes a particular, whether a concrete individual, a specific event, a particular process, and so on" (Khalidi, 2015, p. 2).

[3] A terminological note: the phrase 'cluster theory' can be used in two different ways. When used in contrast to causal theories, the term refers to *non-causal* theories of natural kinds according to which the properties associated with a natural kind simply tend to co-occur. This is the intended sense of the phrase throughout this paper. However, in the literature the phrase is also used in contrast to essentialist theories, and in that case refers to causal theories like those of Boyd and Kornblith according to which kinds do not have essences but are rather causally grounded clusters of properties. This is *not* the intended sense of the phrase here.

[4] As I describe them here, these causal theories share a view about the ground of the epistemic usefulness of kinds, not about their *naturalness*. After all, whatever it is that makes a kind natural need not be the same thing as what makes a kind epistemically useful. But both views often go hand in hand. If, like Boyd for instance, one believes that the naturalness of a kind lies in the accommodation of the epistemic demands of a scientific discipline to the world, then it follows from the causal view on a kind's epistemic fruitfulness that natural kinds "must correspond in relevant respects to the (perhaps unobservable) properties and mechanisms which causally determine the observable properties of the subjects of empirical generalizations" (Boyd, 1980, p. 642). I will not discuss the naturalness of kinds any further in

the rest of this paper, restricting myself to providing an account of the epistemic usefulness of those kinds that are typically considered to be 'natural' ones as well.

[5] The name of this theory and the particular way of loosening up Boyd's *mechanistic* causal view, was first suggested by Craver (2009).

[6] Khalidi's theory is not an essentialist view. The core properties of a kind can be extrinsic or functional, and because these core properties can also constitute a loose cluster, they need not constitute a necessary and sufficient set of properties for kind-membership.

[7] Khalidi also believes that other epistemic uses of kinds depend on their causal ground, saying that: "What enables natural kind categories to play the role that they do in our inductive, explanatory, and taxonomic practices is that they consist of highly connected nodes in causal networks" (Khalidi 2015, 1387). However, how exactly the causal connections between the properties associated with a kind contribute to a kind's explanatory and taxonomic usefulness is not further developed. This is what I am to do in this paper and in doing so provide an argument against the sufficiency of cluster theories.

[8] Slater's paper *Natural Kindness* (2015), in which he defends his metaphysical view of natural kinds, was the recipient of the 2015 BJPS Popper Prize.

[9] Slater casts doubt on META-2 by arguing that there are clusters of properties that support inductive inferences yet that are not causally grounded. The example he discusses are sibling species, namely two species that are very similar to each other but that do not share a biological (interbreeding) mechanism maintaining that similarity (2015, p. 392). I remain agnostic here about the question whether these kinds would constitute proper examples of natural kinds that are not causally grounded. In this paper, I do not aim to argue that natural kinds must be causally grounded but rather aim to argue that two ways in which (many) natural kinds are epistemically useful cannot be accounted for in terms of the clustering of properties.

[10] Slater's description of cliquish clusters does not require that the instantiation of a sub-cluster increases the *ontic* probability that the whole cluster is instantiated. This is compatible, of course, with the requirement that the epistemic probability that we will *observe* the whole cluster upon observing a sub-cluster does increase.

[11] There is also a very different use of kind predicates that *can* be accounted for by cluster theories, in contrast to the use of kind predicates that I am outlining here which cluster theories cannot explain. That other use of kind predicates is when one refers to kinds in explanations because knowledge of their properties allows one to causally explain a phenomenon that is itself *not* typical of the kind. For instance, when explaining why it is that the size of the panda population has declined over the years (although it is now recovering), one would refer both to the ongoing deforestation process and to several projectible properties of pandas, like their diet – they really need a lot of bamboo –, and their birthrate. That the panda population has declined in size is itself not really a projectible property of pandas, but this particular phenomenon can be causally explained based on other properties that pandas tend to have. Often, this is why one refers to a kind in a causal scientific explanation; a typical property of the kind is causally responsible for the phenomenon one aims to explain but which is itself not typical of the kind. This explanatory use of natural kinds can be accounted for by cluster theories as well, since what makes natural kinds apt to be used in this particular way is just that their members tend to have the same properties, namely those that are causally responsible for the explanandum.

