Natural kinds, mind-independence, and unification principles

Tuomas E. Tahko[1] (ORCID)

Abstract

There have been many attempts to determine what makes a natural kind real, chief among them is the criterion according to which natural kinds must be *mind-independent*. But it is difficult to specify this criterion: many supposed natural kinds have an element of mind-*dependence.* I will argue that the mind-independence criterion is nevertheless a good one, if correctly understood: the mind-independence criterion concerns the *unification principles* for natural kinds. Unification principles determine how natural kinds unify their properties, and only those natural kinds that have a mind-independent unification principle should be considered real.

Keywords Natural kinds · Mind-independence · Unification · Properties · Realism · Classification

1 Introduction

What are natural kinds? I favour a realist answer, according to which natural kinds reflect natural divisions in mind-independent reality. Accordingly, I am interested in giving a metaphysical rather than an epistemic account of natural kinds.[1] Metaphysical accounts aim to address what we refer to when using natural kind terms, that is, what kind of entities natural kinds are, if they are entities at all. Epistemic accounts typically focus on the usage of kind terms in scientific practice, e.g., the epistemic fruitfulness of postulating something as a kind. This distinction between metaphysical and epistemic accounts is not fully exhaustive and most accounts encompass both elements, but here I am interested in the metaphysical question of what it means for a natural kind to be a real entity. We can also ask the further question: *what* type of real entities natural

---

[1]On the metaphysical/epistemic distinction, see Lemeire (2021). For a discussion regarding the epistemic value of kinds and their metaphysical status, see Kendig and Grey (2021), who argue that even epistemology-only accounts (e.g., Slater, 2015), require metaphysical commitments.

✉ Tuomas E. Tahko
    tuomas.tahko@bristol.ac.uk

1   Department of Philosophy, University of Bristol, Cotham House, Cotham Hill, Bristol BS6 6JL, UK

kinds are. Possible answers to this question include that natural kinds are universals, reducible to bundles of properties, or *sui generis* entities. I regard natural kinds as *substantial universals*, but here I focus on the former, more general question about the 'objectivity' of natural kinds, namely, the conditions for a natural kind to be real or genuine in a metaphysical sense.[2] A popular answer to this more general question is that natural kinds are mind-independent. This mind-independence criterion can take various forms, but a key part of the criterion is that a *natural* or *real* (I use these terms interchangeably) kind corresponds to some grouping or clustering of properties that reflects a worldly structure (e.g., a bundle of properties or a universal) independent of the interests and actions of human beings (see Bird & Tobin, 2018; Chakravartty, 2007; Devitt, 2005; Lowe, 2006; Psillos, 2002; Tahko, 2015). However, the mind-independence criterion has recently come under scrutiny, given that it seems to rule out many potentially real natural kinds, such as social and mental/psychological kinds (see Ereshefsky, 2018; Franklin-Hall, 2015; Khalidi, 2015, 2016; Magnus, 2012).

The upshot of the paper is that the core of the mind-independence criterion—that real natural kind classifications should not be based on arbitrary choice—is still correct. But if we wish to leave room for social and mental kinds, to not rule them out from the class of natural kinds by definition, we should revise our understanding of what the mind-independence criterion amounts to. Whether or not a kind is called *natural* is not important here. The main goal is to give a general account of the mind-independence criterion in such a way that it can encompass all *natural* or *real* kinds. Accordingly, while the starting point of the paper is that we certainly need to be able to accommodate traditional examples such as physical and chemical kinds, the question is precisely whether the same criterion, when appropriately revised, could be extended to other potentially real kinds, such as social or mental kinds. The upshot is that it *can*, but we must examine the relevant kinds on a case by case basis to determine whether they are indeed real. To put it simply: I am interested in giving a realist account of natural kinds and the question is whether we can include 'higher-level' kinds such as social or mental kinds among the natural kinds on this realist reading of natural kinds. My proposed answer is that, yes, some social, mental/psychological, and other 'higher-level' kinds may be among the real/natural kinds, insofar as they have an objective *unification principle*.

Accordingly, I suggest that we apply the mind-independence criterion to the *unification principles* for natural kinds rather than kinds (or their members) per se. What are these principles? At the highest level of generality, we may define unification principles as follows:

Unification principle (UP): The narrowest common cause for the clustering of properties in members of natural kinds.[3]

---

[2] On the view that natural kinds are substantial universals, see Lowe (2006, 2015), Keinänen and Tahko (2019) and Hommen (2021). On this view, natural kinds are a type of universal distinct from property universals.

[3] On the use of 'common cause', see Ismael and Schaffer (2020) and Godman et al. (2020). In fact, Ismael and Schaffer use 'common ground', but I wish to remain neutral about the 'grounding' terminology, so will stick to the somewhat more general 'common cause' terminology. However, it should be noted here that 'common cause' should not be regarded as a (purely) causal notion. In Godman et al.'s usage, the 'common causes' are what they call 'super-explanatory' properties: properties that 'explain the many shared features

A UP determines the common cause for the clustering of a natural kind's defining properties, thus integrating the kind's instances, and providing a way to distinguish the kind's instances from instances of other kinds. The specification of *narrowest* common cause is important because a given entity may of course feature in various higher-level classifications as well. For instance, *electrons* may also be classified as *fermions* (particles that follow Fermi–Dirac statistics) and as *charged particles*, but the narrowest common cause for the clustering of an electron's properties is presumably something that is unique to *electrons*. I will operate with the assumption that each kind has a unique UP, but this is an assumption that may have to be revisited, given the possibility of *cross-cutting kinds* (Tobin, 2010).[4] There are several different candidates for unification principles, such as causal mechanisms, laws of nature, and essences.

Depending on one's other ontological commitments, only some of these candidate principles might be acceptable, but the point is that there must be *some* objective, realist source of unification, and there may potentially be more than one such source. I will remain largely neutral about which UPs we should accept, although I will give various potential examples in what follows.

The discovery of a UP, i.e., *how* the properties are unified in instances of kinds, is typically an empirical matter. However, the metaphysical basis of UPs, i.e., *why* properties are systematically clustered together in members of natural kinds, is a deeper metaphysical question and might not be fully answerable empirically. I label these the *why* and the *how* questions, in addition to the earlier *what* question; this paper is concerned with the *how* rather than the *why* or the *what*:

(1) *What* are natural kinds, that is, what type of real entities (e.g., universals, bundles of properties…)?
(2) *Why* are properties systematically clustered together in members of natural kinds?
(3) *How* are the clustered properties unified in instances of kinds?

One possible line of reasoning to answer (1)–(3) (modelled after Lowe, 2006) would be to say that:

(1) Natural kinds are substantial (Aristotelian) universals;
(2) which are characterized by attributes (property universals) and instantiated by individual substances (instances of kinds);
(3) and the clustering of the properties is governed (at least partially) by laws of nature featuring the relevant kind universals.

Footnote 3 continued

of things' (2020: 1), and which are metaphysically necessary to those things. As I will explain in Sect. 2, Godman, Mallozzi, and Papineau do seem to treat this as a causal explanation, but in line with Ismael and Schaffer's usage of 'common ground', we could also interpret this as a form of 'metaphysical explanation' rather than causal explanation.

[4] Common examples of cross-cutting kinds come from biology and chemistry. Consider *mammal*. Humans are mammals, and so are *monotremes* such as the platypus. The platypus is also oviparous, meaning that it produces offspring by laying eggs, like birds. But birds and humans cannot be classified together either as mammals or as oviparous. Accordingly, it may be impossible to find a unique UP for cross-cutting kinds. I believe that such cases should be assessed on a case by case basis: some of them may indeed fail to be real kinds, whereas in some cases we may be able to find a unique UP after all. I also leave open the possibility of non-unique UPs, but will not consider these cases here.

