5
DELUSION AND NATURAL KINDS

Richard Samuels

The status of psychiatric kinds has long been a focus of dispute (Szasz 1960). Debates regarding the status of *diagnostic* categories, such as schizophrenia and borderline personality disorder, are testimony to such concerns (Pickard 2009). But so too are discussions of categories of *symptoms* – especially, first-rank symptoms, such as delusion. To a first, very rough approximation, what lies at the heart of such disputes is whether the kinds of psychiatry are, in an appropriate sense, real divisions or groupings in the world, or whether they are arbitrary or merely conventional groupings. Though such issues may be framed in quite different ways, one familiar approach frames the issue in terms of whether the salient psychiatric kind is a *natural kind*.

In this chapter, I consider an instance of this sort of issue – whether *delusion* is a natural kind. In doing so, I make a pair of assumptions. First, I assume there are at least some natural kinds. That is, for present purposes, I disregard eliminativism about natural kinds (Hacking 2007; Ludwig 2018). Unless we do so, it makes little sense to assess the specific claim that delusion is a natural kind. Indeed, and for similar reasons, I assume for present purposes that the paradigm candidates of natural kinds – chemical elements, atomic particles, and biological species, for example – are, in fact, natural kinds. For if we reject such relatively uncontentious candidates, it makes little sense to assess whether a highly contentious candidate, like delusion, is a natural kind.

Second, I assume that it is commonplace for natural kinds to exhibit some form of taxonomic organization. This is apparent in the case of paradigmatic candidates of natural kinds, such as those of chemistry, particle physics, and biological systematics (Ellis 2001).[1] Consider the periodic table. Magnesium is plausibly a natural kind. But metal is also a natural kind – a relatively superordinate kind that has magnesium, iron, aluminium, etc. as subordinate kinds. Moreover, *chemical element* is also a plausible candidate for natural kindhood, albeit one that has magnesium and metal as nested subordinate kinds.

Kinds of delusion also appear to exhibit nesting relations. Notoriously, the clinical literature is replete with many sorts of delusions – persecutory delusions, erotomanic delusions, Othello syndrome, Capgras delusion, and delusions of passivity, to name but a few – which may themselves be organized into relatively superordinate categories – e.g. monothematic and polythematic; bizarre and mundane. Now, we may ask of the above kinds whether they

DOI: 10.4324/9781003296386-7

are natural kinds. But in the present chapter, my primary focus will be on delusion *as such*. That is, I will be concerned with what I call the *NK thesis*: Delusion *as such* is a natural kind.

Clarifying and assessing this thesis involves the coordination of two quite different sorts of issue. First, we must consider an array of broadly empirical issues regarding delusions. Second, we must consider the general matter of what natural kinds are. As we will see, there are several extant accounts, which impose quite different requirements on being a natural kind. In view of this, I propose to index discussion of the NK thesis to specific conceptions of natural kinds and consider the matter of whether the thesis is plausible on *that* construal of natural kinds. More specifically, I consider the thesis in the light of three views about natural kinds: *natural kind essentialism*; Richard Boyd's *homeostatic property cluster* account of natural kinds; and what Muhammad Ali Khalidi calls *the simple casual account*.

Here's how I proceed. In Section1, I consider the plausibility of the NK thesis on the assumption that natural kind essentialism is true. I argue that on this assumption, the thesis is *very implausible*. But in Section 2, I argue that the fault lies with natural kind essentialism, which should be rejected in favour of some alternative account of natural kinds. In Section 3, I consider the merits of the suggestion that delusion is a natural kind in the sense articulated by the *Homeostatic Property Cluster* (HPC) account. Finally, in Section 4, I consider the plausibility of the NK thesis, on the assumption that the simple causal account is correct.

1. Natural kind essentialism and delusion

According to one influential view, natural kinds are to be characterized in terms of *the possession of common essences* (Barnes 1984; Kripke 1972; Putnam 1973). Though *Natural Kind Essentialism* (NKE) may be formulated in different ways, one typical rendering maintains the following:

NKE: A kind K is a natural kind if and only if (and because):
- *Individuation:* All and only the members of K share a common essence E.
- *Necessity:* E is a property, or collection of properties, that all and only K members *must* have.
- *Intrinsicality:* E properties are intrinsic as opposed to extrinsic or relational.
- *Causal Centrality*: E properties cause the instantiation of other properties associated with members of K.[2]

Some philosophers maintain that NKE applies well to the kinds of some sciences. For example, Brian Ellis (2001) maintains that the chemical elements have essences and, hence, are natural kinds in this sense.[3] However, it is implausible to maintain that delusions have essences of the sort articulated by NKE (Ghaemi 2004).

