European Journal for Philosophy of Science (2022) 12:72
https://doi.org/10.1007/s13194-022-00500-x

PAPER IN THE PHILOSOPHY OF THE BIOMEDICAL SCIENCES

Reactive natural kinds and varieties of dependence

Harriet Fagerberg[1]

Received: 18 November 2021 / Accepted: 7 November 2022 / Published online: 7 December 2022

Abstract
This paper asks when a natural disease kind is truly 'reactive' and when it is merely associated with a corresponding social kind. I begin with a permissive account of real kinds and their structure, distinguishing natural kinds, indifferent kinds and reactive kinds as varieties of real kind characterised by super-explanatory properties. I then situate disease kinds within this framework, arguing that many disease kinds prima facie are both natural and reactive. I proceed to distinguish 'simple dependence', 'secondary dependence' and 'essential dependence' between a natural kind and its classification, and argue that a natural kind is only really reactive, in an important sense, under conditions of essential dependence. On this basis, I offer a principled hypothesis for why psychiatric kinds may be more metaphysically unstable than paradigm somatic disease kinds.

Keywords Medicine · Psychiatry · Neuroscience · Natural Kind · Interactive Kind · Biology

1 Introduction

This paper is about disease kinds – the classes of illness, disorder and infirmity which feature in medical science and nosology. I argue that the property clusters featuring in medicine often have an interesting structure: they constitute natural kinds, but are in part explained or impacted upon by our practices of classification. For example, people with breast cancer share properties across instances *both* due to uncontrolled cell-division in their breast tissues *and* due to being classified as 'people with breast cancer' in a particular social context. This peculiar structure of disease kinds prompts the question: when is a natural disease kind *truly* reactive in

This article belongs to the Topical Collection: Reactivity in the Human Sciences
Guest Editors: Marion Godman, Caterina Marchionni, Julie Zahle

✉ Harriet Fagerberg
harriet.fagerberg@hunter.cuny.edu

[1] Hunter College and The Graduate Center, City University of New York, New York, NY, USA

response to our social classifications, and when is it merely correlating with a reactive social kind?

In this paper, I propose an answer to this question. I begin with a permissive account of real kinds and their metaphysical structure – following Millikan (2000, 2017) and Godman et al. (2020) – distinguishing natural kinds, indifferent kinds and reactive kinds. I proceed to situate disease kinds within this structure by arguing, via a number of cases studies, that many disease kinds *prima facie* satisfy the conditions for being natural as well as reactive. On this basis, I raise the question: when is a disease kind *really* reactive, and when is it just accompanied by or intersecting with a distinct social kind of the same name? I distinguish 'simple dependence', 'secondary dependence' and 'essential dependence', and argue that a disease kind is only truly reactive under conditions of essential dependence. I close by offering a principled neuroscientific hypothesis for why psychiatric kinds really may be more metaphysically unstable than other kinds in biomedicine – (some) psychiatric kinds are essentially dependent.

2 Kinds of kinds

In what follows, I provide an approach to real kinds in general, following Millikan (2000, 2017) and Godman et al. (2020), and to natural kinds in particular. On this view, natural kinds are property clusters which are explained by the presence of a super-explanatory property of the sort that feature within the natural sciences. Building on this framework, I distinguish reactive kinds and indifferent kinds. Reactive kinds are property clusters which are in part explained or impacted upon by our practices of classification, whereas indifferent kinds are those where this does not obtain.

2.1 Real kinds

A real kind is a category whose instances share a great many properties in common for some good, non-accidental reason (Millikan, 2000, 2017). Real kinds are sometimes called 'property clusters' for precisely this reason – they are characterised by many properties being co-instantiated across individual instances, such that the presence of *some* characteristic properties of the cluster in question increases the probability that the other properties will also be instantiated in a particular instance (see also Boyd, 1999). Knowledge of which sets of properties cluster in the case of a particular real kind allows us to make inferences, or at least educated guesses, about particular members of the kind. For example, as a matter of empirical generalisation, all instances of the kind *Equus caballus* share many properties in common; they have manes, they trot and gallop, they neigh, they have well-developed flight instincts and so on. Accordingly, if you observe that some particular organism trots and gallops and neighs, you may infer (not infallibly, but with reasonable reliability) that it is a member of the kind 'horse' and, in virtue of this, probably also has a mane and a well-developed flight instinct. One can easily see how inferences of this

kind would be useful to us humans. As Millikan puts it, we live in a 'clumpy world' and our brains have evolved to exploit this fact:

> The world of physical objects is to a large extent filled with clusters each having densely interlocked properties, clusters that are for the most part distinctly though not always perfectly separated from one another. This kind of structure is what underlies the success of ordinary everyday induction, knowing what to expect of yet another member of what one takes to be the same cluster. It allows one to know what might be expected from a cat or a truck or a piano or a cathedral.
> (Millikan p. 12 – 13, 2017)

Real kinds can be contrasted with 'nominal categories'. Nominal categories are categories whose instances are grouped together for conventional, arbitrary or anthropocentric reasons, and which fail to pick out real property clusters. Because instances of nominal categories fail to share a plurality of properties in common, they lack the potential for induction and generalisation which characterises real property clusters.[1]

It is no accident or coincidence that instances of the kind *Equus caballus* have many properties in common. Instances of the kind 'horse' share properties in common because they share a common origin (Godman et al., 2020; see also Khalidi, 2013).[2] Belonging to the same evolutionary lineage *explains* the fact that these properties tend to correlate. Godman et al. call these properties – that is, the non-accidental reasons which cause all the other properties to correlate – 'super-explanatory properties':

> When we have a Kind K whose instances share many different properties G, there will typically be some single property E of their instances that causally explains this multiple commonality. … For example, the atomic constitution of gold explains why all samples of solid gold have the same density, electrical and thermal conductivity, melting and boiling point, and so on. More generally, the molecular constitution of any given chemical substance will explain why its instances share many corresponding properties.
> (p. 319, Godman et al., 2020)

Where Millikan requires simply that the correlating properties of a real kind cluster for some non-accidental reason, Godman et al. narrow down and draw our attention to the peculiar position of super-explanatory properties in the metaphysical structure of real kinds. Many real kinds share properties in common *because* of a shared super-explanatory property which occupies a privileged causal-explanatory position relative to the other properties which are typical of the kind.

[1] For example, consider the category of 'round things', the category of 'vegetables' (parts of plants that are edible to humans, and which are not subsumed under the category 'fruit'), or the category of 'farm animals' (presumably spanning from chickens to Patterdale Terriers, with very little uniting the two).

[2] As argued persuasively by Godman et al. (2020) and Godman and Papineau (2020), however this is controversial (cf. Devitt, 2008, 2021).

On the account offered by Godman et al., super-explanatory properties need not be intrinsic, basic or microstructural. Indeed, the property which explains the cluster *Equus Caballus* – common origin – is relational and historical (Godman et al., 2020; see also Okasha, 2002). In this regard, super-explanatory properties are less philosophically demanding than traditional metaphysical essences (Khalidi, 2013).[3] For present purposes, we shall call the other properties associated with a given real kind – those which are caused by the super-explanatory property but are not themselves super-explanatory properties – 'secondary properties' of the kind in question. For example, if 'having atomic number 79' is the super-explanatory property of the chemical kind gold, then 'being malleable' and 'being highly heat conductive' are among its secondary properties.

Is every real kind characterised by the presence of a super-explanatory underlying property (even providing that they may be higher level or extrinsic)? Godman et al. suggest that we view super-explanatory properties as an illustration of the principle of common cause: if A and B correlate, then either A causes B, B causes A, or A and B have common cause C.

> In general, when we find that some A and B are correlated (in the sense that they are co-instantiated more often than we would expect given their separate probabilities of occurrence), then it will be the case that either A causes B, or B causes A, or both A and B are joint results of some common cause.
> (318 – 319, Godman et al., 2020)

Godman et al. entertain the possibility that some kinds will be characterised, not by a single super-explanatory property, but by a homeostatic feedback loop between the correlating properties of the kind – where A causes B, which causes C, which causes A. Perhaps, they hypothesise, some weather systems and psychiatric disorders fit this bill (Godman et al., 2020; Borsboom et al., 2019; Boyd, 1999). If so, the authors contend, these kinds do not count as being characterised by a super-explanatory property.