I am partially sympathetic to this overall line of reasoning, but now that we have an idea about the relationship between the *what*, the *why* and the *how*, we can proceed with a more focused discussion of the *how*, which will be largely neutral about the *what* and the *why*. In what follows, I shall give a brief overview of the role of the mind-independence criterion with reference to recent literature (Sect. 2), discuss various examples of potentially mind-*dependent* kinds (Sect. 3), the role of projectibility and successful prediction and induction regarding natural kinds in scientific practice (Sect. 4), and finally detail the central idea behind unification principles via a reconsideration of some key examples (Sect. 5), followed by a conclusion (Sect. 6).

2 The role of the mind-independence criterion

Let's start by contrasting the realist view about natural kinds with *conventionalism*. There is a hint of this idea already in Mill's work, in this much-quoted passage:

> In so far as a natural classification is grounded on real Kinds, its groups are certainly not conventional; it is perfectly true that they do not depend upon an arbitrary choice of the naturalist (Mill, 1843, p. 720).

It is the view that Mill here *opposes*—a type of conventionalism—that I wish to highlight. Bird summarises this aptly:

> I take it that by "arbitrary" Mill means something like "not mandated by the relevant empirical facts about natural similarity and difference". For example, the division of garden plants into weeds and flowers does not match a natural difference. (2018, p. 1399.)

This type of conventionalism, explicit in the work of Hacking (1999), comes in weaker and stronger forms. In contrast, the positive view that I hope to defend, which may be compatible with weaker forms of conventionalism (and is not necessarily associated with Mill's broader project), is that real or natural kinds are not arbitrarily selected: the classifications that reflect natural kinds must be objective in some important sense. We need to find criteria for distinguishing between conventional kinds and natural kinds.

I take it that this commitment to objectivity or non-arbitrariness in defining natural kinds is at the core of the mind-independence criterion. However, it is not straightforward to specify what this objectivity amounts to. An important source of supposedly objective classifications is scientific practice, but it seems that many scientific classifications that are successful and useful may not be fully mind-independent. I will discuss several examples of such successful but potentially conventional classifications in Sect. 3; these include non-naturally occurring elements, psychiatric kinds, and social kinds. Since all of these classifications enable us to make predictions and generalisations—they are *projectible*—many would regard them as natural kinds despite their apparent mind-*dependence*.[5] So, it is not surprising that the mind-independence criterion has recently come under severe criticism (e.g., Ereshefsky, 2018; Franklin-Hall,

---

[5] A *projectible* classification is here understood as having the ability to support (scientific) generalisations and be confirmed by 'observation of positive instances' (see Kim, 1992: 11).

2015; Khalidi, 2015, 2016; Magnus, 2012). I will discuss the role of projectibility in Sect. 4.

There is another important and related issue to note before we move on: the relationship between natural kinds and the properties that characterize them. The underlying question here is whether there is any difference between natural kinds and mere groups of objects characterized by a shared (natural) property. For example, chemical elements constitute a plausible example of natural kinds, whereas being green and round is merely a shared property. Thus, we have to distinguish between the idea of grouping some set of entities together on the basis of one or more shared properties, and the idea of grouping them together because they are of the same kind.[6] The point, to use Mill's term (without associating this view to Mill), is that it is *arbitrary* to classify all green and round things together, whereas it is not arbitrary, say, to classify all electrons together. While unit negative charge, a certain rest mass, and a half-integer spin seem to have a special relationship, there is a lack of such a relationship between greenness, roundness, and all green and round things. Interestingly, the idea that properties of instances of natural kinds are linked or unified in an important way is shared both by metaphysical and epistemic accounts of natural kinds (for an epistemic version, see Ereshefsky, 2018, p. 849).

The upshot is that we need to know *which* clusters of properties correspond to real kinds. My view is that properties are systematically clustered together in members of natural kinds *because* the particulars instantiating the clustered properties are members of certain natural kinds. This relates to the *why* question about systematic clustering of properties—which I will set aside—but it does not address the *how* question, i.e., how are the various properties that are clustered together related, what unifies their properties? In this regard, I favour a pluralist answer: there may be many different unification principles (UPs). What matters for a realist theory of natural kinds is that any such UP must be objective or mind-independent. So, instead of focusing on the mind-independence of the *entities* (natural kinds or their members), I propose that we should focus on the mind-independence of the UP that is ultimately responsible for the clustering of properties in instances of natural kinds. This will be the main focus of the paper (rather than the various UPs themselves), but we ought to have some idea about what the relevant UPs could be.

I will discuss UPs in more detail in Sect. 5, but let me give a classic example to get us started. Consider Boyd's (e.g., 1991, 1999) Homeostatic Property Cluster (HPC) theory. According to the HPC theory, the clustering of properties is explained by a shared causal structure among the members of a natural kind; each member of a kind is a property cluster kept in homeostasis by a (often lower-level) causal mechanism. There are many different causal mechanisms and each of them could be considered as a UP: they answer the *how* question regarding, say, the clustering of certain properties in members of *biological species*.[7] So, the relevant unification principle is ultimately

---

[6] We should remain open to the possibility that two members of a natural kind need not share *any* properties, or at least any intrinsic properties. A case in point might be biological species, insofar as they can be defined purely in terms of historical lineage (for discussion, see Godman et al., 2020). But I also leave open the possibility that biological species, being notoriously difficult to define, are not real in the intended sense.

[7] This is of course assuming that the HPC theory is the correct account of biological species. Biological species are a controversial case and there are also attempts to apply essentialist solutions to them; see Devitt

responsible for the systematic clustering of properties in instances of natural kinds. Note that in the case of the HPC theory, the 'common cause' noted with regard to UPs actually is a *causal* basis, since the mechanisms that the HPC theory examines are causal in nature. But this does not mean that UPs must always be causal—or indeed that *any* of them must be causal (e.g., if one denies objective causation) which is precisely why I remain neutral here and favour pluralism when it comes to UPs.

A more recent example, although based on a traditional idea, is provided by Godman et al. (2020). They suggest that natural kinds (or their properties) are typically 'unified by certain super-explanatory core properties' (Godman et al., 2020, p. 2). Such a core property, they propose, is some single property which causally explains the occurrence of the multiple properties shared by instances of a kind. One example they provide is the property of *atomic constitution*, say, of gold, which explains why all samples of solid gold have the same chemical and physical properties. On the face of it, this doesn't sound all that different from the HPC theory, but Godman et al. specify that where they differ from the usual causal accounts is that, on their view, 'the great preponderance of natural kinds owe their clustering of properties, not just to some causal structure or other, but to one single underlying property that serves as the common cause of all the other clustered properties' (2020, p. 6). So, this account as well identifies a causal explanation for the clustering of properties, albeit a rather different one from the HPC theory.

One further account, also very recent, is worth mentioning: the one provided by Hommen (2021; see also Dumsday, 2010; Oderberg, 2011; Keinänen & Tahko, 2019). Hommen's is an account of unification which is closest in spirit to the present account because he attempts to give a general metaphysical basis for unification, by way of a defence of *substantial kind universals*. In brief, Hommen asserts the neo-Aristotelian view according to which natural kinds are substantial universals and it is membership in the kind which (metaphysically) explains why certain properties are co-instantiated in its instances: 'kinds represent *unified ways of being*—both in an individual and in a collective sense: they account for the modal and temporal stability of character both within single particular objects and across what we then call different members of the same kind' (Hommen 2021, p. 297). While I am sympathetic to this suggestion, it mainly concerns the more specific question of what kind of entities natural kinds are, whereas the present paper focuses on the prior question of what it takes for a kind to be real in the first place. In any case, Hommen's account is an example of a non-causal explanation for the clustering of the properties, so we have some idea of the full range of possible UPs.

To be clear, I do not here wish to commit to any of these specific accounts, the details of which are outside the scope of this paper. The primary focus is, instead, to examine the status of unification principles more generally, and specifically with regard to the mind-independence criterion as it is applied to natural kinds.