1.1 Delusions and the individuation condition

One reason to doubt that delusions possess an essence is that there's little reason to suppose they satisfy Individuation. Reflection upon the dismal track record of efforts to define delusion is relevant here.[4] Despite extensive efforts, it's highly doubtful that any extant proposal

provides satisfactory necessary and sufficient conditions for delusion; and certainly, no such proposal garners consensus from the relevant scientific communities.[5]

To illustrate the problem, consider what is arguably the closest thing we have to a consensus definition of delusion – what I call the *Standard Account* – because it's found in many texts, including *DSM-IV*, the glossary of *DSM-5*, and ICD-11. On this view, delusion is a species of *belief*, whose instances possess the following characteristics:

1 *Falsity:* A delusion is a *false* belief.
2 *Conviction:* Delusions are *firmly held* by the patient.
3 *Doxastic Isolation:* A patient's delusion is not accepted by other members of the person's culture or subculture.
4 *Resistance to Rational Persuasion:* Delusions cannot be dispelled by argument – including *good* argument – to the contrary.
5 *Resistance to Incompatible Information:* Delusions are maintained in the face of available, incompatible information.

Satisfying Individuation would require that the Standard Account provide necessary and sufficient conditions for delusion. Yet it does not plausibly do so. First, as many have noted, most – perhaps even all – of 1–5 are unnecessary for delusion:

- Not all delusions are false. For example, there are reported cases of hypochondriacal delusions ('I am ill') and Othello Syndrome ('My partner is unfaithful'), where the belief is true but nevertheless delusional (Gipps and Fulford 2004).
- Not all delusions are doxastically isolated. Sometimes – as in the case of folie à deux – the very same delusion is held by multiple individuals.
- There's ample evidence that the conviction with which delusions are held can vary, and at times fail to be strong (Garety and Freeman 1999).
- Though delusions are often resistant to argument, the efficacy of cognitive therapy casts doubt on the claim that they *always* are. As Richard Gipps and Bill Fulford (2004: 227) note 'in some ways people with delusions can be brought to reflect on their delusional beliefs and, with patience and support, be brought to question them'.

Second, it is far from clear that the Standard Account provides a sufficient condition for delusion. To illustrate, consider the following case:

> *Stubborn Philosopher:* Prof. Smith is deeply invested both personally and professionally in a 'pet theory' (PT) that they alone defend. Unbeknownst to Smith, PT is false, and was arrived at by faulty reasoning. Moreover, despite strong argument and evidence to the contrary, Smith has remained ardent in his commitment to PT. Indeed, he has become quite well-known within his subfield for defending this view.[6]

Philosophy, I suspect, is replete with Smith-like figures; and yet no psychiatrist would return a diagnosis of psychosis for Smith merely because they exhibit the profile sketched above. Mere bullish commitment to a pet theory – even when false, implausible, and singularly held – need not be delusional.

1.2 Delusion and the intrinsicality condition

A second reason to doubt that delusion satisfies NKE is that even if clinicians were to identify a set of properties that satisfy Individuation, there's little reason to suppose that these properties would be *intrinsic*. This is because psychiatric kinds quite generally – delusion, in particular – are almost invariably characterized *relationally*. Again, reflection on the Standard Account is useful here, since the sorts of properties it cites are quite typical of extant approaches to delusion:

- *Doxastic Isolation* requires that a patient's delusion is not accepted by other members of the person's culture or subculture. Yet this quite explicitly articulates a requirement concerning the *relation* between the delusion and the patient's cultural context.
- *Resistance to Rational Persuasion* and *Resistance to Incompatible Information* are also relational. Specifically, they involve relations to accessible arguments and available information.
- *Falsity* is plausibly a relational property. Specifically, on almost every philosophical theory of truth, falsity is characterized in terms of the *absence* of a relation. According to correspondence theories, for example, the falsity of a belief consists in its failure to correspond to the facts; and according to coherence theories, the falsity of a belief consists in its failure to cohere with other beliefs.
- *Belief*: Perhaps less obviously, on almost every account of what beliefs are, being a belief is a relational property. For example, according to functionalism – arguably the dominant approach over the past half century – being a belief is a functional property specified in terms of how beliefs tend to relate to environmental stimuli, behavioural responses, and other mental states. In which case, delusions as a species of belief will of course inherit this relationality.
- *Intentional Content:* Perhaps even less obviously, for delusions to be the sorts of states that could have truth values, they will need to have *intentional contents*: they would need to represent or be *about* something. Yet on almost everyone's view, intentionality is itself relational.

Of course, none of the above precludes the bare possibility of a future psychiatry that characterizes delusion non-relationally. But if extant practice is any guide, this outcome would be quite remarkable.

2. Natural kinds without traditional essentialism

The upshot of Section 1 is that delusion is not plausibly a natural kind by the lights of NKE. Yet reflection on scientific kinds quite broadly has led philosophers of science to view NKE as highly problematic, and to reject it in favour of less demanding accounts of natural kinds. In view of this, the possibility remains that the NK thesis is correct, albeit relative to some more credible account of natural kinds. In this section, I first say more about the dominant assumptions that underwrite most extant views of natural kinds, and then rehearse some of the main reasons for rejecting NKE.