I shall take a slightly more permissive view on this point. Rather than rule out super-explanatory properties in such cases, I hold instead that where a kind K is maintained by some characteristic pattern of some mutually reinforcing properties A, B, and C, the *interaction* between A, B and C *is* the super-explanatory property of kind K. Indeed, the pattern of mutually reinforcing properties which characterises the kind may in turn be the cause of other correlating properties of K (such as properties D, E and F) which correlate with each other (and with A, B and C) precisely

[3] According to Godman et al. (2020), super-explanatory properties are special, not just in virtue of their particular causal explanatory relation to the secondary properties, but in that they are (in contrast to the secondary properties) metaphysically necessary. Godman et al. go on to hypothesise that the modal necessity of super-explanatory properties derives from the special causal relation which obtains between the super-explanatory property and the secondary properties of the kind. For my purposes here, it the special causal and definitional role of super-explanatory properties and the conversation this enables which are of importance, and I shall remain agnostic as to the modal issues.

because of the interaction between A, B and C. In these cases, at least, the interaction itself should count as super-explanatory.[4]

Consider, as an example of this effect, drug abuse or, as it is termed within the DSM-5, Substance Use Disorder (SUD). It is sometimes hypothesised that SUD is caused by a maladaptive conditioned feedback loop between the reward system, the anti-reward/stress system and the executive system in the brain (see Koob & Simon, 2009; Koob & Volkow, 2016). To which of these neural systems and sub-systems can we attribute the complex behavioural symptoms associated with Substance Use Disorder, such as an inability to attend to important life commitments or to stop using drugs in spite of serious adverse consequences?[5] It seems natural to say that these secondary behavioural properties of substance misuse are attributable precisely to the characteristic *interaction* between these altered systems in the brain, rather than to any one property in isolation. This interaction is, on my view, super-explanatory. One way to think of the metaphysics here is that there are four properties present when the interaction between A, B and C sustains a kind – (1) A, (2) B, (3) C and (4) their pattern of interaction.[6]

What motivates including feedback loops, or characteristic interactions, among the class of super-explanatory properties? As will become clear, I intend for this framework to apply to psychiatric kinds as well as somatic disease kinds, and many psychiatric kinds are hypothesised as being constituted by characteristic interactions or feedback loops, as Godman et al. (2020) themselves note. Accordingly, it is necessary to widen our notion of what sorts of processes can count as super-explanatory. Moreover, feedback loops may not be unique to psychiatry. For example, they have been implicated in systemic lupus and Alzheimer's disease (see Shlomchik et al., 2001; Doig, 2018). As such, this view of super-explanatory properties is more friendly to thinking about disease kinds in general.

Because super-explanatory properties occupy a special causal-explanatory position in the metaphysical structure of real kinds, they also play special roles in our definitions and classifications thereof. Chemical elements, for example, are classified according to their particular atomic constitution (such as an atomic number of 79), and not by their secondary properties like hardness, colour or melting point. Likewise, species and higher taxa in biology are defined by their evolutionary

[4] In my view, it is not *necessary* that an interaction has these sorts of effects for it to count as super-explanatory. It is enough that the interaction explains the correlation of the mutually reinforcing properties (and any effects they have individually). However, the interaction should *at the very least* count as super-explanatory when it has effects which are not straightforwardly reducible to the action of any of the constitutive properties. Thanks to an anonymous reviewer for prompting me to clarify my position in this regard.

[5] The DSM-5 (Diagnostic and Statistical Manual of Mental Disorders) recognises eleven criteria for the diagnosis of Substance Use Disorder, including items such as "Continuing to use, even when you know you have a physical or psychological problem that could have been caused or made worse by the substance" (American Psychiatric Association, 2013).

[6] I see no reason to deny this, other than an overly restrictive view of which features of a phenomenon rightly count as 'properties'. One might, for example, think of interactions between properties as complex properties made up of more basic elements (see Cunningham, 2001 for discussion). However, I won't commit to any particular view of properties here.

lineage. As we shall see, a similar principle also underlies medical theorising and classification.

The view of kinds I outline here differs more in emphasis than in substance from the 'simple causal view' advanced by Khalidi (2013) and Craver (2009). However, I agree with Godman et al. that "an undifferentiated appeal to causal structure misses the widespread significance of super-explanatory properties" (p. 321, 2020). Moreover, for reasons that will become clear, explicitly postulating properties which occupy a special causal explanatory position within the kind – rather than vague "causal relations" (Khalidi, p 81, 2013) – will be useful in the present context to enable certain distinctions I wish to make. Provided we are not too demanding about what sorts of properties may occupy this special position (if we permit that they may be interactions, and need not be basic or microstructural), this model can provide a useful template for thinking about real kinds in general and disease kinds in particular.

The picture of real kinds which emerges looks something like this. In Fig. 1, SP represents a super-explanatory property, while each S represents a secondary property, and each arrow represents a causal explanatory relation.

2.2 Natural kinds

Having defined real kinds as property clusters characterised by super-explanatory properties, we now turn to natural kinds. As I use the term here, natural kinds are a sub-set of real kinds defined by super-explanatory properties of the sort which feature within the natural sciences (broadly construed), for example common evolutionary lineage, shared atomic structure or a peculiar strain of fungal infection. Many contemporary philosophers, contrary to my usage in this papeer, use natural kind to mean real kind or property cluster. If so, copies of Alice in Wonderland – in that they form a real property cluster – would qualify as a natural kind.[7] This is in the end a terminological issue, and I have no substantive quarrel with this alternative usage of the term 'natural kind'.[8] The definition of a natural kind which I shall employ here is, in this sense, mostly operational.

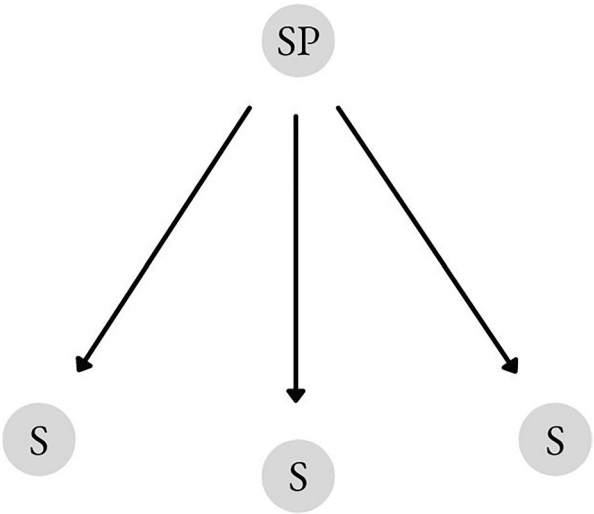
Within the broad category of natural kinds, we can draw more fine-grained distinctions according to the nature of the super-explanatory property characterising the kind in question, and the domain within which that property features as an object of study.[9] So, for example, gold is a chemical kind – not because the *kind* gold features

[7] A number of instances can share a great many properties in common, and as such constitute a very tight and 'clumpy' property cluster, for some reason that *is* super-explanatory, but that is not by any measure a 'natural' property. Groups of artefacts may share a great many properties because they descend from some original – for example all the world's copies of Alice in Wonderland (Millikan, 2017; Godman et al., 2020). Similarly, Godman has argued that specific gender categories – like 'Japanese Woman' – are culturally reproduced historical kinds, instances of which share properties due to cultural inheritance (Godman, 2018, Godman et al., 2020).

[8] Nor am I committed to the view that 'natural kinds' themselves constitute a natural kind (cf. Hacking, 2007).

[9] Thanks to Cecily Whiteley for helpful conversations that lead me to think of distinctions between kinds in this way.

Fig. 1 A real kind



as an object of study *only* in chemistry (clearly it features as an object of study in other domains as well – e.g., goldsmithing, dentistry and economics) – but because the super-explanatory property of gold (atomic number 79) is the sort of property which features in chemistry. Similarly, the kind *Equus caballus* constitutes a biological kind because (as hypothesised) the property which explains the presence of all the other properties associated with *Equus caballus* is a common evolutionary origin (which is the sort of extrinsic property which features as an object of study in biology). As we shall see, many disease kinds – the sorts of kinds that feature in medical nosology – will qualify as natural kinds *qua* biological kinds on this view.[10]

The picture of a natural kind which emerges looks something like Fig. 2, where N is a natural property, each S is a secondary property of N and each arrow (as before) represents a causal explanatory relation.

2.3 Indifferent kinds and reactive kinds

Having defined real kinds (*contra* nominal categories) and natural kinds (*contra* real kinds that do *not* count as natural, such as copies of Alice in Wonderland), we shall now make one further distinction. Sometimes, the nature of our theories, concepts and classifications impact upon the objective properties of members of the category under description, such that the nature of the properties associated with the kind change. In turn, these changes may feed back into our theories, concepts and classifications thereof. In these cases, our category is 'reactive'. Where there is no such

[10] Similar distinctions can be drawn within the category of real kinds, such as between social kinds (kinds which share some social property, such as a classification, profession or social role) and artefactual kinds (common intended function or deriving from some original design-template) and, still more fine-grained distinctions (see also Millikan, 2017).