Footnote 7 continued
(2008) and Austin (2018) and Godman et al. (2020) for three different approaches. Again, I also leave open the possibility that biological kinds are not real or natural kinds, and will not discuss them in detail here. It is worth mentioning though that Boyd as well could be regarded as being critical of the mind-independence criterion (cf. Lam 2020).

3 Mind-dependent kinds

An obvious challenge to the usefulness of the mind-independence criterion comes from the possibility of mind-*dependent* kinds, and especially those that would nevertheless seem like candidates for real kinds. Potential examples include non-naturally occurring transuranic elements such as the element with atomic number 99 (*Einsteinium*), psychiatric kinds such as mental disorders, and social kinds such as money.[8] I think that there are plausible ways to treat such cases without compromising the mind-independence criterion as applied to unification principles. However, before I demonstrate how the revised mind-independence criterion handles these cases, I will outline the challenge to the traditional mind-independence criterion.

Various authors (e.g., Ereshefsky, 2018; Franklin-Hall, 2015; Khalidi, 2016) have considered different ways to understand the mind-independence criterion. These authors have made important progress and I agree with them that the traditional mind-independence criterion, which is applied to entities, i.e., the natural kinds or their members, is not satisfactory. For instance, Khalidi thinks that even if were able to distinguish social and psychological kinds from cases such as artificial elements, such a distinction would not provide a good ontological account for distinguishing real kinds from *ersatz* kinds. More generally, he concludes that mind-independence should be considered irrelevant for the issue of realism about kinds.[9]

Regardless of the overall agreement on the failure of the traditional mind-independence criterion, it will be useful to briefly analyse what the source of the problem is. To this end, Khalidi (2016, p. 228) has put forward a helpful discussion of five possible ways to distinguish between innocuous and threatening senses of mind-independence. Khalidi's interest here is to determine whether these distinctions provide a way to distinguish between innocuous and problematic cases of mind-independence in such a way that physical, chemical, and biological kinds are on one side (innocuous) and psychological and social kinds are on the other (problematic). The reason to search for such a distinction is that, traditionally, friends of the mind-independence criterion have argued that psychological and social kinds are not real because they

---

[8] See Khalidi (2015, 2021) for an effort to classify different types of social and etiological kinds, which have an element of mind-dependence.

[9] It may be helpful to compare Khalidi's account to Ereshefsky's, who takes a somewhat different line, suggesting that 'we should worry about mind dependent classifications when such dependence does not promote the epistemic aims of scientific classification' (2018, p. 847). This reflects the difference between Khalidi's metaphysical account and Ereshefsky's epistemic account of kinds. A third possible approach is suggested by Franklin-Hall, who summarises her understanding of mind-independence as follows: 'For facts about which kinds are natural to be mind-independent is for those facts to be independent of all aspects of us—including our evaluative attitudes, conceptual schemes, and cognitive capacities' (2015, p. 928). But she is in fact advocating for a (partly) anti-realist account of natural kinds, so she is happy to abandon the mind-independence criterion. Of these three accounts, only Khalidi's is a realist, metaphysical account of kinds, which is why I am focusing on his account of mind-independence as well. It should also be mentioned that Mallon (2016) has put forward an important defence of a realist approach to human kinds. (I'd like to thank an anonymous reviewer for reminding me about this.) Mallon's social constructionist (albeit realist) project differs from the present one in an important way: he is specifically interested in human kinds such as race and gender and does not aim to give an overall account of natural kinds. In contrast, I aim to formulate a version of the mind-independence criterion that applies to all natural kinds equally. Accordingly, Khalidi's metaphysical account makes for the best comparison to the present project.

are mind-dependent, hence producing a criterion for distinguishing between 'real' and 'non-real' kinds (as Khalidi puts it). As already noted, Khalidi denies this, but it will be useful to see why. The dialectic of the present paper dictates that I will grant Khalidi's conclusion, namely, that the following candidate distinctions do not provide a satisfactory way to distinguish between real and non-real kinds. Moreover, my own goal is to provide a mind-independence criterion that is able to accommodate not just physical, chemical, and biological kinds, but also psychological and social kinds. But there is an important difference in my approach, as I will not apply the mind-independence criterion to *entities*. Since Khalidi dedicates a whole paper to these five distinctions, it will not be possible to do full justice to his arguments here. However, I will indicate some (further) reasons to dismiss these distinctions as being definitive of mind-independence regarding natural kinds. Here are the five candidate distinctions:

(1)   Mind-dependence of the kind vs. its instances
(2)   Causal vs. constitutive mind-dependence
(3)   Contingent vs. necessary mind-dependence
(4)   Mind-dependence vs. theory-dependence
(5)   Mental sustenance vs. initial manifestation

(1) concerns the distinction between a natural kind and its instances, i.e., individual members of a natural kind. It is easy to see that kinds like *Einsteinium* will count as innocuous on this distinction, given that the kind itself could be considered mind-independent even if its instances, i.e., the specific Einsteinium atoms that humans have synthesised, are, in a sense, mind-dependent. Even if we had synthesised no *Einsteinium* atoms, the fact that this specific element has a degree of stability (where 'stability' means that we are at least able to synthesise such elements, not necessarily that they have a long half-life) suggests that the natural kind itself (if we regard it as a separate entity) is mind-independent. However, the case of the various psychological and social kinds is more controversial as one might think that without any minds at all such kinds could not exist. So, on the face of it, (1) may appear to be a reasonable candidate for upholding the distinction between real and non-real kinds, provided that one is happy to dismiss psychological and social kinds. However, the details of this solution will depend on whether one believes that kinds can exist without any instances, i.e., whether they are indeed separate entities. If they cannot, then even supposedly real kinds like *Einsteinium* and other non-naturally occurring elements may appear threatened, because if we had not synthesised any samples of such kinds then they would likely not have come into existence. On the other hand, if kinds *can* exist without instances (say, in a Platonic realm), then there's no obvious reason to think that psychological and social kinds couldn't do so as well. Accordingly, (1) does not appear to give us a way to distinguish real from non-real kinds in the intended way.

As to (2), the thought here is that human minds may be causally involved in the creation of certain natural kinds, but these kinds may nevertheless be *constitutively* mind-independent if no part of their nature is itself mind-dependent. Clearly, kinds such as *Einsteinium* would count as real on this criterion, as no part of the constitution of its instances is mind-dependent. However, kinds such as mental disorders might

not count as real because their instances would seem to essentially involve mind-dependent constituents, i.e., the minds that are afflicted. So, this distinction as well may appear promising at the outset. But Khalidi correctly points out that some 'artificial or synthetic kinds too can be said to be constituted by human minds, since their makeup and structure have been engineered to affect and influence human minds' (Khalidi, 2016, p. 231). He has in mind drugs such as *methylphenidate* (*Ritalin*), which was designed with the purpose of influencing human mental states. While no part of the chemical *methylphenidate* actively depends on minds, it is at least arguable that the entity would not be what it is—or had not been synthesised at all—if it did not have the capacity to affect human neurochemistry (by increasing dopamine levels). Admittedly, the case isn't quite conclusive, but I am willing to grant to Khalidi that (2) also fails to deliver a plausible way to distinguish between real and non-real kinds.

(3) brings in explicitly the idea mentioned in connection to (1), namely, that certain kinds, such as mental disorders or certain artefactual or synthetic kinds, could not have come about without human minds (hence, the instances of these kinds are *necessarily* mind-dependent). In contrast, we can perhaps allow for the possibility that instances of kinds like *Einsteinium* could have been created spontaneously in the natural world (this may be controversial though, given the focused efforts required to create such kinds). Khalidi suggests that even many imaginary or fictional kinds, such as *phlogiston* or *fairies*, could have been instantiated without human minds, making them contingently rather than necessarily mind-dependent. I find this to be quite controversial, although I am willing to grant it to Khalidi for the sake of the argument.[10] A somewhat more plausible case against (3) comes from the idea that many chemical kinds (again, such as some drugs like synthetic opioids, e.g., *fentanyl*) have been purposely designed by humans and would be unlikely if not impossible to emerge on their own. This is of course a point closely related to the issue raised in connection to distinction (2); no kind that necessarily requires human intervention in order to come into being is mind-independent in the intended sense. While these objections are far from conclusive, it does appear that distinction (3) as well leaves something to be desired when it comes to demarcating real from non-real kinds.