2.1 Some dominant assumptions about natural kinds

Though the notion of a natural kind has had a chequered intellectual history, it regained some semblance of respectability among philosophers of science, largely due to its utility in

understanding core aspects of scientific practice (Boyd 1991; Quine 1969). Such philosophers very typically adopt the following three assumptions.[7] First, they assume that the division of phenomena into kinds – i.e. classification and taxonomy – is a central aspect of scientific practice: one that's important in its own right, but also a prerequisite for other core aspects of science, including the production of generalizations, theories, models, and explanations. Second, they commonly assume a form of *methodological naturalism*. Methodological naturalists typically suppose that science is our best – albeit defeasible – guide to empirical reality. For present purposes, the crucial and widely shared assumption is that scientific classification is our best guide to what *kinds* of things there are in the Universe. Finally, following John Venn (1866), contemporary philosophers of science typically construe natural kinds as the denotations of scientific classificatory terms.[8] Thus chemical elements, biological species, and subatomic particles (e.g. leptons, neutrons, and electrons) are paradigmatic natural kinds, since they are among the kinds acknowledged by the classificatory schemes of apparently successful science.

Given the above assumptions, a good first step in developing a philosophical account of natural kinds is to identify common features of the (putative) kinds that scientific terms denote. In doing this, philosophers of science have proposed several candidate features, with the following related ones being perhaps the most widely endorsed:

> *Inductive potential*: Natural kinds are denotated by terms that can enter into successful inductive inference. As such, natural kind *terms* are *projectible*
> *(Goodman 1955; Magnus 2012)*

> *Fecundity:* Natural kinds are, at least pragmatically, the subject of many scientifically relevant, empirical generalizations
> *(Machery 2005; Mill 1843/1882; Quine 1969)*

> *'Stickiness':* Natural kinds are 'sticky' in the sense that they are associated with clusters of properties (and relations) that, while logically unrelated, reliably covary (Khalidi 2023). As a result, instances of each specific natural kind – e.g. specific samples of gold, or individual electrons – tend to have lots of properties (and relations) in common
> *(Boyd 1991; Khalidi 2023; Mill 1843/1882)*

Though we need not address the precise relationship between these three features, they will each be relevant to later discussion.[9]

2.2 Reasons to reject NKE

In view of the above comments, the principal reason to reject NKE is easily stated. If scientific classification is our best guide to what natural kinds there are, then the conception of natural kinds afforded by NKE is manifestly *too restrictive*. For while it might hold of (say) the elements of the periodic table, it clearly does not apply more broadly to the kinds denoted in scientific classifications. One reason is that many such kinds are not plausibly characterizable in terms of their intrinsic properties. This is true of biological kinds, such as species, but also true of the kinds of psychology, materials science, and arguably physics (Boyd 1991; Griffiths 1997; Khalidi 2023; Magnus 2012; Samuels 2009). A second reason is that even if we reject intrinsicality and allow essences to contain relational properties, the

modified view is still overly restrictive. Specifically, it's doubtful that natural kind membership is always definable in terms of the causally central characteristics of kind members. This is plausibly true, for example, of the kinds cell and neuron. Finally, as Boyd (1991) and others have observed, there is not the slightest reason to suppose that all the (presumed) natural kinds denoted by scientific terms can be defined by sets of individually necessary and jointly sufficient conditions – i.e. that Individuation is true. Boyd's parade case is biological species; but the point is almost certainly true of many other plausible candidates for natural kindhood, including the kinds of psychology, anatomy, and ecology.

The above has clear implications for the present discussion of delusion. First, if NKE is unacceptable as an account of natural kinds, then the considerations rehearsed in Section 1 provide no grounds for rejecting the NK thesis. But second, if NKE is unacceptable, then we would do well to consider the plausibility of the NK thesis in the light of other, more plausible accounts of natural kinds. In the remainder of this chapter, I do just this. Specifically, I focus on a pair of views about natural kinds that are often regarded as more plausible characterizations of special science kinds, especially those that figure in the brain and behavioural sciences. In Section 3, I consider the plausibility of the NK thesis in light of Boyd's well-known *homeostatic property cluster* (HPC) view; and in Section 4 I consider it in the light of what's sometimes called the *simple causal theory* (SCT) of natural kinds (Craver 2009; Khalidi 2013, 2023).

3. Delusion as a homeostatic property cluster

The HPC view of natural kinds is often adopted by philosophers interested in assessing the naturalness of psychological and psychiatric kinds (Beebee and Sabbarton-Leary 2010; Griffiths 1997; Machery 2005; Samuels 2009). In what follows I briefly sketch the view and consider the NK thesis in the light of this view.

3.1 The HPC account of natural kinds

According to the HPC account,[10] K is a natural kind if:

> H1. K is associated with a property cluster P: a contingently co-varying collection of properties that tend to be co-instantiated by instances of the kind, but need not be necessary conditions for kind membership.
>
> H2. There is *homeostatic mechanism*[11] – an empirically discoverable causal mechanism(s) that explain why members of K reliably exemplify the property cluster P.
>
> H3. To the extent that there is any real definition of what it is for something to be a member of K, it is not the property cluster, as such, but the presence of the underlying homeostatic mechanism that defines membership of K.[12]

Consider an illness such as influenza. Influenza plausibly meets conditions H1–3: (a) it is associated with a range of characteristic symptoms – coughing, elevated body temperature, etc.; (b) there's a causal mechanism – roughly, the influenza virus – which explains the occurrence of such symptoms; and (c) to the extent that the kind, influenza, has a definition, it's not symptoms as such, but the virus that characterizes the kind.