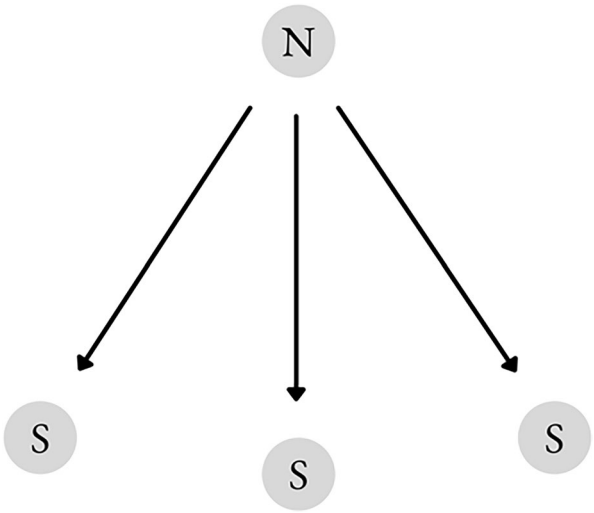


Fig. 2 A natural kind

relationship between the kind (or category) and its classification, our kind (or category) is 'indifferent'.

Hacking sometimes implies that awareness, or the possession of a self-concept, is necessary for a reactive effect between members of a kind and its classification to *count* as reactive (Hacking, 1999; see also Tekin, 2014). I take a much broader view of this. There seems no principled reason to deny that interactive effects which bypass members' conscious awareness of the classifications themselves, but act on them via other routes – such as through education, policy or the behaviours of other organisms – should not count as interactive (for a supporting argument, see Cooper, 2004; Khalidi, 2010 and Khalidi, 2013). Relatedly, I shall use 'classification' to mean the full social apparatus associated with a particular category, including official classification systems (such as the DSM-5 and the Periodic Table) but also legal constructions (refugee status), as well as more implicit social factors such as common biases, conceptions and socially accepted ways of thinking about the world.

Nominal categories like 'round things' and 'gemstones' are what we might call 'indifferent nominal categories'. Indifferent nominal categories are categories which do *not* pick out real kinds upon first being classified, and which are *not* subject to reactive effects in response to being classified. For example, the act of categorising aesthetically pleasing samples of minerals, rocks and organic matter together does not yield a feedback loop between the category 'gemstone' and the samples under description such that where before there was no real property-cluster, one eventually emerges. In this sense, the category of 'gemstone' is indifferent to our classification practices. Natural kinds too can be indifferent. Indeed, Hacking hypothesises this to be a distinguishing feature of genuinely natural kinds (Hacking, 1999; cf. Cooper, 2004). Consider again the natural kind 'gold'. Whether we classify, conceptualise and understand gold as 'a chemical element with atomic number 79' or a divine 'king of metals' with special celestial connections and healing powers is

of no relevance to the objective properties (density, melting point) caused by gold's atomic structure.[11]

Not all nominal categories persist in being nominal, however. Consider Hacking's famous case of Multiple Personality Disorder (MPD). If Hacking's story is to be believed, MPD started out as a nominal category, just like 'round things'.[12] There was no pre-existing real property cluster which MPD picked out upon first being coined. Instead, there were individual disparate instances of unhappiness and, perhaps, other types of mental ill-health, primarily among young females, upon which a name, a conception and a set of expectations were imposed. So far, MPD is like 'gemstone' and 'farm animal' – an 'arbitrary' grouping of individuals that reflects our interests, rather than a plurality of objective similarities across these instances in the world. However, those diagnosed with MPD did not react to the imposition of this arbitrary label in the same way as would a gemstone or a farm animal.

> Around 1970 there arose a few sensational paradigm cases of strange behaviour similar to phenomena discussed a century earlier and largely forgotten. A few psychiatrists began to diagnose multiple personality. It was rather sensational. More and more unhappy people started manifesting these symptoms. At first they had the symptoms they were expected to have. But then they became more and more bizarre. First a person had two or three personalities. Within a decade the mean number was seventeen. This fed back into the diagnoses, and entered the standard set of symptoms. It became part of the therapy to elicit more and more alters. The psychiatrists cast around for causes, and created a primitive, easily understood pseudo-Freudian aetiology of early sexual abuse, coupled with repressed memories. Knowing this was the cause, the patients obligingly retrieved the memories. … In 1983 I confidently said that there could never be split bars, analogous to gay bars. In 1991 I went to my first split bar.
> (Hacking, p. 296, 2007; see also Hacking, 1998)

A label is allocated to a set of individuals. A standard of behaviours, feelings, expectations, self-perceptions and narratives is imposed on those labelled, and put to work in our institutions, our media and culture, our social and economic systems, and our systems for knowledge production. Those classified then adjust accordingly.

Note that Hacking's suggestion is not (or not primarily) that those classified adjust to labels imposed upon them by *pretending* to display the objective properties associated with the label – in this case, MPD. The suggestion, rather, is that the individuals thus classified come to *actually possess* some properties, such as behaviours, feelings, dispositions etc., in virtue of, or in response to, the imposition of a classification. Changes to the properties of the subjects being classified then feed back into classification systems, which in turn re-inform the experience of the individuals

[11] As the alchemists would have it. Kauffman cites Arnold of Villanova: "[I]t harbours specific virtues which are due to celestial influence. In its stability and permanence, gold is itself like a star of heaven." (p. 74, Kauffman, 1985).

[12] There is naturally some controversy as to whether this story *should* be believed. I shall just assume Hacking's account for present purposes. See Hacking (1998) for discussion.

being classified. This is Hacking's famous looping effect: "[O]ur investigations interact with the targets themselves, and change them. And since they are changed, they are not quite the same kind of people as before" (Hacking, p. 293, 2007). One might worry, therefore, that a kind which is subject to these sorts of effects is an unstable object of scientific knowledge, changing under our feet, or, as Hacking suggests, a 'moving target' (2007). As we shall see, not all types of reactivity are equal threats in this regard.

As a consequence of this sort of reactive process, an originally nominal category can cease to be nominal and become real – what I shall term a 'nominal reactive kind'. Even though these classifications did not track a real property cluster upon first being coined, through reactivity, individuals classified in this way may come to actually share properties in common, and thus the category may come to support the sorts of inductions and generalisations associated with real kinds. Where before there was nothing real to track, there is now a group of people who genuinely share characteristics – feelings, behaviours, dispositions, experiences and, indeed, memories. In other words, members of a nominal reactive kind do share a plurality of properties, and they share a plurality of properties *because* they share a super-explanatory property – they share the relational, extrinsic property of being classified in a particular way by their social and cultural collective. In other words, a classification (C) can, in the right circumstances, qualify as a super-explanatory property.

In Fig. 3, C represents a super-explanatory classificatory property and, as before, each S a secondary property, and each arrow a casual-explanatory relation.

Do nominal reactive kinds – i.e. real kinds that are real *only* in virtue of our classifications – count as natural kinds? This matter has been disputed in the literature. If you take constituting a natural kind to be nothing over and above constituting a real kind, and you take nominal reactive kinds to (sometimes at least) constitute real kinds, then nominal reactive kinds can count as natural kinds (see Cooper, 2004). According to

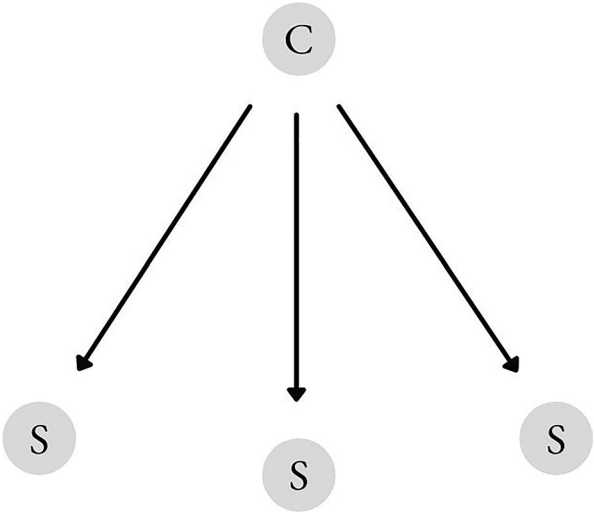


Fig. 3 A reactive, classificatory kind

the framework I have outlined here, however, they do not. What qualifies a real kind as a *natural* kind is the nature of its super-explanatory property, not its secondary properties. As such, even if different instances of a nominal reactive kind share objective (even biological) properties in common, they do not qualify as natural kinds, because the super-explanatory property they share – being classified in a particular way by the social collective – is not the sort of property which features in the natural sciences. They may, however, qualify as a real social and classificatory kind – that is, a real kind which is explained by the presence of a super-explanatory social classification.

3 Disease kinds

Having distinguished real kinds from nominal kinds, natural kinds from real kinds, reactive kinds from indifferent kinds, we now turn to disease kinds. What are disease kinds like? Disease kinds are not generally mere nominal categories (non-kinds) or nominal reactive kinds (real kinds, but mere products of our classification systems). Nor are disease kinds, generally speaking, indifferent natural kinds. Disease kinds, as I shall argue, seem *prima facie* to have both natural and reactive features. In Sect. 3.1., I argue that medical nosology, often enough, picks out real kinds, and that many of these will satisfy the conditions for being natural kinds. In Sect. 3.2. I shall argue that, in addition to satisfying the conditions for being natural kinds, many disease kinds also have a reactive element in that their properties are in part explained, or impacted upon, by our systems of classification.