The distinction suggested in (4) attempts to make the case for mind-independence in terms of ruling out social kinds such as *money* because the existence of these entities is dependent merely on there being a relevant theory (i.e., economics), whereas more fundamental kinds presumably do not require such a theory. Khalidi questions this, noting that it's certainly not the case that all social kinds are theory-dependent. This seems right: for instance, people may be (and have been) discriminated against based on an observed social kind such as race well before the relevant theories existed. So, (4) is a more clear-cut case, and also fails to reliably distinguish between real and non-real kinds.

Finally, according to (5), the distinction between real and non-real kinds might be specified in terms of whether the instances of the relevant kinds require human mental activity to *sustain* them as instances of those kinds. If such a need exists, these entities would seem mind-dependent in a problematic way, and hence the kind would be non-real. This could be a way for us to distinguish social kinds like money from kinds such

---

[10] Thanks to an anonymous reviewer for highlighting this controversy.

as transuranic elements because instances of the latter can carry on existing after their initial manifestation (albeit sometimes quite briefly) without any help from humans, whereas instances of social kinds would appear to require an active contribution of human minds in order to carry on existing. The existence of money, for example, is based on a type of social contract, where we exchange goods and services based on the agreement that the monetary instrument used will be valid for the purchase of such goods and services in the future. On the face of it, (5) seems promising, and Khalidi accepts that it does allow us to distinguish, say, between chemical and social kinds. Yet, Khalidi insists that even (5) will not give us a good criterion to distinguish between real and non-real kinds. The reason for this is that the problematic non-real kinds (he mentions *fairies* and *cold fusion*) do not cease to be when there are no human minds to sustain them, because they were never real kinds to begin with: 'there are no instantiations of such kinds because there is nothing in reality that would correspond to the specifications or characteristics that members of those kinds are posited to have' (Khalidi, 2016, p. 242). The point he is making, I take it, is that *fairy* is not a real kind, which is something that all parties in this debate would presumably accept, but the non-reality of *fairies* has nothing to do with how their existence is sustained. So, *fairy* would not be a real kind even if it did not require human mental activity to sustain it because there simply are no such entities as fairies in reality.

This brief overview of possible ways to adapt the mind-independence criterion shows us that the traditional mind-independence criterion on its own is unlikely to provide a systematic way to distinguish between real and non-real kinds. Notice that all the candidate distinctions are focused on the involvement of human minds in somehow bringing about, sustaining, or constituting the relevant *entities* that are the members of these natural kinds (or, exceptionally the natural kinds themselves). So, the traditional mind-independence criterion that Khalidi targets is solely focused on entities.

Recall that Khalidi as well must believe in the real/non-real distinction, because he defends a metaphysical, realist account of natural kinds. Importantly, this means that Khalidi must have some alternative criterion of reality for kinds in mind, given that he dismisses the mind-independence criterion. Indeed, he defends a 'causal criterion of reality' for kinds (Khalidi, 2018), which amounts to the idea that natural kinds are clusters of causal properties or 'nodes in causal networks'. The thought is that if this criterion works, then we could safely abandon the problematic mind-independence criterion and instead use the causal criterion to determine which kinds are natural or real. So, before I present my own approach, let's evaluate this alternative approach. This brings us to the next section, which focuses on the importance of *projectibility*, a central notion in Khalidi's account.

4 Natural kinds, projectibility, and entrenchment

What is projectibility? It is often invoked in connection to laws; as Kim puts it, a projectible law needs to have 'the ability to be confirmed by observation of positive instances' (Kim, 1992, p. 11). Kim argues that psychological states aren't projectible in this sense and hence that there are no special science laws concerning them, whereas Fodor (e.g., 1997) insists that the states 'pain' and 'believes that P' both 'express real

states, about which all the available evidence suggests that there are real laws' (Fodor, 1997, p. 150). If there are real laws about such states, then *pain* and the like are arguably real kinds—this is the link between projectibility and natural kinds. There is a venerable tradition tying projectibility and natural kinds together, and Khalidi is tapping into this tradition:[11]

> If I want to know whether there is a real kind *fairy*, I am primarily concerned with the question as to whether there is a group of individuals, all of whom share such properties as: smallness of stature (relative to humans), aerial flight, disposition to live in woodlands, ability to perform magic, and so on. They may not all have the exact same set of causal properties, but there must be enough similarity among them to participate in the same or similar causal processes. Accordingly, the categories that correspond to these kinds are projectible, feature in inductive inferences, and figure in scientific explanations (whether in the natural or social sciences). (Khalidi, 2016, p. 243.)

We may conclude that *fairy* is not a real kind; the categories that correspond to it are not projectible. Compare this with Ereshefsky's epistemic account, which requires natural kind classifications to be *defeasible*: 'if natural kind classifications are to help us understand and manipulate the world, then they should be vulnerable to disconfirming evidence' (2018, p. 846). The thought here is that only defeasible classifications are subject to empirical investigation and hence potentially of interest to science. So, we can see an important similarity between Khalidi's metaphysical account and Ereshefsky's epistemic account of natural kinds, the core here being scientific explanation and classification.

How do these criteria (i.e., projectibility, featuring in inductive inferences, and figuring in scientific explanations) fare with other potential natural kinds? The worry is that if a kind may be conjured by the mind willy-nilly and somehow end up in scientific practice, then all sorts of spurious kinds might qualify. Consider a fictional example: *C-rocks*. Let us suppose that I like to collect rocks from a specific beach in Cornwall. I select them based on a loose set of aesthetic properties, such as relative roundness and a reddish hue. But neither their roundness nor reddish hue is necessary for these rocks to end up in my *C-rock* collection; some are green and round, some are reddish and rectangular. The only necessary property is that these rocks are to be found on a specific beach in Cornwall (no matter how they ended up there). Surely, no one would think *C-rocks* belong to any well-defined natural kind—certainly not on the basis of the features I've just given. So, these rocks form an entirely arbitrary collection, even if they do share certain properties and, to my mind, form an aesthetic category. Yet, one may predict whether a given rock is suitable for my rock collection based on its location (a beach in Cornwall) and its properties (either circular or reddish). If you think, quite reasonably, that this is not enough for *C-rocks* to be considered a genuine kind, consider this: I decide that I am not going to collect very many *C-rocks* on my own, so I promise to buy a beer to everyone who brings me a *C-rock*. Soon, thirsty people line up to deliver *C-rocks* to me and, once the word in the local community

---

[11] The tradition goes further back than Kim's and Fodor's exchange from the 90s, in particular, to Goodman (e.g., 1983).

gets around, there is going to be a projectible generalisation concerning the delivery of *C-rocks* to me and the subsequent quenching of one's thirst by free beer.

This scenario is compatible with Ereshefsky's defeasibility requirement: as he specifies (2018, p. 852fn1), a classification may become a natural kind when defeasible property relations among the members of a category are introduced. In the case of *C-rocks*, the relevant defeasible property relation would be the connection between *C-rocks* and behavioural patterns related to free beer. It's not difficult to see that despite the silliness of the example, this could conceivably have even scientific interest; one day, sociologists may write about the notorious *C-rock* incident in Cornwall, and the relevant kind term could figure in scientific explanations.