As P. D. Magnus observes, the HPC view isn't plausible as a *definition* of natural kind – that 'natural kind = HPC' (Magnus 2012: 147). This is because some paradigmatic natural kinds

appear not to be HPC kinds. Most obviously, the account won't apply to the kinds of fundamental physics – e.g. electron and quark. Although these are among the paradigm examples of natural kinds, it's hard to see how H2 could be true of such kinds since, q*ua fundamental* entities, the covariation of their associated properties could not be explained by some causal mechanism, process, or structure. Nonetheless, as noted earlier, many have supposed the Boyd's proposal offers a plausible characterization of special science kinds – especially the kinds of biology and the behavioural sciences. In view of this, it makes sense to ask how plausible the NK thesis is, on the assumption that delusion must be an HPC kind.

3.2 Prima facie grounds for optimism about the NK thesis

Assuming the HPC account of natural kinds, the best evidence for the NK thesis would be an empirically attested account of the causal mechanism responsible for the co-variation of properties associated with delusion. Evidently, we have no such account. Nevertheless, in previous work, I suggest prima facie reasons to take seriously the view that delusion is an HPC kind, including the following (Samuels 2009):

> *Consideration 1: Delusion is associated with a contingently covarying property cluster.* Earlier, when discussing the Standard Account of delusion, I noted that none of the five conditions it imposes are necessary for delusion and that collectively these conditions are insufficient. Nevertheless, it is very often the case that delusions instantiate *all* five properties. But this is precisely what we would expect if delusion is an HPC kind. In short, the suggestion is that, as per HPC, delusion has an associated covarying property cluster, and that the Standard Account specifies at least some of the properties which constitute it.
>
> *Consideration 2: Empirical Regularities*. If the NK thesis is true, then we should expect appropriate methods of enquiry to yield a body of empirical generalizations concerning delusion. Though it is too early to tell with any certainty, there are grounds for optimism on this score. Over the past few decades, a wide array of results has emerged regarding delusions. For example, there is considerable evidence of cognitive abnormalities in the reasoning, attention, metacognition, and attributional tendencies of delusional patients (Bell et al. 2006; Garety & Freeman 20013). To illustrate, first consider the manifestation of a reasoning bias, known as *the jumping to conclusions (JTC) bias*. This bias consists in gather very limited information – far less than typical subjects – when making judgments and decisions. Moreover, it is strongly statically associated with the manifestation of delusions (Dudley et al. 2016). For a second, very different sort of illustration, consider the convergent neuroscientific evidence regarding the role of the dopamine system in the formation and maintenance of delusions (Weinberger 2022).

Evidently, such considerations do not provide strong grounds for accepting the NK thesis. Nevertheless, they may provide some prima facie reason to take the claim seriously.

3.3 Some challenges for the view that delusion is a HPC kind

There are also several potential challenges to the view that delusion is an HPC kind. Here are two.[13]

3.3.1 Continuity objections

For delusion to be a HPC kind, it needs to be a *kind*. However, some have challenged the categoricity of delusion, instead viewing delusion as *continuous* with non-delusions (Lincoln 2007; van Os 2003). It's important to see that the complaint is not merely that delusions comprise a kind with *vague* boundaries. After all, there's no obvious problem with kinds having borderline cases; and as Andrea Scarantino and Paul Griffiths (2011) note, Boyd's proposal is, in part, designed to accommodate the borderline cases of special science kinds. Rather, the claim is that delusion should be construed as a range on a continuum for which there exists no well-motivated boundaries – vague or otherwise – between delusion and other sorts of state. For example, it has been suggested that psychotic delusion is continuous with the sorts of 'delusional ideation' that occurs in ordinary experiences of the general population (Linscott and van Os 2013; Verdoux et al. 1998). In one review, for example, it was revealed that more than 8 per cent of the general population report psychotic experiences (van Os et al. 2009).

Whether such considerations pose a significant challenge to the NK thesis is unclear. This is because the observation that members of nonpathological populations have delusion-like experiences is wholly consistent with the NK thesis. First, the NK thesis in no way implies that delusions are always pathological. By widespread consensus, having a psychopathology presupposes that some *evaluative* condition is met – such as being *harmful* (Wakefield 1992). In which case, for all we have said so far, the delusion-like experiences of nonpathological subjects many be *bona fide* delusions – albeit nonpathological, because they fail to meet the relevant evaluative condition (for more on delusion and pathology, see Petrolini, Chapter 1). Second, the sort of continuity that the available evidence supports is a continuity of *symptoms*. Roughly put, non-pathological subjects have mental states whose properties are very similar to, or overlap with, those associated with psychotic delusion (Verdoux and van Os 2002). But such *symptomatic* continuity is wholly consistent with the NK thesis. After all, it's not symptoms, as such, but the instantiation of the salient homeostatic mechanism that determines category inclusion. And this, of course, permits many phenomena that are superficially similar to delusion and yet not genuine delusions. (Compare: influenza is symptomatically like many other illnesses, e.g., the common cold. But that does not mean that a continuum theory of influenza should be adopted.)

3.3.2 The unity problem

Perhaps the most serious challenge to the claim that delusion is a HPC kind is what I elsewhere call the *Unity Problem* (Samuels 2009). According to the HPC view, natural kinds are individuated by their associated homeostatic mechanisms (H3). But if this is so, then the absence of a shared homeostatic mechanism would be reason to deny that delusion is an HPC kind. The problem is that we currently have no well-developed account of what common mechanism(s) might be responsible for the formation and maintenance of delusion. Rather, the issue remains an ongoing and substantially unresolved empirical matter.