3.1 Disease kinds as natural kinds

According to the framework outlined in earlier sections, natural kinds are real property clusters characterised by super-explanatory properties of the sort which feature within the natural sciences. Biological kinds, I suggested, constitute a subcategory within the broader category of natural kinds, characterised by super-explanatory biological properties. I also outlined a permissive notion of super-explanatory properties, according to which characteristic *interactions* between properties can count as super-explanatory. Do disease kinds fit this picture? We can split this question into three: (1) do medical kinds pick out *real* property clusters, (2) are these real property clusters characterised by super-explanatory properties, and (3) are the super-explanatory properties 'natural'?

Real kinds, as we have defined them, following Godman et al., Millikan, Boyd and others in this area, are categories the instances of which share a great many properties in common – in other words, real kinds are property clusters. Assuming medical science and nosology are not severely off track, instances which are classed together in medicine do often share objective, real, projectible similarities in common.[13] These similarities between instances, in turn, inform useful scientific

[13] There is some reasonable concern, at least in psychiatry, that medical nosology really *is* off track, but this problem is a relatively peculiar to psychiatry and its ongoing so called 'crisis of classification' (see e.g. Poland & Tekin, 2017).

generalisations and inductive inferences between cases. Real similarities between specific cases of disease allow us to infer (not infallibly, but reasonably reliably) from fatigue, hyperglycaemia and unexplained weight loss that the patient probably has diabetes and would respond to insulin. To put this differently, the success of medicine as a scientific endeavour is built upon the assumption of projectible similarities between instances of disease kinds. If we did not expect there to be real similarities between one person with testicular cancer and another, what would motivate giving a sample of them an experimental drug in a clinical trial (and generalising from here to the wider treatment population)? The existence of real property clusters in medicine is what underlies medical knowledge.[14]

Many medical kinds, then, are real property clusters. But are they characterised by super-explanatory properties? It is interesting to note at this stage that medicine is premised upon an etiology/symptomology distinction which closely mirrors that between super-explanatory and secondary properties. For example, people with scurvy tend to share the symptoms (or secondary properties) of scurvy – anaemia, bleeding, myalgia, gum disease – across instances *because* they share the underlying super-explanatory property of deficient levels of vitamin C. Similarly, Graves' Disease is an autoimmune disease caused by the immune system producing antibodies that attack cells in the thyroid gland. Because people with Graves' Disease share this characteristic dysfunctional immune response, they tend to share a number of other secondary properties in common as well, such as weight loss, rapid heartbeat, fatigue and trouble tolerating heat. Or consider infectious diseases. The symptoms of tuberculosis – such as fever, cough, loss of appetite – are all *caused* by mycobacterium tuberculosis infecting the body's cells. In this sense, the infection explains the other properties associated with the kind, and thus counts as super-explanatory.

The etiology/symptomology distinction is also reflected in medical nosology, where systems which individuate diseases according to etiology – that is, the super-explanatory cause of the disease – are considered the gold standard. Indeed, where a particular symptom-cluster (say, fever) is later discovered in fact to be causally accounted for by a number of distinct underlying disease processes (tuberculosis, heatstroke, inflammation of the joints etc.), this usually results in calls for its re-classification into more fine-grained disease-types which better reflect the syndrome's actual causal underpinnings in each kind of case.[15] Recall that we are employing a more permissive notion of the kinds of properties which may count as super-explanatory, spanning characteristic interactions and higher-level properties, and so may count a wider range of etiologies (or specific domino dysfunctions) as super-explanatory properties. Given these stipulations, the distinction between super-explanatory and secondary properties seems an apt model for medical kinds in general.

[14] That is not to say that *all* medical kinds are real kinds. Some may turn out to be spurious or overly heterogenous.

[15] Of course, the distinction between (super-explanatory) etiology and (secondary) symptomology is salient in medicine for clinical as well as theoretical reasons – ideally, we would intervene medically on the property which is *causing* all the other properties (the super-explanatory pathoetiology) so as to alleviate *all* the problematic symptomology with a single 'silver bullet' cure, rather than treat each disparate symptom individually in a therapeutic game of whack-a-mole.

So medical kinds are, often enough, real kinds characterised by shared super-explanatory properties. But are these properties natural? We have defined natural properties as the sorts of properties which feature within the natural, including the biological, sciences. Thus, if a disease kind is characterised by a particular underlying biological property, state or process – such as a disordered neural circuit, peculiar bacterial infection or uncontrolled cell division in the breast tissues – it would qualify as a natural kind. And indeed, the usual case of a disease kind is just such a case.[16]

3.2 Disease kinds as reactive kinds

As we have seen, many disease kinds qualify as natural kinds in virtue of being real property clusters which are caused by super-explanatory biological properties. However, unlike most other natural kinds – gold, *Equus caballus* – disease kinds are also divisions between *people*, and as such form an interesting class. Other than disease kinds, there are very few natural divisions *within* the category of human beings. *Homo sapiens* is of course a natural kind, but few distinctions within our species form genuine natural, biological kinds of their own (with the possible exception of biological sex).[17] The intra-human property clusters which are important tend mostly to be explained, not by super-explanatory natural properties, but rather by cultural, legal and economic factors – teachers, Palestinians, fathers, members of the middle-class, and so on. In that natural disease kinds also map onto human kinds, they have within them enhanced potential to enter into interactive relations with their classifications – just as other human kinds do.[18]

The claim I shall go on to defend is not that *every* natural disease kind has an element of reactivity. I do not think we are in a position to make any blanket claims here and, in any case, it is not important for my purposes to do so. Instead, I shall offer four cases – ranging from the biomedical, to the neurological, to the microbial – which, taken as a whole, motivate the position that medical kinds often have a reactive element.

3.2.1 Breast cancer

Let us first consider breast cancer. Those who develop breast cancer share many properties in common due to a particular shared pathophysiological process and

[16] Are there exceptions to this general rule? An obvious candidate would be MPD or another such transient mental illness where the characteristic symptomology is caused by an extrinsic social property such as a classification. However, whether these cases should count as *true* disease kinds is plausibly up for debate. We might also wonder about cases where characteristic symptomology appears to be caused by an extrinsic past event such as exposure to trauma (as in traumatic head injury or post-traumatic stress disorder) or past infection (as in the case of 'Long Covid'). In these cases, there will usually be some intermediate property (say, damage to the neural tissues or inflammation) causing the symptomology in a proximal sense. However, nothing about the framework I have outlined here prohibits extrinsically constituted biological kinds from constituting natural kinds, so I need not rule these cases out entirely.

[17] For discussion, see Khalidi (2020).

[18] That is not to say *unique* potential, as we shall see.

its secondary properties. Uncontrolled cell-division in the breast tissues (the super-explanatory property) causes lumps in the breast or lymph nodes, changes to the structure of the skin, changes to the size and shape of the breast, and eventually more serious symptomology, including mortality. In other words, breast cancer is a natural kind (per our permissive view thereof).[19] However, this is not the only reason why people with breast cancer share properties in common. People with breast cancer also share properties as a consequence of being classified, conceptualised and treated in a particular way within a particular social context. As is particularly often the case with 'women's diseases', a diagnosis of breast cancer carries particular social meanings and cultural significance – in some respects, it is a category 'on the move'. Discussing the emergence of the category as an identity in the early 20th Century, Klawiter writes:

> In addition to the stigmas that adhered to cancer in general, the stigma of breast cancer in particular … was intensified and inflected by the cultural power of women's breasts. But although gender, heterofemininity, and the cultural significance of women's breasts were deeply implicated in the stigma of this disease, the stigma itself circulated in the form of discourses that were not "carried" by specific individuals or particular subsets of the female population. With rare exception, no one publicly identified as a woman with breast cancer, as an ex-breast cancer patient, or even as a breast cancer victim. No one claimed these [identities] or had them thrust upon her. For the most part, women with breast cancer histories "passed" as normal women and, as a social formation, breast cancer-related identities did not yet exist. … Whereas people with AIDS and HIV were publicly 'outed' and disparaged, women with breast cancer were publicly 'closeted' and pitied.
> (p 8 - 9, Klawiter, 1999)

Contrast Klawiter's account of being a person with breast cancer in the early 20th Century with how this label features in our current cultural milieu. Since the Pink Ribbon movement, breast cancer awareness month, cancer screening programmes, breast cancer survivor networks and a multitude of corporate campaigns and sponsorships, the properties associated with instantiations of breast cancer have changed. Breast cancer has taken on a new form, indeed perhaps emerged as a "way of being as person" (p. 303, Hacking, 2007). Women with breast cancer share properties – identities, narratives, stigma, experiences – due to being classified as people with breast cancer. Being thus classified is what explains why all these properties are co-instantiated – it is, in this sense, a super-explanatory factor.[20]

Nor are the natural facts and the classificatory facts entirely causally unrelated here. Stigma and 'closeting' of those affected by the disease may contribute to

[19] See Plutynski (2018) for an argument that cancer is *not* a natural kind. Plutynski appears to assume a rather more reductive and traditionally essentialist position than the one I am proposing here – for example, she entertains the possibility that what defines cancer cells is "their distinctive interactions with neighboring cells" (p. 43, 2018) which would appear to be compatible with my account (if not with hers). In any case, a rebuttal is beyond my scope here.