Khalidi is aware of this type of counterexample to the causal criterion of reality, and he tackles it head on. The cases that he compares are illuminating, take *race* and *witch*. We may assume that *race* does not correspond to a biological kind, but based on the above criteria it does seem to constitute a social kind—Ereshefsky (2018, p. 847) explicitly counts it among scientifically useful social kinds. Among other things, people are unfortunately discriminated against based on race, so it does have real and projectible causal effects, where 'it' refers to the socially constructed kind. Similarly, *witch* may not be a social kind anymore, but it did also have causal effects when belief in witches was common. This leads Khalidi to conclude that *witch* was a real social kind in the same way that *race* is a real social kind now, and it may someday cease to be a real kind in the same way that witches have ceased to be a real kind. But one may ask: is this really the result that we wanted? What exactly differentiates kinds like *race* and *witch* from my *C-rocks*? Khalidi has an answer:

> The difference would seem to be based on how widespread the effects are and how robust, long-lasting, and entrenched the causal profile. If the category in question only affects the behaviour of a single person, leading to minimal social influence, then there is little reason to regard it as a social kind (Khalidi, 2016, p. 243).

Accordingly, as beautiful and precious as my *C-rock* collection may be, it would presumably not qualify as a natural kind while it only affects my own behaviour. But once I introduce the free beer… well, then the situation may change.

Strikingly, Ereshefsky (2018, p. 852) attempts something very similar to Khalidi's line of reasoning in response to the worry that his defeasibility requirement is too inclusive: he clarifies that it's a necessary rather than a sufficient condition for natural kinds and that we would need to add some additional condition to make it sufficient. What could such an additional condition be? He suggests that natural classifications should be 'well tested and not defeated' (*ibid.*), where 'well tested' is a placeholder for scientists to fill in. Ereshefsky adds that 'If the category tall people was investigated by science, then it would be a good candidate for being a natural kind' (2018, p. 853). This looks tantamount to Khalidi's idea of an entrenched causal profile, at least if we add that the causal profile associated with a category needs to be investigated by science. In fact, in this regard Ereshefsky distances himself from Khalidi's account, claiming that Khalidi places too much emphasis on causality (Ereshefsky, 2018, p. 854). Ereshefsky thinks that this makes Khalidi's account too *restrictive* rather than too lenient (because he thinks that the relevant relations among properties of a

kind need not be causal). To an extent, this highlights the difference between Khalidi's metaphysical and Ereshefsky's epistemic account, but it's obvious that both regard it important that we deal with well tested and entrenched scientific classifications. My worry is that both accounts may be much too lenient.

The move that Khalidi and Ereshefsky both seem to be attracted to is well-rehearsed: projectibility and the relative *entrenchment* of natural kind predicates has been proposed as a criterion for the 'reality' of kinds before, or at least as the 'epistemic marker for the metaphysical relation of causality' (Khalidi, 2013, p. 80) that underlies this reality. For instance, Antony, in a paper discussing Kim's and Fodor's views, suggests that (natural kind) predicates 'can only get entrenched *if the world cooperates*' (Antony, 2003, p. 12). If monetary exchanges don't really have something important in common despite their myriad physical dissimilarities, then 'monetary exchange' could not have the role in our language that it in fact has. Antony notices that entrenchment by itself cannot be a sufficient criterion for reality. That's because there are plenty of projectible predicates that are suitably entrenched in human language, but which we do not wish to count among real natural kinds. Somewhat ironically, she mentions *witches* and *angels* as such entrenched but non-real kinds, where the problematic understanding associates supernatural properties to these kinds. In contrast, Antony suggests that 'one can find real similarities in the socio-political circumstances of a large group of women who were accused and convicted of witchcraft, and so can explain the utility of "witch"' (Antony, 2003, p. 14). So, the entrenchment of "witch" is explained by the utility of the socio-political grouping rather than any supernatural properties. Accordingly, *witches* and indeed *angels* may be a social kind on Antony's criteria too.

How did we get here? It now seems that social kinds can come into existence quite easily, and also go out of existence, simply because a sufficiently large group of people believe in them (and this belief gets picked up by scientific practice). Even Khalidi does want to rule out some kinds, such as *fairy*. Yet, there is a large group of children who believe in the *tooth fairy* and indeed this belief sometimes results in teeth being exchanged for cash. There is clearly a lot of vagueness in deciding just how robust, long lasting, well tested, and entrenched the causal profile of a kind needs to be in order to be counted as a real natural kind. The problem is that there does not appear to be an objective unification principle that underlies these classifications: on Khalidi's view, we determine the reality of (social) kinds based on their usefulness in predicting social interaction. This is far from the type of robustness that many of us (i.e., realists) require from natural kinds. In fact, Khalidi does not specify what kind of 'robustness' he has in mind in the passage that I have quoted above (from Khalidi, 2016, p. 243). Accordingly, I can only speculate how he intends to understand the notion. Instead, I offer an understanding of robustness which is well-supported in the literature on laws of nature, where the notion is used to demarcate real laws from universal generalizations: 'This counterfactual robustness that the laws enjoy—their 'inevitability', 'stability', or 'preservation', as it has variously been called—is the very feature of the laws that distinguishes them from merely accidental universal generalizations' (Tan, 2020, p. 577). I will put this idea to use below.

Let me attempt to summarise why I think that kinds like *witch* or *tooth fairy* cannot be real.[12] In both cases, the projectibility associated with the supposed kind seems to be derived from predictions that we make about human behaviour based on the *beliefs* or other similar psychological states that drive the behaviour of a social group. Children's belief in tooth fairies explains certain social practices (putting teeth under one's pillow and expecting money in return), and it is these and related social practices (such as adult storytelling) that anthropologists study and associate with *tooth fairy*. So, it is not of course the fictional entity, tooth fairy, that has causal relevance in this case, just as it is not the supernatural entity, witch, that has causal relevance (it is instead the social kind *witch*). Rather, what anthropologists will find interesting are the myriad social responses to the belief that there are witches or tooth fairies, and the associated human behaviour related to these beliefs and other relevant mental states. Perhaps, then, we should accept both *witches* and *tooth fairies* as natural kinds in this revised sense?

Unfortunately, I think that this the wrong result, and the reason for this is precisely because there is no objective, robust unification principle at hand in either case. It's true that in both cases we can predict and explain a variety of behaviours, but I contend that the real explanation for these behaviours cannot be captured in terms of kinds like *witch* or *tooth fairy*. Consider witches: some people who discriminated against women labelled as witches no doubt had the belief that these people were performing black magic and so on. But many others were simply following a group mentality, e.g., by participating in lynching, and others were well aware that no black magic was being performed, but simply used the situation to get rid of their enemies. All of these behaviours are clustered around the notion of 'witch', but the psychological processes behind them are quite varied: some are based on false beliefs, some on envy, hate, political or religious motives, and so on. Examples are easy to find (see, e.g., Bever, 2002), so I will not labour the point further.

Of course, one could try to isolate further sub-kinds here; maybe there is a kind *witch** that just captures those social practices that are based on, say, the false belief that witches perform black magic. But how could we possibly hope to isolate the unification principle for *that* kind? Is it not more plausible that what is really doing the work here is not a *social* kind at all, but, perhaps, a variety of psychological kinds associated with human behaviour in complex social situations? At least some of these kinds may indeed be real kinds, but in order to determine whether or not they are, we would need to engage in further research—I will return to this point in Sect. 5.2 in connection to psychological kinds.

The upshot of this discussion is that if we are unable to identify the underlying unification principle, then there seems to be nothing on offer that would differentiate natural kinds from (bundles of) causally efficacious properties, such as those that we may associate with the myriad behavioural responses associated with the notion of 'witch'. The causal criterion that Khalidi proposes was supposed to give us just this, but it is evident that it is not, after all, enough that there is a causal profile associated with a kind—it has to be suitably entrenched as well. The good news is that I think we are now in a position to identify the problem more accurately, and to re-assert

---

[12] I'd like to thank an anonymous reviewer for pushing me to justify this further.

the importance of the (revised) mind-independence criterion for realism about natural kinds.