In view of this predicament, what might be said about the mechanisms responsible for the formation and maintenance of delusion? For the moment, I restrict myself to a single observation:[14]

> *Determinable Type Unity:* We should expect the Unity Problem to be resolved by identifying a determinable, mechanism type, as opposed to some highly determinate mechanism.

Sometimes Unity Problems in the sciences are resolved by identifying some very specific sort of mechanism or process that produces instances of the kind. Consider, for example, the mechanism responsible for a typical form of colour-blindness, deuteranopia. This form of red-green colourblindness is explained by a relatively specific mechanism – the presence of functional retinal L-cones and S-cones, accompanied by an absence of functional M-cones ('green' cones). The relative specificity of this mechanism contrasts with the case of retinal colourblindness more broadly. Here, the type of primary mechanism consists in the reduction of function for *some* subset of retinal cone types. But crucially, the specific types of cones that are affected, and the degree to which they are impaired varies, thereby producing different forms of colour blindness (Bartolomeo 2021). In the case of retinal colourblindness, then, there is a kind of unity at the level of mechanism. But it takes the form of a *determinable* mechanism type, of which there are different *determinates* for different sorts of colour blindness.

What does any of this have to do with delusion? The suggestion is that, in view of the vast array of different sorts of delusion, a solution to the Unity Problem will likely be analogous to the case of retinal colourblindness. That is, the Unity Problem, if it is to be solved, will likely involve the specification of a determinable mechanism type.

One extant proposal that fits well with the above suggestion come from the Bayesian Brain research program (Adams et al. 2015). According to this influential approach, a central function of the brain is *predictive processing*. That is, a central function of the brain is to minimize *prediction error*: the disparity between the sensory signals the brain 'expects', based on its model of the world, and the sensory signals it actually receives (Bongiorno and Corlett 2024). Moreover, because of the brain's structural organization, especially in sensory cortex, proponents of the Bayesian Brain hypothesis often suggest that predictive coding is hierarchically organized into interacting layers where prior beliefs and hypotheses are conveyed via backward connections to predict inputs from the layer below.

Against these background commitments, several theorists defend what Sterzer and colleagues (2018) call the *canonical account* of delusion – that 'the relevant disturbance consists in an overweighting of sensory precision as compared to prior beliefs' with the result that 'prediction errors call for unneeded and sometimes profound revisions in an agent's model of the world' (Bongiorno and Corlett 2024). To illustrate, consider the case of Capgras delusion, where the patient believes that a loved one, such as a wife, has been replaced by an imposter. As Federico Bongiorno and Philip Corlett (2024) explain, according to the canonical account:

> In Capgras delusion, the connection between the face recognition system and the autonomic nervous system is damaged, so that seeing your wife's face does not elicit the appropriate level of autonomic response, as it does in healthy subjects (Ellis, Lewis, et al. 2000). The ensuing mismatch between the expected response and the actual lack of response gives rise to a prediction error indicating that your internal model is mistaken and needs to be updated. According to the canonical account, the misleading error signal would be afforded excessive precision and allotted undue influence on model revision. This would cause rejection of the hypothesis that is most compatible with their prior beliefs ('this person is my wife') in favour of one which is delusional but more observationally adequate ('this person is an imposter').

The point of the present illustration is not to argue that the canonical account is correct. Rather, my point is that the canonical account, construed as a general account of

delusion, fits well with the suggestion that the Unity Problem could be resolved at the level of determinable mechanism types. In the case of Capgras, it is the connection between face recognition and the autonomic nervous system that is disrupted, and the salient predictive hierarchy will involve these systems. In other cases, e.g. passivity phenomena or delusions of grandeur, the contents of the delusions will be quite different, and the salient predictive hierarchies will likely be different as well. Nevertheless, the mechanism that results in delusion will still, according to the canonical account, involve an overweighting of sensory precision as compared to prior belief. That is: we have sameness of determinable mechanism type, with a difference of determinate mechanism.

4. Delusion and the simple causal theory of natural kinds

Though the HPC account has been highly influential in recent debates about the status of special science kinds, it is not the only option. In this section, I close by considering the NK thesis in the light of an alternative view, which, following Carl Craver (2009), Khalidi calls the *Simple Causal Theory* (SCT) (Khalidi 2013, 2023).