[20] The case of HIV, to which Klawiter alludes, may constitute an analogous case of reactivity.

impoverished knowledge of the condition and a lack of research investment, and deter treatment seeking behaviour. This in turn has implications for the mortality rate of the disease, as well as the stages of pathology the disease is permitted to reach prior to intervention. Our classification – through our knowledge, our experts, our institutions and our conceptions – interacts with the pathophysiology of members of the kind to produce certain outcomes. These effects in turn feed back into what it means to have breast cancer – is it a relatively treatable disease from which you may emerge a noble 'survivor', or a shameful death sentence?

3.2.2 Autism

Let us move on to consider a rather different case, one that has featured prominently in Hacking's later work. High-functioning autism, Hacking contends, predates our classification thereof: "[I]f, as is widely supposed, autism is a congenital neurological deficit, then there were certainly autistic children who were dismissed as retarded, feeble-mind, and so on, a long previous litany of dismissive epithets." (p 304, 2007). In other words, high-functioning autism existed as a real natural (biological) kind prior to our idea of it as such.[21] In our terminology then, unlike MPD, autism is not a nominal reactive kind, a mere consequence of our classification practices; it would have been real regardless.

However, as Hacking goes on to argue, even if high-functioning autism always existed as a mind-independent natural kind – a neurodevelopmental disorder – it did not always exist as a 'way to be a person':

> Before 1950, maybe even before 1975, high-functioning autism was not a way to be a person. There probably were a few individuals who were regarded as retarded and worse, who recovered, retaining the kinds of foibles that high-functioning autistic people have today. But people did not experience themselves in this way, they did not interact with their friends, their families, their employers, their counsellors, in the way they do now. Later this did become a way to be a person, to experience oneself, to live in society … This was a looping effect: a few of those diagnosed with autism developed in such a way as to change the very concept of autism. They brought into being the idea of a high-functioning autistic person.
> (p. 303 – 304, Hacking, 2007)

Hacking argues that, even if there were people with the congenital disorder of autism prior to the coining of 'high-functioning autism' as a construct, there was no one who was a member of the particular social kind 'high-functioning autism'.[22] The properties associated with that kind are in part explained and impacted upon by

[21] Whether autism constitutes a real, uniform kind – or even a disease – is controversial both philosophically and scientifically (Chapman, 2020). I do not have any particular commitments here.

[22] What is the super-explanatory property in this case? We do not know exactly, because there is empirical uncertainty as to what causes autism and, indeed, whether there is a uniform cause (Walsh et al., 2011). The assumption in Hacking's account, however, appears to be that there is some particular neural dysfunction ('a congenital neurological deficit') which causes the symptoms associated with autism.

our peculiar systems of classification – in this case the classification 'high-functioning autism'. As such, before that classification existed, nor did the cluster of properties caused by its reactive effects.[23]

3.2.3 Ischemic stroke

Reactivity in the medical realm can also come in the form of self-fulfilling prophecies, aided by placebo and related effects (the mechanisms of which remain largely mysterious) (Benedetti, 2020; Cavanna et al., 2007). Think of ischemic stroke. Until relatively recently the received view was that the brain was essentially 'set' after adolescence, and that if something broke it was therefore doomed to stay broken:

> Just 50 years ago, the idea that the adult brain could change in any way was heretical. Researchers accepted that the adolescent brain was malleable, but also believed that it gradually hardens, like clay poured into a mould, and, therefore, that any damage or injuries it sustains cannot be fixed.
> (p. 2, Costandi, 2016)

However, this consensus had been turned on its head in the past few decades, with increasing evidence that the brain possesses a remarkable neuroplastic ability to reorganize itself in response to injury and novel environmental demands. This finding is also informing our theories of brain damage. Where damage to neural tissues was previously thought to be permanent and immutable, it is now recognised that the plasticity of the brain can be harnessed to regain function and compensate for impairments.[24] By encouraging those affected to think of their brains as capable of healing themselves, over time, with the aid of therapeutic exercises, cognitive therapies and increased activity levels, perhaps patients may come to have more of this capacity as well. As noted by researchers, "[M]otivation and attention can be critical modulators of plasticity" (Cramer et al., p. 1603, 2011). As such, patients who see their brains as able to heal, rather than permanently and immutably broken, may be more likely to heal – which in turn informs our understanding of the nature of stroke and its pattern of recovery.[25]

[23] It should be noted that Hacking does not employ the language of 'social, reactive kind' versus 'natural kind', but I think it is reasonably to infer that this is the metaphysics which would make sense of his insistence that high functioning autistic people did, and yet did not, exist prior to classification.

[24] I assume, here, that damage to the neural tissues is the super-explanatory property. Of course, the damage itself is, in a more distal sense, caused by vascular dysfunction. However, the original disruption to the blood supply is no longer causative of the symptoms after it is dissolved. It is the damage caused by the disruption to the oxygen supply which remains, and which causes the symptoms.

[25] This is at least a plausible hypothesis. In addition to motivation, neuroplasticity in stroke recovery is modulated by factors such as depression and degree of social engagement, indicating that belief in one's capacity for recovery could make a difference via these metrics (Chaturvedi et al., 2020; Tracy et al., 2014). Stinear also notes that a stroke patient's perceived poor prognosis may risk "the outlook becoming a self-fulfilling prophecy" (p. 1230) by lowering investment in therapeutic measures (2010). Moreover, it has also been argued that, in learning contexts, positive narratives may stimulate neuroplasticity, whilst negative thinking may increase stress and decrease plasticity in the brain (Cozolino & Sprokay, 2006).

3.2.4 Covid-19

Finally, let us consider a case which is *prima facie* rather different. It has been pointed out by various participants in the philosophical debate surrounding reactivity and its relationship to natural kinds that some pathogens, like bacteria or viruses, may in the right circumstances count as reactive. Hacking writes:

> Microbes, not individually but as a class, may well interact with the way in which we intervene in the life of microbes. We try to kill bad microbes with penicillin derivatives. We cultivate good ones such as the acidophilus and bifidus we grow to make yogurt. In evolutionary terms, it is very good for these benevolent organisms that we like yogurt, and cultivate them. But some of the malevolent organisms ones do pretty well too. Disease microbes that we try to kill may as a class, a species, respond to our murderous onslaught.
> (Hacking, p. 106, 1999)

That is, our theories and conceptions and the behavioural, scientific and political implications of our theories and conceptions – broadly, our classifications – can under the right conditions act as selection pressures influencing the evolution of the pathogen, such that the properties of the disease-kind are altered.

Perhaps we are living through one such process right now. Sars-Cov-2, the novel Coronavirus, was discovered in Wuhan in the latter months of 2019, and was quickly identified as a major threat to human life, health and societal functioning. In response, countries locked down and instituted control measures intended to reduce transmission. New strains of the virus emerged that were better able to evade our measures by being more transmissible. It is possible that, in a similar vein to Hacking's microbes, our classifications are acting as selection pressures in the evolution of the virus such that Covid-19 – this disease kind – is altering in response to our classifications (see Khalidi, 2013).[26] We have since produced vaccines against the virus and rolled them out on a large scale. How might Sars-Cov-2 evolve next?

4 Reactive natural kinds and varieties of dependence

As argued in the previous section, we have reason to believe that many disease kinds satisfy conditions for being natural kinds whilst being, to some extent, reactive in response to our classification practices. This prompts the question: under what conditions is a natural disease kind really a *reactive natural biological kind*, and when does it simply co-occur or intersect with a corresponding social, reactive kind?

[26] The super-explanatory property of the disease Covid 19 is infection with the Sars-Cov-2 virus. The secondary properties are the symptoms of Covid-19 (e.g. loss of the sense of taste and smell). This case is analogous to the tuberculosis case considered in Sect. 3.1.

In what follows, I shall argue that we can provide a principled answer to this question which is conditional on the variety of dependence relation which holds between the natural kind and its classification. Where the peculiar pattern of correlating properties within a natural medical kind is partly explained by, or impacted upon, by our practices of classification, there are three varieties of dependence which may obtain. Each has different implications for the metaphysics of the kind in question. In cases of simple dependence, there is simply a natural kind and an associated classificatory kind (assuming that the classification yields a social kind) and no real reactivity. In cases of secondary dependence, there is a natural kind and an intersecting classificatory kind, and thus a limited form of reactivity, but not the kind of reactivity which stands to render the natural kind metaphysically unstable and 'a moving target' of scientific inquiry. In cases of essential dependence, the natural kind in question really is an unstable reactive natural kind.[27]

4.1 Simple dependence

Simple dependence is so called because it is ubiquitous and represents no real puzzle. According to my definition, simple dependence obtains between a natural kind and its classification when the classification depends upon the properties of the natural kind. For example, our classification of gold reflects or depends upon the underlying super-explanatory property of the chemical kind gold – i.e. 'having an atomic number of 79'. This is the sort of dependence relation which obtains in successful science.