5 Unification principles for natural kinds

5.1 Mind-independence reconsidered

We have seen that each of the many ways to understand the mind-independence criterion has problems, but so do the various versions of the causal criterion which is sometimes suggested as a replacement. I propose that there is still mileage in the mind-independence criterion if we understand it to concern the unification principles (UPs) for natural kinds rather than entities, such as the members or instances of natural kinds or natural kinds themselves. The UPs concern the *properties* of the instances of natural kinds: we wish to understand how certain properties are systematically and reliably clustered together in instances of natural kinds. I suggest that we may identify real kinds in terms of the mind-independence of their relevant unification principle as follows[13]:

> **Mind-independence criterion (MIC):** A kind is natural or real if and only if there is a mind-independent *unification principle* that is responsible for the clustering of properties being tracked by the relevant kind term.

In contrast to the candidate distinctions regarding mind-independence that were discussed in Sect. 3, MIC does not concern entities but rather the principle by which entities (i.e., the properties of the kind's members) come together. Notice that we should not focus on any individual instances of clustering here, because that leads us back to focusing on the entities involved in the specific case. Instead, our focus should be on the general principle of clustering, i.e., the general unification principle (UP). Accordingly, the situation is partially analogous to the first type of distinction that we discussed in Sect. 3, from Khalidi: mind-dependence of the kind vs. its instances. However, in this case it is not the kind that we are focusing on but rather the unification principle. The best way to explicate MIC is to consider how it fares with some of the supposedly mind-*dependent* kinds that we have been discussing above (Sect. 3 in particular). What we must ask is whether the relevant UP objectively unifies the properties of the kind.

Take the relatively easy case of *Einsteinium* first. Rather than focusing on the mind-independence of instances of *Einsteinium* or the constitution of those instances, as discussed in Sect. 3, we should ask whether the clustering of properties in instances of *Einsteinium* is objectively determined. This is effectively a question about the principles which determine the formation of *Einsteinium* atoms—the common cause of the clustering of certain properties (e.g., nuclear charge, mass) in instances of *Einsteinium*. The relevant UP will certainly involve the strong force, which holds nuclei together, but more generally we may regard the narrowest UP for elements in general to be nucleosynthesis. The precise procedure of nucleosynthesis is a complex

---

[13] Thanks to an anonymous reviewer for helping to pin down the best possible formulation of this principle.

matter and research is still on-going: nuclear fusion in stars is capable of producing elements up to iron, but heavier nuclei need to be created by different processes, such as by neutron capture in more massive stars. For even heavier elements, especially superheavy elements, the most recent research suggests that neutron star mergers might be the most likely source.[14] Of course, *Einsteinium* itself might not be produced even by these processes, given that it does not occur in nature (as far as we are aware). Still, the same basic principles—and certainly the same laws of nature—that enable nucleosynthesis in stars also apply in our particle accelerators. Accordingly, for the purposes of identifying an objective UP in the sense of MIC, it makes little difference whether the process is initiated by us in particle accelerators or by neutron star mergers or supernovae. It is not mind-independence of the *process* or *event* (such as nucleosynthesis in stars) in question that matters, but rather the underlying *laws* that enable that process.

One upshot of the preceding discussion is that there are several different UPs that need to be specified under the more general heading of 'nucleosynthesis', given that the nuclear fusion process in stars only works for elements below iron. The reason for this limitation is that the fusion of iron would subtract energy rather than provide it. We can set these complications aside though. The main point of this brief case study is to clearly establish that the relevant unification principle for *Einsteinium* is indeed objective in the required sense, and as we have seen, the fact that instances of *Einsteinium* have been created by humans in a particle accelerator is quite irrelevant here. It also follows that some subsets of elements (e.g., those lighter than iron in the periodic table vs. those heavier than iron) may have different UPs. Does this entail that *element* is not a genuine natural kind? Not necessarily: while the narrowest UP when discussing individual elements may differ for different elements, there is still the general UP of nucleosynthesis that is the narrowest common UP for all elements. This does of course raise questions about whether the higher-level UP, nucleosynthesis, reduces to a number of narrower, lower-level UPs, but it is a further question how we deal with this issue, i.e., does reducibility entail unreality? In some cases, it may well do so, but the case of elements is not straightforward: even though details of nucleosynthesis may vary for different elements, there may still be sufficient cohesion in the underlying process to identify a unique UP for elements in general (instead of a "wildly disjunctive" one, cf., Kim, 1992, p. 10).

Moving on to trickier examples, what should we say about UPs regarding psychological or social kinds? I think that there is no reason why there couldn't be objective (and non-reducible) unification principles for such kinds. But some cases certainly lack them. Consider *tooth fairy* again: the properties and the causal profile that we associate with tooth fairies is based on stories that we tell children. Yes, there is a fairly robust and entrenched causal profile that may be associated with *tooth fairy*—entrenched enough to be of interest to anthropologists (e.g., Wells, 1997), but that causal profile is not held together by anything real; *tooth fairy* could just as well have a *different* causal profile. So, in contrast to the robustness that Khalidi (2016, p. 243) may have in mind—which we may associate even with *tooth fairy*—there is another, important type

---

[14] See the special issue on superheavy elements edited by Düllmann et al. (2015).

of robustness that *tooth fairy* lacks, namely the type of modal or counterfactual robustness that is typically associated with laws of nature, as specified earlier (again, see Tan, 2020, p. 577). The causal profile of *tooth fairy*—even when considered as a social kind associated with human behaviour rather than as a fictional kind—is thoroughly contingent, lacking counterfactual robustness, because it has been arbitrarily selected. Now, contingency doesn't automatically entail arbitrariness. However, as the level of contingency increases, so does the risk of arbitrariness, and hence the violation of MIC. Presumably, even the causal profile associated with *tooth fairy* is not completely arbitrary because it will be importantly related to the social and cultural practice of human storytelling and hence it is influenced by our own psychology. Accordingly, any robustness that one might associate with *tooth fairy* is fully reducible to (and may ultimately be explained by) the (supposed) robustness of these more fundamental features, such as human psychology.

One further clarification is in order here: it may appear that MIC suggests that any grouping produced by an intentional action will automatically fail the mind-independence criterion.[15] This is not the necessarily the case. Rather, a grouping whose UP is produced by an intentional action would be at risk of violating the mind-independence criterion. The case of *fairy* is a perfect example. The relevant intentional content here are the properties that our stories associate with *fairy*, properties (mentioned by Khalidi) such as smallness of stature, aerial flight, disposition to live in woodlands, ability to perform magic, and so on. But the reason that these properties are associated with *fairy* are nothing to do with the kind *fairy*, but rather something to do with the history and psychology of human story telling (again, see Wells, 1997). So, if there is an objective UP at play here, it will not be associated with the kind *fairy*, but rather with some much more general kind concerning human psychology. The issue that remains is that even this type of UP may lack the type of robustness that we have come to associate with natural kinds.

We are now faced with the difficult choice of drawing a line: how much robustness is enough, or how much contingency is too much for a UP to count as being objectively determined? In fact, I don't think that this is the right question to ask because robustness is merely a heuristic, it provides *prima facie* evidence for the reality of a kind, but it does not settle it. A kind is natural or real if it satisfies MIC, whereas counterfactual robustness is a helpful guide: roughly, the more robustness there is, the more likely it is that the UP is objectively determined. To get a better sense of what is required for MIC to be satisfied, the best practice is to consider further examples.

So, let's briefly recap the main issues, before discussing a new example in the next section. What should we say about *witches* or *race*? The same strategy applies. The fact that there is, was, or could be systematic discrimination based on some set of properties that certain groups of people share is not enough, given that this discrimination may be based on a myriad of psychological responses, as discussed at the end of Sect. 4 (again, see also Bever, 2002). It is precisely because there is no mind-independent unification principle in these cases that we should discount them, no matter how long lasting or entrenched their causal effects may be.

---

[15] I'd like to thank an anonymous reviewer for drawing my attention to this.