4.1 The simple causal theory of natural kinds

One way to construe the SCT is that it results from relaxing a requirement imposed by Boyd's HPC account. Specifically, we drop the requirement for a homeostatic mechanism that explains the covariation of the kind's associated property cluster while insisting that

> natural kinds are the kinds appearing in generalizations that correctly describe the causal structure of the world regardless of whether a mechanism explains the clustering of properties definitive of the kind.
>
> *(Craver 2009: 579)*

As Khalidi explains, the suggestion is not that the *mere* presence of a property cluster suffices for natural kindhood. It would not suffice, for example, if the associations were merely conventional.[15] Instead, what's required is that 'causation … glues these properties together' (20023: 27). In this regard, the SCT is like the HPC account, which also supposes that the covariation of properties associated with a natural kind is a result of causation. But in contrast to Boyd's view, the SCT does not insist that the salient causal relations obtain in virtue of a homeostatic mechanism. Instead, it suffices that K and its associated properties be 'nodes' in 'recurring causal networks'.[16] We thus end up with something like the following view: K is a natural kind if:

> S1. K has an associated cluster of properties, $P_1 \ldots P_n$ that contingently co-vary, and tend to be co-instantiated by instances of K.
>
> S2. The covariation relations between K and $P_1 \ldots P_n$ obtain in virtue of their causal relations to each other – the fact that they are all nodes in a 'recurring causal network'.
>
> S3. To the extent that there is any real definition of what it is for something to be a member of the natural kind, K, it is to be an instance of, K, where this kind is itself individuated by the salient recurring causal network within which $P_1 \ldots P_n$ are nodes.

4.2 Assessing the NK thesis in light of SCT

Suppose we adopt the SCT of natural kinds. What are the implications for the NK thesis? Let's start with the grounds for optimism. First, assuming SCT, the prima facie grounds for optimism rehearsed in Section 3.2 carry over. That is: the fact that (a) delusion is associated with a contingently covarying property cluster, and (b) researchers have identified various empirical regularities regarding delusion, comports well with the proposal that delusion is, by the lights of SCT, a natural kind.

Second, if we reject the HPC view in favour of the SCT, then a primary challenge to the NK thesis – the Unity Problem – no longer seems pressing. Recall, on the HPC account, natural kinds are individuated by an underlying homeostatic mechanism. Thus, the failure to identify such a mechanism for delusion is prima facie grounds to doubt the NK thesis itself. In contrast, since the SCT relaxes the demand for a homeostatic mechanism, failure to identify such a mechanism for delusion is no longer grounds to doubt the NK thesis. Thus, the switch from the HPC account to the SCT of natural kinds is accompanied by a reduction in the demands on defending the NK thesis. Defending this thesis no longer seems to require the specification of a homeostatic mechanism for delusion.

Nevertheless, the adoption of the SCT is accompanied by another, albeit quite different, problem. Specifically, it appears to give rise to indeterminacy worries for the NK thesis. The problem may be formulated as a dilemma:

> Horn 1: As typically presented, SCT is implausibly profligate as an account of natural kinds. As such, on such presentations, delusion will count as a natural kind, but only because SCT is implausibly profligate.
>
> Horn 2: If SCT is not to be implausibly profligate, then certain restrictions must be imposed – of which more below. But now – and until some plausible restriction is identified – we have *no idea* whether delusion is a natural kind.

Either way, the transition to SCT appears not to aid us in determining whether the NK thesis is true.

To appreciate the first horn of the dilemma, consider Craver's original gloss on the SCT: 'natural kinds are the kinds appearing in generalizations that correctly describe the causal structure of the world' (2009: 579). Literally interpreted, this formulation generates an implausibly profligate conception of natural kinds. Almost any grouping will count as a natural kind. To illustrate, consider *constellation*. By broad consent, constellation – Orion, Leo, Aries, etc. – is not a natural kind. It is a kind (or category) that consists of arbitrary groupings of stars that we created, presumably because we can see gestalt patterns in arbitrary configurations of entities. Yet there are causal generalizations involving the kind constellation – e.g. that on looking at the night sky people with the requisite knowledge and intentions tend to make judgements regarding the presence of constellations. Since nothing prevents an arbitrary grouping of things from being a cause, the above characterization of SCT turns arbitrary groupings into natural kinds.

Of course, those sympathetic to SCT never thought that constellation counts as a natural kind. Among other things, they assume that bona fide natural kinds figure in *lots* of causal generalizations, not just a few. Or to put the point in terms of causal networks: They assume that natural kinds are 'highly connected vertices in directed causal graphs' – that

they involve causal links to *many other* properties (Khalidi 2023). But now we confront the second horn of our dilemma. If the difference between natural kinds and other kinds is to be understood in terms of the *richness* of their causal connections, then we appear to need some principled way of specifying when the connections are rich *enough*. And this is something that proponents of SCT are yet to provide.

Now, proponents of SCT are fully aware of this sort of issue.[17] Moreover, the failure to specify such a principled distinction need not be a problem for many purposes. Suppose, for example, we are concerned with the kind *electron*. Here we might suppose that the richness of the causal connections is *so great* as to render it as clear-cut a case of a natural kind as anything is. But the situation is very different when we turn to delusion. As noted earlier, researchers have identified some causal regularities involving delusion. Yet extant formulations of the SCT provide no guidance regarding whether these regularities are rich *enough* to confer natural kind status on delusion. That is, without precisification of the SCT, the status of the NK thesis remains indeterminate.

5. Conclusion

In the light of different conceptions of natural kinds, this chapter explored the issue of whether delusion is a natural kind – whether the NK thesis is true. I first argued that, on well-known essentialist accounts of natural kinds, the NK thesis is almost certainly false, but that this is due to the implausible stringency of natural kind essentialism. Next I argued that on Boyd's more plausible HPC account of natural kinds, there are *prima facie* reasons to take the NK thesis seriously. But I also noted that the thesis faces some significant explanatory challenges, most notably the Unity Problem. Finally, I argued that, on extant formulations of the recently popular SCT of natural kinds, the status of the NK thesis remains indeterminate. Combining the above, the status of the NK thesis remains unresolved. The issue of whether delusion is a natural kind will only be resolved in the light of further empirical study of delusion, but also – and crucially – sustained philosophical attention to the issue of what natural kinds are.