In these cases, precisely because our classification (C) reflects the nature of the natural kind (N), it is also the case that if some particular instance as a matter of fact *is* N then that is going to increase the probability of that instance being labelled as C. For example, the fact that the cufflinks I've just handed to a jeweller *are* made of gold – that is, that they constitute a sample of an element with atomic number 79 – is going to increase the probability that the label 'gold' will be applied to them (assuming the jeweller is any good at her job). As such, if the classification C has some secondary effects, then the properties caused by the natural kind (the properties S springing from N, in Fig. 4) and the properties caused by our classification thereof (the properties S caused by C) will *still correlate* – even if there is no reactivity to speak of between the natural kind and our classification thereof. For example, suppose our classification of gold includes some culturally contingent symbolism around marriage and romance, and that this causes samples of gold to be frequently formed into rings. If so, then the secondary properties of the natural kind gold (e.g. having a melting point of 1064 °C) will correlate with the secondary properties caused by our classification

[27] Although I have focused in this paper on disease kinds, I take this framework to – in principle – have broader application.

of gold (e.g. being moulded into rings) despite there being no interaction of note. As such, the secondary properties of C and N will still *cluster*, even if there is no real reactivity at play.

Some property clusters in medicine likely constitute just such cases of associations between natural kinds and classificatory kinds, where the nature of the classification depends upon the nature of the kind in a simple sense. A number of properties correlate across instances, some of which are products of our classifications (that is, caused by C) and some of which are products of the underlying natural disease kind (caused by N). In cases that fit this pattern, simple dependence obtains, and the natural kind is not *really* reactive (in that no properties of the natural kind are changing in response to our classifications) – the natural kind is merely *associated* with a social phenomenon.

Determining which real-world disease kinds fit this pattern will, in practice, be very complicated but, by way of illustration, let us consider a hypothetical example. Suppose a natural disease kind – let's call it N1 – causes weakness in the arms and legs and bruising of the skin. Through scientific investigation, we discover that N1 is a particular sexually transmitted pathogen, the presence of which can be established by a simple laboratory test. The classification of N1 – let's call it C1 – now encompasses this knowledge. Due to N1's association with sex, people who are classified as having N1 (that is, that have the diagnostic label C1 applied to them) also tend to giggle and be embarrassed. Because C1 reflects N1, those persons that *actually* instantiate N1 now have an increased likelihood of being classified as such, that is, of having C1 applied to them. Because C1 causes embarrassment and giggling, those identified as infected with N1 now instantiate those properties, in addition to weakness in the arms and legs and bruising of the skin. Thus, the properties of the natural kind and the properties of the classificatory kind come to correlate – and as such (in a sense) form a cluster of properties – without any reactivity in the natural kind. N1 is just correlated with C1.

Simple dependence is represented below. As before, N is a super-explanatory natural property, C is a super-explanatory classificatory property, and the downward arrows represent causal relations between the super-explanatory properties (N, C) and their secondary properties (S). The arrow leading from N to C represents the dependence relation between the nature of the natural kind and the nature of our classification thereof. Assuming that C is super-explanatory, and itself the super-explanatory property at the heart of a classificatory, social kind (in the sense outlined in previous sections) what we now have is a natural kind which is associated with a classificatory, reactive kind.[28] Their properties will correlate for this reason, but that does not mean that the natural kind itself is reactive.

[28] It is not clear that C will *always* be super-explanatory in the sense of causing many other secondary properties to be instantiated. If so, it is just the property C which will be correlated with N and the other secondary properties of N.

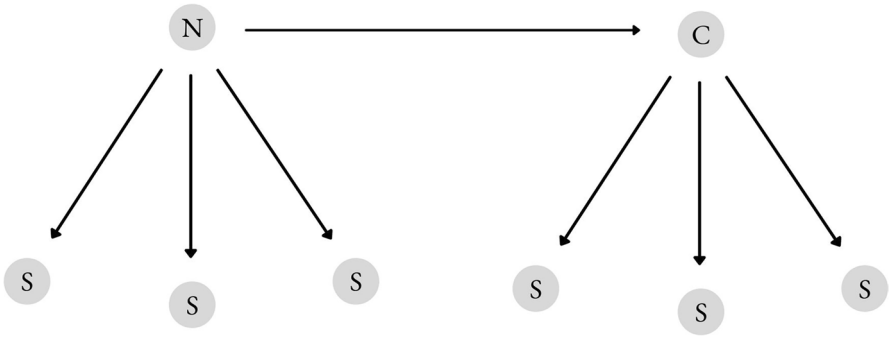


Fig. 4 Simple dependence between a natural kind and a classificatory kind

4.2 Secondary dependence

In *The Social Construction of What?* and elsewhere, Hacking rather confusingly considers the possibility that some particular kind may count *both* as an indifferent kind *and* as an interactive kind. Given that indifferent kinds are defined, by myself but also by Hacking, in opposition to reactive kinds – "All I want is a contrast to interactive kinds. Indifferent will do." (p. 105, Hacking, 1999) – Hacking's contentions in this regard have caused some understandable bewilderment in the literature: "Given Hacking's manner of defining interactive kinds and indifferent kinds as 'classifications that affect their objects of study' and 'classifications that do not affect their objects of study', respectively, he is not entitled to maintain that a classification such as autism can be both interactive and indifferent." (Tsou, p. 334, 2007). However, as we shall see, the phenomenon of secondary dependence, which I shall outline in this section, makes good sense of the manner in which a particular (in this case, disease) kind may simultaneously both interact with its classification and not.

Secondary dependence obtains where the classification (C) impacts upon secondary properties (S) which are *also* caused by the kind's super-explanatory natural property (N). Let us return to the example of breast cancer. Suppose that there is a great deal of stigma associated with breast cancer, and that this impacts upon the extent and timing of treatment seeking behaviour which, in turn, impacts upon the mortality of the disease. Of course, the mortality of the disease is not solely a product of our classifications. Mortality rates for breast cancer are also products of the pathophysiology of the disease itself – that is, uncontrolled cell division in the breast tissues (N). In this sense, the secondary properties of breast cancer (S) depend *both* upon the super-explanatory natural property of the disease kind (N) and on our practices of classification (C). In this sense, breast cancer is a case of secondary dependence.

We are now in a position to provide a clear solution to Hacking's puzzle. How can a disease kind be *both* indifferent and reactive? The answer is that the secondary properties of the disease kind may be responsive to our classifications, but without any reactivity in the super-explanatory natural property which characterises the natural kind. Some secondary properties of breast cancer may be reactive to our classifications, but the super-explanatory property – the property which characterises the kind and

occupies a special, causal-explanatory role in the metaphysical structure of the kind – is not. As such, there is a sense in which breast cancer is both indifferent and reactive. The super-explanatory natural property is indifferent, but some secondary properties of the natural kind in questions *are* reactive in response to our classifications.[29]

So, assuming C is a super-explanatory property of a classificatory kind, under conditions of secondary dependence we have a natural kind which intersects with a social kind to produce certain outcomes which are products *both* of the natural kind and the classificatory kind (see Fig. 5). In this sense, the properties associated with the disease will not (as in the case of simple dependence) be entirely reducible to the separate contributions of the natural and the social kind – some properties are influenced by both. As such, there is a sense in which the natural kind N is reactive, but it is a limited form of reactivity which does not threaten its stability in response to our classification.

4.3 Essential dependence

Essential dependence between a natural kind and its classification occurs where the super-explanatory natural property at the heart of the natural disease kind is responsive to our classifications (Fig. 6). In other words, a natural kind is essentially dependent if the property (N) of the relevant disease kind is causally impacted by our practices of classification (C). Let us consider again the novel coronavirus. Sars-Cov-2 is the virus which is responsible for the respiratory disease Covid-19. As I have argued above, it is possible that our classification of Sars-Cov-2 as a major threat to human health and well-being, and the resultant policy measures – are impacting upon the evolution of the virus. As such, the super-explanatory natural property of the disease kind Covid-19 *is* responsive to our classifications. In this sense, Covid-19 is essentially dependent.