Indeed, these cases look like the case of *C-rocks* from Sect. 4. We may identify a causal profile for these supposed kinds and if things develop suitably, they may become entrenched in society to the extent that we can make accurate predictions about people's behaviour—these classifications may be useful (or harmful) and projectible. Hence, there would also be a modest level of robustness. But this type of *epistemic* criterion is not a reliable guide to real kinds, even if it is sometimes a reliable guide to useful classifications based on some shared properties. If we cannot point to an objectively determined UP, which would provide a metaphysical explanation for the projectibility of such classifications, then MIC will be violated. In other words, the robustness associated with these classifications is only based on the fleeting stability of our social interactions and biases. Again, such classifications can be epistemically useful, but they lack the counterfactual robustness of real kinds. Accordingly, an epistemic account of natural kinds, such as the one provided by Ereshefsky (2018), is not directly competing with the present account.

5.2 Unification principles for higher-level kinds

I hope to have made it clear with the previous examples that, on my view, we should be less liberal about natural kinds than many other recent accounts suggest. Yet, I do wish to leave room, at least in principle, for genuine higher-level natural kinds, including psychological and social kinds. Plausible examples are not easy to find, but let us consider one interesting example: psychiatric kinds, such as *depression*.

On the face of it, a kind like *depression* may seem to lack an objective unification principle, as it is defined in DSM-5 on the basis of psychological and somatic symptoms as opposed to their causes (it is specified that the symptoms must not be caused by substance abuse or other medical conditions). There is also an element of arbitrariness, because in order to be diagnosed with depression, an individual must be experiencing five or more of the eight diagnostic symptoms (such as depressed mood, fatigue, and suicidal thoughts) associated with depression during the same two week period, and at least one of the symptoms should be either (1) depressed mood or (2) loss of interest or pleasure. Although *depression* is defined quite loosely, this does not entail that there could not be an objective UP that underlies the mutual occurrence of its symptoms. But it is quite difficult to accurately diagnose conditions like depression as we do not truly know (all of) its causes. At the same time, we do need classifications of phenomena like depression: we can make important predictions about people's behaviour and assign treatment based on them. So, there is also a modest level of robustness. This robustness gives some weak *prima facie* evidence in support of there being an objective UP for depression, but if we are unable to identify it then we are not in a position (epistemically) to conclusively determine whether depression has an objective UP or not.

The case of depression is interesting, as the symptoms associated with it, such as loss of interest or pleasure, could be the result of any number of different causes. So, while it is useful to have the general category of *depression*, which is reliably associated with several symptoms, the true causes of these symptoms may remain elusive (even though we may be able to treat the symptoms). The upshot is that in

the case of *depression* it may be best to remain cautious, given the modest, defeasible evidence available to us—it could end up being similar to *witch* in that the causal profile associated with it is ultimately based on convention rather than an objective UP.

But what about the fact that the clustering of depression's symptoms is surely sustained by the depressed individual's psychology, i.e., doesn't the UP appear to be to a large extent mind-*dependent* (because it is sustained by the individual's psychology)?[16] Well, recall that our focus here should not be on any individual instances of clustering, because that would lead us back to focusing on entities and their dependence on the specific mind that we are focusing on, as this line of reasoning suggests. It shouldn't matter if, say, the clustering of depression's symptoms is sustained by the depressed individual's psychology. What matters, instead, is whether the general unification principle that underlies this clustering has an objective source, such as a causal mechanism, law of nature, or essence.

Does any of this matter if *depression* is nevertheless a useful classification? Here is one reason to think that it does: if *depression* is a real kind, it is also useful because the underlying UP is itself mind-independent and this would allow for very accurate predictions—presumably more accurate than the predictions available to us previously. If *depression* is not a real kind, then there are, presumably, more accurate, lower-level natural kinds that are jumbled together into *depression* merely because of our epistemic limitations, making the unification principle mind-*dependent*.[17] In that case, we may be missing out on the more accurate predictions. It may be interesting to note that this distinction roughly corresponds to the distinction between a *syndrome* and a *disease*, where the former is a term used to describe a collection of symptoms which may or may not be attributable to some disease and generally do not have a clear (known) causal basis. Diseases, in contrast, typically have a known set of symptoms that can be attributed to them and often there is also a clear causal basis, be it pathogenic, hereditary, or something else (see Williams, 2011). So, just like we would generally like to know whether there is a singular disease behind a set of correlated symptoms, we should also like to know whether there is a singular natural kind that unifies a cluster of properties, even if the motivation in the latter case is different. Indeed, there is also another, albeit more arcane reason to care about this, namely, the desire to get the ontology of natural kinds right!

5.3 Unification principles for fundamental kinds

What further potential unification principles are there besides the ones already proposed? It is surprisingly difficult to find good or uncontroversial candidates, but as I have already hinted, it is plausible that various laws of nature are responsible for clustering of properties at lower levels. For instance, when one carbon atom and four hydrogen atoms form a *methane* molecule through covalent bonding, this process is (in

---

[16] Thanks to an anonymous reviewer for raising this objection.

[17] One way to read this is that if it turns out that there is no objective UP for *depression*, then we could say that the higher-level kind *depression* is reducible to the lower-level kinds that are responsible for the symptoms that we've classified together (for epistemic purposes) as depression.

part) governed by electrostatic interaction. So, to simplify a little, we might consider electrostatic interaction to be one type of UP.

In many cases it may be very difficult to identify what the relevant UP is. The question is of course closely tied to empirical research, so it faces the same epistemic limitations. We can make a little more progress by reiterating what the UP actually explains: modal and temporal robustness (but, importantly, not entrenchment). Hommen, again, puts this nicely: 'kinds represent *unified ways of being*—both in an individual and in a collective sense: they account for the modal and temporal stability of character both within single particular objects and across what we then call different members of the same kind' (2021, p. 297). In general, it is only when we observe such modal and temporal robustness that we postulate the existence of a natural kind. But we often do so without knowing what the underlying UP is. Nowhere is this more striking than in the case of (supposedly) fundamental natural kinds, such as *electron*. Such cases are notoriously difficult for some theories of natural kinds. In particular, causal theories like the HPC theory rely on identifying a causal mechanism that explains the clustering of properties. In the case of fundamental kinds this is unlikely to be very useful, as there are no obvious causal connections between, say, unit negative charge, half-integer spin, and the rest mass of electrons, even though all three of these are presumably definitive of, and unified in, the natural kind *electron*. This is another reason why the 'common cause' in the initial definition of UPs may need to be interpreted non-causally (e.g., in terms of a 'common ground' as Ismael & Schaffer, 2020 put it).

The problematic status of fundamental kinds, and the lack of clear causal mechanisms in particular, has led some to suggest that at least fundamental natural kinds unify their properties as a matter of *brute fact*. One version of this idea may be found in Chakravartty's work (2007, p. 171), who suggests that fundamental natural kinds, such as electrons, have their core properties (mass, charge, spin) as a matter of brute fact. Importantly, this is not supposed to entail that such fundamental kinds are not real; presumably, the brute fact as well could be an objective unification principle. In any case, I think it would be a mistake to conclude that there is no UP just because some cases of unification look to *us* like brute facts. It is another matter whether or not we are in a position to *know* what the relevant UPs are. So, what should we say about cases like *electron*? Well, the fact that the properties of electrons are very robustly clustered together strongly suggests that *electron* is a real kind, but this is, just as in other cases, *prima facie* evidence. If we cannot identify the relevant UP, then it could still turn out that *electron* is not a real kind after all—this would be a very surprising result, but the evidence in favour of *electron* being a real kind is nevertheless defeasible. However, even in this case, *something* would of course be responsible for the clustering of the properties that we have identified as an electron, perhaps an even more fundamental kind. One line of thought that might go some way toward supporting the latter suggestion comes from quantum field theory and the fact that we can interpret electrons as the excitation states of quantum fields—perhaps there are in fact no particles in quantum field theory at all, in which case the true natural kinds would not be things like *electron*—at least not if considered as a *particle* (for discussion, see Fraser, 2008; Ladyman, 2015). To be clear, I am not suggesting that this is really the case, but in the spirit of fallibilism we should be able to accept even this type of result, and I regard it is an advantage of the present theory that it can be accommodated.