Notes

1 Unlike Ellis (2001) I won't assume that organization into a taxonomic hierarchy is criterial for natural kinds.
2 For a formulation of this view see, for example, Khalidi (2023).
3 To illustration, if the kind *hydrogen atom* is a natural kind, then: (a) All and only members of the kind share a common property – e.g. having the atomic number one; (b) something *must* have this property to be a member of the kind; (c) this property will be intrinsic; and (d) having this atomic number, will cause the instantiation of other properties associated with hydrogen.
4 For a useful review, see Garety and Hemsley (1994). See also Radden (2011).
5 This is perhaps reflected in the most recent editions of the Diagnostic and Statistical Manual of Mental Disorders – DSM-5 and DSM-5-TR – in which little effort is made to provide a definition of delusion.
6 For discussion of a related phenomenon in science see Maher (1988: 20–22).
7 It should be noted that the sort of approach to natural kinds discussed here is not the only one. See Chapter 4 of Cooper (2007) for other approaches.
8 At any rate, the good ones. After all, not all the terms that figure in scientific classifications successfully denote. The terms 'caloric', 'phlogiston' and 'aether' readily come to mind.

9 One thought is that Stickiness explains Fecundity and Fecundity explains Inductive Potential. In more detail: It's because the instances of natural kinds reliably share a cluster of contingently related properties (and relations) that natural kinds tend to be subject of many true empirical generalizations. And it is because they are the subject of many such generalizations that one can reliably make inductive inferences about the instances of natural kinds.
10 For more extensive characterizations of the homeostatic cluster view see, for example, Boyd (1991, 1999).
11 Although Boyd often speaks of homeostatic mechanisms, it clear that he has an expansive conception of mechanisms that incorporates lots of things that are often called processes, states and structures.
12 In Boyd's words, '[t]he natural definition of …homeostatic property clusters kinds is determined by the members of a cluster of often co-occurring properties and by the ("homeostatic") mechanisms that bring about their co-occurrence' (Boyd 1999: 141).
13 For more extensive discussion see Samuels (2009).
14 In earlier work I further suggested that the salient mechanisms are likely cognitive mechanisms that are multiply realized by various neural processes (Samuels 2009).
15 To illustrate: *constellation* won't be a natural kind since our grouping of stars into constellations – Orion, Leo, Aries, etc. – is (merely) conventional.
16 Though I won't explore the issue here, the present view is clearly formulated in the light of recent developments in casual modelling. For details see, for example, Pearl (2000) and Woodward (2003).
17 Khalidi appears to view this sort of indeterminacy – or at any rate something very much like it – as more-or-less inevitable. See Khalidi (2023: 28).

References

Adams, R. A., Brown, H. R. and Friston, K. J. (2015) "Bayesian Inference, Predictive Coding and Delusions", *Avant* 5: 51–88.

American Psychiatric Association (APA) (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. 4th ed. Washington, DC: American Psychiatric Association.

American Psychiatric Association (APA) (2013) *Diagnostic and Statistical Manual of Mental Disorders*: DSM-5. Washington, DC: American Psychiatric Association.

Barnes, J. (Ed.) (1984) *The Complete Works of Aristotle*. Aristotle, categories. Princeton, NJ: Princeton University Press.

Bartolomeo, P. (2021) "Color Vision Deficits", *Current Neurology and Neuroscience Reports* 21: 1–7.

Beebee, H. and Sabbarton-Leary, N. (2010) "Are Psychiatric Kinds Real?" *European Journal of Analytic Philosophy* 6(1): 11–27.

Bell, V., Halligan, P. W. and Ellis, H. D. (2006) "Explaining Delusions: A Cognitive Perspective", *Trends in Cognitive Sciences* 10(5): 219–226.

Bongiorno, F. and Corlett, P. R. (2024) "Delusions and the Predictive Mind", *Australasian Journal of Philosophy*, 1–16. https://doi-org.proxy.lib.ohio-state.edu/10.1080/00048402.2023.2293825

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. Wilson (ed.), *Species: New Interdisciplinary Essays*. Cambridge: MIT Press, 141–186.

Cooper, R. (2007) *Psychiatry and Philosophy of Science*. Stockfield: Acumen.

Craver, C. F. (2009) "Mechanisms and Natural Kinds", *Philosophical Psychology* 22(5): 575–594.

Dudley, R., Taylor, P., Wickham, S. and Hutton, P. (2016 May) "Psychosis Delusions and the 'Jumping to Conclusions' Reasoning Bias: A Systematic Review and Metaanalysis", *Schizophrenia Bulletin* 42(3): 652–665.

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

Ellis, H. D., Lewis, M. B., Moselhy, H. F. and Young, A. W. (2000) "Automatic without Autonomic Responses to Familiar Faces: Differential Components of Covert Face Recognition in a Case of Capgras Delusion", *Cognitive Neuropsychiatry* 5: 255–269.