My claim is that conditions of essential dependence are the *only* conditions under which a natural disease kind is *truly* reactive in an important sense, with implications for the stability of the kind and our classification thereof. If the super-explanatory natural property underlying the kind changes in response to our classification practices, then there is a fundamental instability at the heart of the natural disease kind in question.[30]

[29] In a critique of Hacking, Tsou makes a distinction between 'strong' and 'weak' implications of looping effects which bears some relation to my distinction between secondary and essential dependence. The strong implication, according to Tsou, is that the looping effect causes the definitional criteria of a classification to change, where 'definitional criteria' are cashed out in terms of law-like biological regularities. The weak implication is that people's behaviours simply change in response to their classification. Although I am sympathetic to Tsou's line of argument here, I take my own to offer a higher degree of precision. Firstly, it is not clear that 'definitional criteria' are what Tsou is really after – recall that the formal definitional criteria for psychiatric disorders (within ICD and DSM) are generally behavioural rather than biological. Moreover, Tsou makes the curious assumption that biological regularities *ipso facto* cannot be mutable in response to looping effects: "To establish [that objects of study in psychiatry are unstable], Hacking would need to show that *the typical biological or physiological process that leads to abnormal behaviour is changed because of looping effects*. There is no good evidence for thinking that this is a possible consequence of looping effects." (Tsou, 2007, emphasis original; see also Tsou, 2013). As we shall see, this inference does not hold.

[30] I have in mind here direct changes to the nature of the super-explanatory natural property in question, not merely changes to its *frequency* of instantiation in the population (which would not have implications for the stability of our classifications).

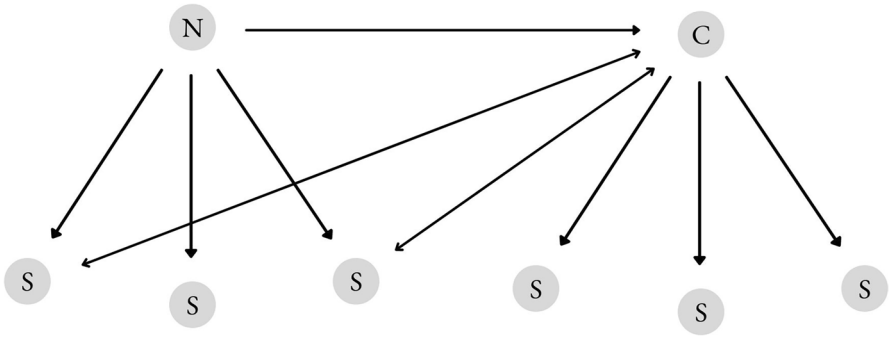


Fig. 5 Secondary dependence between a natural kind and a classificatory kind

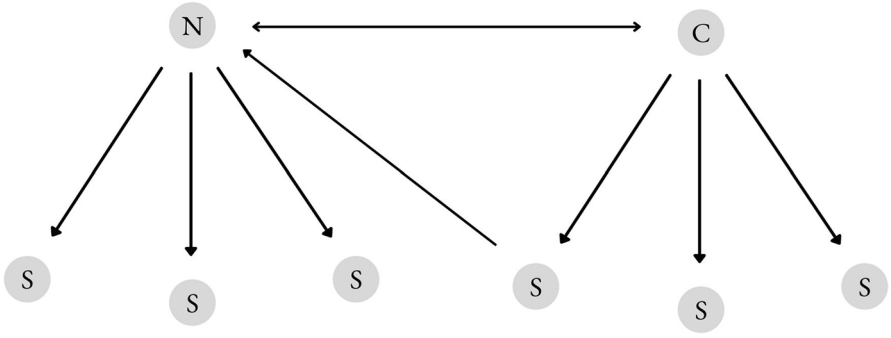


Fig. 6 Essential dependence between a natural kind and a classificatory kind

The special super-explanatory property of the natural kind in question – that which is explanatory and definitional – really *does* change in response to our classification such that, once classified in a particular way, the phenomenon of interest really *is not what it once was* prior to classification, in the manner Hacking suggests at his more radical: "[S]ince they are changed, they are not quite the same kind … as before. The target has moved." (p. 293, 2007). In what follows, I shall offer a hypothesis as to when essential dependence may obtain, and why this may occur more often in neuropsychiatric disorders.

5 Essential dependence, adaptation and neuropsychiatry: a hypothesis

Tsou has argued, *contra* Hacking, that neuropsychiatric kinds (like depression, schizophrenia and even suicide) are not as unstable as Hacking suggests, because they are in fact characterised by 'stable' neurobiological regularities across instances (Tsou, 2007, 2013; see also 2021). And yet neuropsychiatric kinds of various sorts are often cited, by Hacking and others in this literature,

as paradigm cases of reactivity. Psychiatry contrasts with biomedicine in this respect as biomedical kinds – arthritis, myocarditis or sickle cell anaemia – are rarely invoked as examples of reactivity (although, as I have argued, in the limited sense of secondary dependence, they often are). How to explain this apparent disanalogy? In what follows, *contra* Tsou, I shall offer a hypothesis as to why, even if they are underpinned by biological realities, kinds in psychiatry (and possibly to a lesser extent in neurology) may be more susceptible to essential dependence and thus to instability. This view is motivated not by some spooky dualism about the mental, but by a very real biological disanalogy between the brain and the rest of the physical body – the relative plasticity of the brain.

Under what conditions can essential dependence obtain? Essential dependence can obtain where the super-explanatory natural property is susceptible, via some discernible mechanism, to directed change, en masse, in response to our classification practices. For example, the virus which underlies Covid-19, Sars-Cov-2, is literally evolving and so, when our classification practices act as selection pressures in its evolution, the super-explanatory property at the heart of Covid-19 reacts to our classification. As such, it is an unstable and changeable kind.[31] However, this change in no way threatens its status as a *natural* kind – after all, it is still a real kind explained by the presence of a super-explanatory natural property. It's just that that this super-explanatory property is susceptible to *adaptation.* In the same vein, other diseases that are evolving – such as those caused by bacteria, fungi and viruses – may be similarly reactive under the right social conditions.

The primacy of adaptation to essential dependence points to an interesting implication. It may be that neuropsychiatric kinds characterised by dysfunctions in complex or higher-level neural traits are susceptible to essential dependence as well, because they too are susceptible to adaptation.[32] The reductionist programme in psychiatry – which optimistically hypothesised that psychiatric disorders would turn out be neatly reducible to simple, basic pathologies (as was the case for, say, neurosyphilis) – has largely given way to a programme according to which, even if psychiatric disorders *are* brain disorders, many of them are likely to be disruptions in complex neural functions or, as the National Institute of Mental Health's Research Domain Criteria project would have it: psychiatric illnesses are "neural circuit disorders" (p. 499, Insel & Cuthbert, 2015).

Complex higher level neural processes differ from functions in the rest of the body in some interesting respects. In particular, relative to functions of the biological body beyond the brain (and perhaps some very basic brain functions, such as

[31] My claim is not that essential dependence absolutely obtains in the case of Covid-19 or any other particular case. More modestly, I hope to establish the possibility of these sort of dependence relations between our classifications and causes of disease that are, in some sense, subject to adaptation. Thanks to an anonymous reviewer for pushing me on this point.

[32] I accept that talk of 'higher-level' and 'complex' neural functions is somewhat vague. Roughly speaking, what I have in mind are the sorts of complex abstract functions which feature, for example, in the RDoC matrix (Morris & Cuthbert, 2012; see also Hyman, 2000). Although it is beyond my scope here, much can be done to elaborate on this notion.

the internal functioning of neurons), higher level neural processes have a remarkable potential for neuroplastic adaptation to external stimuli and novel environmental demands. When these novel environmental demands *include* our classification practices, theories and conceptions, it seems possible that the underlying super-explanatory neurobiological property (that is, the dysfunctional neurocircuitry causing the psychiatric disorder) may change, systematically, in response to our practices of classification. That is, like our evolving pathogens in the above, neural circuits are amenable to *adaptation* (broadly construed).[33]

Whether and when this *in fact* occurs in particular cases is going to be difficult to ascertain. However, by way of illustration, let us consider a (hypothetical) example. Suppose Substance Use Disorder is caused by a maladaptive feedback loop between the reward system, the stress system and the executive system in the brain. If so, this characteristic *interaction* between neural systems is the natural super-explanatory property which causes the secondary properties of addiction, such as neglect of important life goals (as suggested in previous sections). Suppose further that addiction is reconceptualised and reclassified as a *disease*, rather than as moral deviance (as it was previously understood). Assuming addiction is realised by a dysfunctional neural circuit with the capacity to adapt plastically to environmental factors it seems possible that this change in our classification (which, after all, forms part of the social, political and material environment within which our brains must operate) could affect the super-explanatory neural processes. If the natural property underlying the disease is a neural process with the potential for adaptation in response to environmental (including social) factors, there seems no principled reason to assume that our classification practices could *only* ever impact upon the kind's secondary properties (such as behavioural symptomology).[34]

To be clear, I am not committed to essential dependence obtaining for this, or any other, particular psychiatric disorder. For now, I seek merely to establish that there are principled reasons why psychiatric disorders may be more essentially dependant on our classifications than paradigm somatic disorders, *even* if they are underpinned by a super-explanatory natural property such as characteristic neurobiology. Establishing that there are biological regularities underpinning particular psychiatric disorders, as Tsou attempts to do, is insufficient grounds to establish that these super-explanatory properties are *not* the sorts of properties that could be reactive in response to classifications. We must be careful to avoid the fallacious inference that biological equals immutable and psychological equals changeable. Psychiatric disorders are biological; this is not the issue. The question is, rather, what *sorts* of biological properties underpin psychiatric disorders, and do they have the potential to change systematically in response to our classifications? I believe that we have principled and empirically informed reasons to think the answer here is 'yes'.