[12] "Although the Chinese do not recognise a difference, the term 'jade' applies to two minerals: jadeite and nephrite. Chemically, there is a marked difference. Jadeite is a combination of sodium and aluminium. Nephrite is made of calcium, magnesium and iron. These two quite different microstructures produce the same unique textural qualities!" (Putnam, 19751, p. 24)

[13] That the causal core properties of kinds also constitute their identification condition is epistemically justified. If kind-categorization is to be useful for predictive and explanatory purposes, it is the instantiation of the core properties that is to be decisive given that it is these core properties that are causally responsible for the other properties associated with the kind.

[14] That is not to say that the observation of more symptoms would not increase one's confidence in a diagnosis but rather that when diagnosing an illness, one does not aim to infer the presence of these other symptoms. Rather, diagnosis is a matter of inferring the presence of the causal core properties responsible for the observed symptoms.

[15] Rehder (2003) provides a formalization of the probabilistic causal model that people rely on when making these categorization judgments. His results show that, in many cases, people have a specific model in mind when making categorization decisions, and that they for the most part obey the rules of probability theory when reasoning about these models.

**References**

American Psychiatric Association (APA) (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Pub.

Beebee, H. (2011). Necessary connections and the problem of induction. *Noûs*, *45*(3), 504-527.

Boyd, R. (1980). Scientific realism and naturalistic epistemology. In *PSA: Proceedings of the biennial meeting of the Philosophy of Science Association* (pp. 613-662). Chicago: Philosophy of Science Association

Boyd, R. (1991). Realism, Anti-foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies, 61*, 127-148.

Boyd, R. (1999a). Kinds, Complexity and Multiple Realization. *Philosophical Studies, 95*, 67-98.

Boyd, R. (1999b). Homeostasis, Species, and Higher Taxa. In Wilson, R.A. (Ed.), *Species: New Interdisciplinary Essays*, (pp. 141-185). Cambridge, MA: MIT Press.

Boyd, R. (2000). Kinds as the "Workmanship of Men": Realism, Constructivism, and Natural Kinds. In Nida-Rümelin, J. (Ed.), *Rationalität, Realismus, Revision: Vorträge des 3. Internationalen Kongresses der Gesellschaft für Analytische Philosophie*, (pp. 52-89). Berlin: De Gruyter.

Chakravartty, A. (2007). *A metaphysics for scientific realism: Knowing the unobservable*. Cambridge, MA: Cambridge University Press.

Churchland, P. M. (1985). Conceptual progress and word/world relations: In search of the essence of natural kinds. *Canadian Journal of Philosophy*, *15*(1), 1-17.

Craver, C. F. (2009). Mechanisms and natural kinds. *Philosophical Psychology*, *22*(5), 575-594.

Ellis, B. (2001). *Scientific Essentialism*. Cambridge, MA: Cambridge University Press.

Franklin-Hall, L. R. (2015). Natural kinds as categorical bottlenecks. *Philosophical Studies*, *172*(4), 925-948.

Goodman, N. (1954/1979), *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Häggqvist, S. (2005). Kinds, projectibility and explanation. *Croatian journal of philosophy, 13,* 71-87.

Hempel, C. (1965). *Aspects of Scientific Explanation*, New York, Free Press.

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders?. *Psychological medicine*, *41*(6), 1143-1150.

Khalidi, M.A. (2013). *Natural Categories and Human Kinds. Classification in the Natural and Social Sciences*. New York: Cambridge UP.

Khalidi, M.A. (2015) Natural Kinds as Nodes in Causal Networks. *Synthese*, 1-18.

Kornblith, H. (1993). *Inductive Inference and its Natural Ground*. Cambridge, MA: MIT Press.

Maung, H. H. (2016). Diagnosis and causal explanation in psychiatry. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *60*, 15-24.

Millikan, R.G.  (1999). Historical kinds and the 'special sciences'. *Philosophical Studies*, *95*, 45–65.

Putnam, H. (1975). "The Meaning of 'Meaning'", *Minnesota Studies in the Philosophy of Science*, 7: 215–271.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1141.

Slater, M. H. (2013). *Are Species Real? An essay on the metaphysics of species*. London: Palgrave-Macmillan.

Slater, M.H. (2015). Natural kindness. *The British Journal for the Philosophy of Science*, *66*(2), 375-411.

Strevens, M. (2008). *Depth: An account of scientific explanation*. Harvard University Press.

Strevens, M. (2012). The Explanatory Role of Irreducible Properties. *Noûs*, *46*(4), 754-780.

Wilkerson, T.E. (1988). Natural Kinds. *Philosophy*, *63*, 19-42.

Wilson, R. A., Barker, M. J., & Brigandt, I. (2007). When traditional essentialism fails: biological natural kinds. *Philosophical Topics*, *35*(1/2), 189-215.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.