6 Conclusion

The previous examples serve to highlight that just because a grouping or classification may be based on a mind-*dependent* unification principle does not mean that it could not be useful or could not track causal structures. Such a classification or the associated set of properties may satisfy some looser sense of cohesion or unity, say, given a particular epistemic goal. Moreover, the result that some kinds themselves may be mind-dependent is not the problem; rather, problems arise if we do not recognize that the clustering of properties reified as a kind is picked up on the basis of our epistemic interests rather than some mind-independent unification principle. We may conclude that the mind-independence criterion for natural kinds is still valid—in the context of a metaphysical account of natural kinds—but it should be understood as concerning the underlying unification principle rather than the kind itself. The epistemic question concerning which kinds are genuine or natural must be settled by a careful case by case analysis and it will generally also involve empirical work.

Let me conclude by addressing one worry that may remain.[18] I have repeatedly appealed to potential unification principles such as causal mechanisms, laws of nature, and essences. But if the mind-independence of natural kinds is deferred to these other notions, then have we really explained what the mind-independence regarding natural kinds amounts to? Moreover, at least causal mechanisms would not always seem to be fully mind-independent, as humans are also involved in causal interactions. We can indeed debate the objectivity of each of these candidates for unification (causal mechanisms, laws of nature, essences), but I would insist that insofar as we wish to be realists about natural kinds, some objective source for unification is required—this is the core claim of the paper. However, I wish to remain neutral in this paper about these broader ontological issues, even though some combinations of views are certainly more plausible or at least easier to defend than others.

In conclusion, the revised mind-independent criterion proposed in this paper does not need to commit us to any specific account of the underlying ontology. Ultimately, my proposal here is that unless an objective source of unification can be found, be it in terms of an objective sense of causal mechanisms, laws of nature, essences, or some combination of these, then natural kind realism collapses. Fortunately, I think that prospects for finding at least some unification principles on these lines are promising.[19]

---

[18] Thanks to an anonymous reviewer for articulating this worry.

[19] For further discussion, see Tahko (2015, 2021).

References

Antony, L. (2003). Who's afraid of disjunctive properties. *Philosophical Issues, 13*, 1–21.

Austin, C. J. (2018). *Essence in the age of evolution: A new theory of natural kinds*. Routledge.

Bever, E. (2002). Witchcraft, female aggression, and power in the early modern community. *Journal of Social History, 35*, 955–988.

Bird, A. (2018). The metaphysics of natural kinds. *Synthese, 195*, 1397–1426.

Bird, A., & Tobin, E. (2018). Natural kinds. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy* (Spring 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2018/entries/natural-kinds/.

Boyd, R. (1991). Realism, anti-foundationalism, and the enthusiasm for natural kinds. *Philosophical Studies, 61*, 127–148.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 141–185). MIT Press.

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

Devitt, M. (2005). Scientific realism. In F. Jackson & M. A. Smith (Eds.), *Oxford handbook of contemporary philosophy* (pp. 767–791). Oxford University Press.

Devitt, M. (2008). Resurrecting biological essentialism. *Philosophy of Science, 75*, 344–382.

Düllmann, C. E., Herzberg, R. D., Nazarewicz, W., & Oganessian, Y. (2015). Special issue on superheavy elements. *Nuclear Physics A, 944*, 1–690.

Dumsday, T. (2010). Natural kinds and the problem of complex essences. *Australasian Journal of Philosophy, 88*, 619–634.

Ereshefsky, M. (2018). Natural kinds, mind independence, and defeasibility. *Philosophy of Science, 85*, 845–856.

Fodor, J. (1997). Special sciences: Still autonomous after all these years. *Philosophical Perspectives, 11*, 149–163.

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

Fraser, D. (2008). The fate of "particles" in quantum field theories with interactions. *Studies in History and Philosophy of Modern Physics, 39*, 841–859.

Godman, M., Mallozzi, A., & Papineau, D. (2020). Essential properties are super-explanatory: Taming metaphysical modality. *Journal of the American Philosophical Association, 6*, 316–334.

Goodman, N. (1983). *Fact, fiction and forecast* (4th ed.). Harvard University Press.

Hacking, I. (1999). *The social construction of what?* Harvard University Press.

Hommen, D. (2021). Kinds as universals: A neo-Aristotelian approach. *Erkenntnis, 86*, 295–323.

Ismael, J., & Schaffer, J. (2020). Quantum holism: Nonseparability as common ground. *Synthese, 197*, 4131–4160.

Keinänen, M., & Tahko, T. E. (2019). Bundle theory with kinds. *Philosophical Quarterly, 69*(277), 838–857.

Kendig, C., & Grey, J. (2021). Can the epistemic value of natural kinds be explained independently of their metaphysics? *British Journal for the Philosophy of Science, 72*(2), 359–376.

Khalidi, M. A. (2013). *Natural categories and human kinds: Classification in the natural and social sciences*. Cambridge University Press.

Khalidi, M. A. (2015). Three kinds of social kinds. *Philosophy and Phenomenological Research, 90*, 96–112.

Khalidi, M. A. (2016). Mind-dependent kinds. *Journal of Social Ontology, 2*, 223–246.

Khalidi, M. A. (2018). Natural kinds as nodes in causal networks. *Synthese, 195*, 1379–1396.

Khalidi, M. A. (2021). Etiological kinds. *Philosophy of Science, 88*(1), 1–21.

Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research, 52*, 1–26.

Ladyman, J. (2015). Are there individuals in physics, and if so, what are they? In A. Guay & T. Pradeu (Eds.), *Individuals across the sciences* (pp. 193–206). Oxford University Press.

Lam, K. H. (2020). The realism of taxonomic pluralism. *Metaphysics, 3*, 1–16.

Lemeire, O. (2021). No purely epistemic theory can account for the naturalness of kinds. *Synthese, 198*(Suppl 12), 2907–2925.

Lowe, E. J. (2006). *The four-category ontology: A metaphysical foundation for natural science*. Oxford University Press.

Lowe, E. J. (2015). In defence of substantial universals. In G. Galluzzo & M. J. Loux (Eds.), *The problem of universals in contemporary philosophy* (pp. 65–84). Cambridge University Press.

Magnus, P. D. (2012). *Scientific enquiry and natural kinds: From planets to mallards*. Palgrave Macmillan.

Mallon, R. (2016). *The construction of human kinds*. Oxford University Press.

Mill, J. S. (1843). *A system of logic, ratiocinative and inductive*. In J. M. Robson (ed.), *Collected works of John Stuart Mill*. University of Toronto Press and Routledge and Kegan Paul.

Oderberg, D. S. (2011). Essence and properties. *Erkenntnis, 75*, 85–111.

Psillos, S. (2002). *Causation and explanation*. McGill-Queen University Press.

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

Tahko, T. E. (2015). Natural kind essentialism revisited. *Mind, 124*, 795–822.

Tahko, T. E. (2021). *Unity of science. Elements in philosophy of science*. Cambridge University Press.

Tan, P. (2020). Ideal laws, counterfactual preservation, and the analyses of lawhood. *Australasian Journal of Philosophy, 98*(3), 574–589.

Tobin, E. (2010). Crosscutting natural kinds and the hierarchy thesis. In H. Beebee & N. Sabbarton-Leary (Eds.), *The semantics and metaphysics of natural kinds* (pp. 179–191). Routledge.

Wells, R. (1997). The making of an icon: The tooth fairy in North American folklore and popular culture. In P. Narváez (Ed.), *The good people: New Fairylore essays* (pp. 426–446). University Press of Kentucky.

Williams, N. E. (2011). Putnam's traditional neo-essentialism. *Philosophical Quarterly, 61*, 151–170.