[33] Adaptation needs to be broadly construed here as the mechanisms by which the brain adapts and that through which a virus adapts are rather different (for some relevant discussion, see Garson, 2019).

[34] Tsou appears to draw this erroneous inference.

6 Conclusion

Under what circumstances is a natural disease kind truly reactive in response to our classification, and when is it merely co-instantiated with or intersecting with a corresponding social and reactive kind? I have argued that a disease kind is only really reactive in the sense of being unstable or 'on the move' under conditions of 'essential dependence', that is, where the super-explanatory natural property characterising the kind is changing systematically in response to our practices of classification. I went on to argue that there are principled theoretical reasons to think that this sort of dependence can only occur in cases where the super-explanatory natural property is amenable to certain forms of directed change or 'adaptation' (broadly construed) in response to environmental factors.

In what cases can this occur? I have argued that essential dependence is a possibility in two sets of cases: 1) diseases caused by pathogens which are literally evolving (fungi, bacteria, Sars-Cov-2) and 2) diseases which are caused by dysfunctions in higher level neural processes with the capacity for neuroplastic adaptation to environmental conditions. This implies a possible, and heretofore underappreciated, disanalogy between neuropsychiatric disorders and paradigm biomedical disorders. If the hypothesis advanced in this paper holds, then special sorts of interdependencies may obtain between (certain types of) neuropsychiatric disorders and our classifications of these disorders. These sorts of interdependencies may undermine the prospects for a science of, and a scientific classification system for, psychiatry which is strongly premised on biomedicine.

Acknowledgements I would like to thank David Papineau, Marion Godman and participants at the workshop 'Reactivity and Categorisation in the Human Sciences' for valuable feedback on this research. I am also grateful to participants at Interdisciplinary Philosophy and Medicine Workshop at King's College London for their insightful comments and questions. This research was conducted while in receipt of a studentship from the London Arts and Humanities Partnership.

Funding This research was conducted while receiving a studentship from the London Arts and Humanities Partnership (LAHP).

Declarations

Ethical approval N/A. Work is entirely theoretical and includes no participants.

Informed consent N/A. Work is entirely theoretical and includes no participants.

Conflict of interest None.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Benedetti, F. (2020). *Placebo effects*. Oxford University Press.

Borsboom, D., Cramer, A. O., & Kalis, A. (2019). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences, 42*, e2.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 141–185). MIT Press.
Cavanna, A. E., Strigaro, G., & Monaco, F. (2007). Brain mechanisms underlying the placebo effect in neurological disorders. *Functional Neurology, 22*(2), 89.
Chapman, R. (2020). The reality of autism: On the metaphysics of disorder and diversity. *Philosophical Psychology, 33*(6), 799–819.
Chaturvedi, P., Tiwari, V., Singh, A. K., Qavi, A., & Thacker, A. K. (2020). Depression impedes neuroplasticity and quality of life after stroke. *Journal of Family Medicine and Primary Care, 9*(8), 4039.
Cooper, R. (2004). Why Hacking is wrong about human kinds. *The British Journal for the Philosophy of Science, 55*(1), 73–85.
Costandi, M. (2016). *Neuroplasticity*. MIT Press.
Cozolino, L., & Sprokay, S. (2006). Neuroscience and adult learning. *New Directions for Adult and Continuing Education, 110*, 11–19.
Cramer, S. C., Sur, M., Dobkin, B. H., O'Brien, C., Sanger, T. D., Trojanowski, J. Q. ... Vinogradov, S. (2011). Harnessing neuroplasticity for clinical applications. *Brain, 134*(6), 1591–1609.
Craver, C. F. (2009). Mechanisms and natural kinds. *Philosophical Psychology, 22*(5), 575–594.
Cunningham, B. (2001). The reemergence of'emergence'. *Philosophy of Science, 68*(3), S62–S75.
Devitt, M. (2008). Resurrecting biological essentialism. *Philosophy of Science, 75*(3), 344–382.
Devitt, M. (2021). Defending intrinsic biological essentialism. *Philosophy of Science, 88*(1), 67–82.
Doig, A. J. (2018). Positive feedback loops in Alzheimer's disease: The Alzheimer's feedback hypothesis. *Journal of Alzheimer's Disease, 66*(1), 25–36.
Garson, J. (2019). *What biological functions are and why they matter*. Cambridge University Press.
Godman, M. (2018). Gender as a historical kind: A tale of two genders? *Biology & Philosophy, 33*(3), 1–16.
Godman, M., Mallozzi, A., & Papineau, D. (2020). Essential properties are super-explanatory: Taming metaphysical modality. *Journal of the American Philosophical Association, 6*(3), 316–334.
Godman, M., & Papineau, D. (2020). Species have historical not intrinsic essences. In A. Bianchi (Ed.), *Language and reality from a naturalistic perspective* (pp. 355–367). Springer.
Hacking, I. (1998). *Rewriting the soul: Multiple personality and the sciences of memory*. Princeton University Press.
Hacking, I. (1999). *The social construction of what?* Harvard University Press.
Hacking, I. (2007). Natural kinds: Rosy dawn, scholastic twilight. *Royal Institute of Philosophy Supplements, 61*, 203–239.
Hyman, S. E. (2000). Mental illness: Genetically complex disorders of neural circuitry and neural communication. *Neuron, 28*(2), 321–323.
Insel, T. R., & Cuthbert, B. N. (2015). Brain disorders? precisely. *Science, 348*(6234), 499–500.
Kauffman, G. B. (1985). The role of gold in alchemy. Part II. *Gold Bulletin, 18*(2), 69–78.
Khalidi, M. A. (2010). Interactive kinds. *The British Journal for the Philosophy of Science, 61*(2), 335–360.
Khalidi, M. A. (2013). *Natural categories and human kinds: Classification in the natural and social sciences*. Cambridge University Press.
Khalidi, M. A. (2020). Are sexes natural kinds?. In *Current controversies in philosophy of science* (pp. 163–176). Routledge.
Klawiter, M. E. (1999). *Reshaping the contours of breast cancer: From private stigma to public actions*. University of California.
Koob, G. F., & Simon, E. J. (2009). The neurobiology of addiction: Where we have been and where we are going. *Journal of Drug Issues, 39*(1), 115–132.
Koob, G. F., & Volkow, N. D. (2016). Neurobiology of addiction: A neurocircuitry analysis. *The Lancet Psychiatry, 3*(8), 760–773.
Millikan, R. G. (2000). *On clear and confused ideas: An essay about substance concepts*. Cambridge University Press.
Millikan, R. G. (2017). *Beyond concepts: Unicepts, language, and natural information*. Oxford University Press.
Morris, S. E., & Cuthbert, B. N. (2012). Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior. *Dialogues in Clinical Neuroscience, 14*(1), 29–37. https://doi.org/10.31887/DCNS.2012.14.1/smorris
Okasha, S. (2002). Darwinian metaphysics: Species and the question of essentialism. *Synthese, 131*(2), 191–213.

Plutynski, A. (2018). *Explaining cancer: Finding order in disorder*. Oxford University Press.
Poland, J., & Tekin, S. (Eds.). (2017). *Extraordinary science and psychiatry: Responses to the crisis in mental health research*. MIT Press.
Shlomchik, M. J., Craft, J. E., & Mamula, M. J. (2001). From T to B and back again: Positive feedback in systemic autoimmune disease. *Nature Reviews Immunology, 1*(2), 147–153.
Stinear, C. (2010). Prediction of recovery of motor function after stroke. *The Lancet Neurology, 9*(12), 1228–1232.
Tekin, Ş. (2014). The missing self in Hacking's looping effects. In H. Kincaid & J. A. Sullivan (Eds.), *Classifying psychopathology: Mental kinds and natural kinds*. Boston Review.
Tracy, J. I., Hampstead, B. M., & Sathian, K. (Eds.). (2014). *Cognitive plasticity in neurologic disorders*. Oxford University Press.
Tsou, J. Y. (2007). Hacking on the looping effects of psychiatric classifications: What is an interactive and indifferent kind? *International Studies in the Philosophy of Science, 21*(3), 329–344.
Tsou, J. Y. (2013). Depression and suicide are natural kinds: Implications for physician-assisted suicide. *International Journal of Law and Psychiatry, 36*(5–6), 461–470.
Tsou, J. (2021). *Philosophy of psychiatry* (Elements in the philosophy of science). Cambridge University Press.
Walsh, P., Elsabbagh, M., Bolton, P., & Singh, I. (2011). In search of biomarkers for autism: Scientific, social and ethical challenges. *Nature Reviews Neuroscience, 12*(10), 603–612.