Garety, P. and Freeman, D. (1999) "Cognitive Approaches to Delusions: A Critical Review of Theories and Evidence", *British Journal of Clinical Psychology* 38: 113–154.
Garety, P. A. and Freeman, D. (2013) "The Past and Future of Delusions Research: From the Inexplicable to the Treatable", *British Journal of Psychiatry* 203(5): 327–333.
Garety, P. A. and Hemsley, D. R. (1994) *Delusions: Investigations into the Psychology of Delusional Reasoning*. Oxford: Oxford University Press.
Ghaemi, S. Nassir (2004) "The Perils of Belief: Delusions Reexamined", *Philosophy, Psychiatry, and Psychology* 11(1): 49–54.
Gipps, R. G. and Fulford, K. W. M. (2004) "Understanding the Clinical Concept of Delusion: From an Estranged to an Engaged Epistemology", *International Review of Psychiatry* 16(3): 225–235.
Goodman, N. (1955/1983). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.
Griffiths, P. E. (1997) *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.
Hacking, I. (2007) "Natural Kinds: Rosy Dawn, Scholastic Twilight", *Royal Institute of Philosophy Supplement* 61: 203–239.
Khalidi, M. A. (2013) *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. New York: Cambridge University Press.
Khalidi, M. A. (2023) *Natural Kinds*. Cambridge: Cambridge University Press.
Kripke, S. (1980) *Naming and Necessity* Cambridge, MA: Harvard University Press.
Lincoln, T. M. (2007) "Relevant Dimensions of Delusions: Continuing the Continuum Versus Category Debate", *Schizophrenia Research* 93(1–3): 211–220.
Linscott, R. J. and van Os, J. (2013) "An Updated and Conservative Systematic Review and Meta-Analysis of Epidemiological Evidence on Psychotic Experiences in Children and Adults: On the Pathway from Proneness to Persistence to Dimensional Expression across Mental Disorders", *Psychological Medicine* 43(6): 1133–1149.
Ludwig, D. (2018) "Letting Go of 'Natural Kind': Towards a Multidimensional Framework of Non-Arbitrary Classification", *Philosophy of Science* 85: 31–52.
Machery, E. (2005) "Concepts Are Not a Natural Kind", *Philosophy of Science* 72: 444–467.
Magnus, P. (2012) *Scientific Enquiry and Natural Kinds: From Planets to Mallards*. New York: Palgrave-Macmillan.
Maher, B. A. (1988) "Anomalous Experience and Delusional Thinking: The Logic of Explanations", in T. F. Oltmanns and B. A. Maher (eds.), *Delusional Beliefs*. New York: John Wiley & Sons, 15–33.
Mill, J. S. (1843/1882) *A System of Logic* (8th edition). New York: Harper & Brothers.
Pearl, J. (2000) *Causality*. Cambridge: Cambridge University Press.
Pickard, H. (2009) "Mental Illness Is Indeed a Myth", in M. Broome and L. Bortolotti (eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford: Oxford University Press, 82–102.
Putnam, H. (1973), "Meaning and Reference", *Journal of Philosophy* 70, 699–711.
Quine, W. V. O. (1969) *Ontological Relativity and Other Essays*. New York: Columbia University Press.
Radden, J. (2011) *On Delusion*. New York: Routledge.
Samuels, R. (2009) "Delusions as a Natural Kind", in M. Broome and L. Bortolotti (eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford: Oxford University Press, 49–79.
Scarantino, A. and Griffiths, P. (2011) "Don't Give Up on Basic Emotions", *Emotion Review* 3(4): 444–454.
Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M. and Corlett, P. R. (2018) "The Predictive Coding Account of Psychosis", *Biological Psychiatry* 84(9): 634–643.
Szasz, T. S. (1960) "The Myth of Mental Illness", *American Psychologist* 15(2): 113–118.
van Os, J. (2003) "Is There a Continuum of Psychotic Experiences in the General Population?" *Epidemiologia e Psichiatria Sociale* 12: 242–252.
van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P. and Krabbendam, L. (2009 Feb) "A Systematic Review and Meta-Analysis of the Psychosis Continuum: Evidence for a Psychosis Proneness-Persistence-Impairment Model of Psychotic Disorder", *Psychological Medicine* 39(2): 179–195.

Venn, J. (1866/1888) *The Logic of Chance* (3rd edition). London: Macmillan.
Verdoux, H., Maurice-Tison, S., Gay, B., Van Os, J., Salamon, R. and Bourgeois, M. L. (1998) "A Survey of Delusional Ideation in Primary-Care Patients", *Psychological Medicine* 28(1): 127–134.
Verdoux, H. and van Os, J. (2002) "Psychotic Symptoms in Non-Clinical Populations and the Continuum of Psychosis", *Schizophrenia Research* 54(1–2): 59–65.
Wakefield, J. C. (1992) "Disorder as Harmful Dysfunction: A Conceptual Critique of DSM-III-R's Definition of Mental Disorder", *Psychological Review* 99(2): 232–247.
Weinberger, D. R. (2022) "It's Dopamine and Chizophrenia all Over again", *Biological Psychiatry* 92(10): 757–759.
Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.