An ideal disorder? Autism as a psychiatric kind

Daniel A. Weiskopf

Abstract: In recent decades, attempts to explain autism have been frustrated by the heterogeneous nature of its behavioral symptoms and the underlying genetic, neural, and cognitive mechanisms that produce them. This has led some to propose eliminating the category altogether. The eliminativist inference relies on a conception of psychiatric categories as kinds defined by their underlying mechanistic structure. I review the evidence for eliminativism and propose an alternative model of the family of autisms. On this account, autism is a network category defined by a set of idealized exemplars linked by multiple levels of theoretically significant properties. I argue that this network model better captures the empirical phenomena, the historical growth of the category, and the ways the category has been shaped by social norms and interests. Finally, I defend a realist interpretation of network categories against the challenge from eliminativists.

1. Introduction

Psychiatry attempts to organize mental disorders by imposing schemes of classification that sort them in ways that serve a range of ends, including inference, explanation, and, ultimately, interventions and cures. One output of psychiatric theorizing should be a taxonomy of conditions that groups them according to the distinctive sort of cognitive and emotional breakdown that they manifest. Here I’ll argue that one prominent model of psychiatric kinds fails to capture the dynamics of how disorders—at least some of them—are categorized. I take as my paradigm one of the most exhaustively studied mental disorders, namely autism spectrum disorder (ASD), including autism itself, Asperger’s syndrome, and their kindred pervasive developmental disorders (PDDs).¹

Autism has been regarded as the model of a biologically grounded neuropsychological disorder, a standard by which to measure our understanding of psychiatric disorders as

¹ Although I mostly refer to autism as a “disorder” throughout this discussion, I recognize that there are those who reject this label, preferring the language of difference instead (Silberman, 2015). In some cases, at least, it is clear that persons with autism can display patterns of abilities that are significantly superior to those of neurotypical individuals, and it is a complicated question whether to regard these departures from the “norm” as disorders (Lorenz & Heinitz, 2014).
manifestations of pathology occurring in underlying cognitive and neural systems. In recent years, however, cracks have begun to appear in this seemingly ideal case. At least one researcher, Lynn Waterhouse (Waterhouse, 2009, 2013; Waterhouse & Gillberg, 2014), has argued that autism is so heterogeneous that it does not actually exist as a single disorder, or even a set of subgroups of disorders. There are autistic symptoms, caused by a variety of mechanisms in much the way that fever is a symptom of different underlying conditions, but there is no disease or set of diseases that might felicitously be called “autism” or “the autisms”. In philosophical terms, Waterhouse espouses eliminativism about autism, arguing that this nosological revision will improve our diagnostic, inferential, and therapeutic practices; hence, the only viable scientific path forward is “to abandon the use of the diagnosis of autism” (p. 39).

A similar skepticism has been voiced by Sam Timimi, a child psychiatrist, and his co-authors Neil Gardner, and Brian McCabe, both of whom have been diagnosed with ASD (Timimi, Gardner, & McCabe, 2011). They conclude their extensive review, which surveys much of the same ground as Waterhouse’s, by straightforwardly declaring that “there is no such thing as autism and the label should be abolished,” adding that “it is hard to see any scientific, clinical or even research value for the construct of autism” (p. 286).

The conclusion that autism is not a unitary disorder and therefore should be eliminated from psychiatric classification rests on two key assumptions, both of which I will challenge. The first is that failure to find the right sort of unifying mechanism is sufficient for something’s not being a kind of disorder. In the next section I will sketch the picture of kinds that lies behind this assumption. The second is that scientific classifications should not include non-disorders, or that these should be relegated to secondary status, perhaps in virtue of failing to perform the same epistemic and explanatory work as categories of disorders do. This assumption is required
because it might be that autism is not a kind of disorder but is nevertheless a condition that plays a different role in psychiatric practice. If so, autism might not be a single disorder, but this would not imply wholesale eliminativism about it.

2. Psychiatric kinds and the HPC model

Among philosophers, the most plausible candidate for being the received view of kinds within the biological and biomedical sciences is Richard Boyd’s Homeostatic Property Clusters (HPC) account (Boyd, 1999). An HPC kind has three core aspects:

(1) a family of contingently clustered properties that exist in a homeostatic equilibrium, where this clustering may be statistical or otherwise imperfect;

(2) an underlying causal mechanism that both maintains this homeostatic state and explains the co-occurrence of the property cluster; and

(3) the causal importance of these clusters within a larger disciplinary matrix, meaning that they produce “theoretically or practically important effects” in an established context of inquiry.

For Boyd, what matters for kindhood is the existence of explanatorily significant statistical property clusters linked by an underlying causal mechanism. These clusters allow exceptions, so kind members can resemble each other imperfectly; and because the boundaries of the kind are drawn not by the statistical clusters alone, but also by the causal mechanism that produce them.

---

2 Such at least is my impression; see Wilson, Barker, and Brigandt (2007) for evidence to this effect, and also Magnus (2012), Ch. 6. Note that the claim here is that philosophers tend to reconstruct theories and practices in the biological sciences in terms of the HPC model, not that scientific practitioners themselves tacitly endorse this model. Some have made this latter claim, though I have my doubts about it and for present purposes I leave it an open question.
HPC kinds are partially independent of their merely diagnostic characteristics. The existence of an underlying mechanism is a guarantee that the observed clustering is non-accidental. If these mechanisms operate independently of our classifying activities, HPC kinds achieve a degree of objectivity.

Boyd’s account is tailor made for biological species, but the HPC account has also been adopted as a model for understanding psychiatric taxonomy. Jonathan Tsou, for example, argues that schizophrenia, depression, and suicide are all psychiatric natural kinds in virtue of their conforming to the HPC pattern (Tsou, 2013, 2016). In each case, we have a set of correlated symptoms that are produced and sustained by an underlying mechanism; depression, for instance is partially constituted by anhedonic mood accompanied by disturbed eating and sleeping, fatigue and impaired concentration, feelings of guilt and worthlessness, and suicidal ideation, all of which are thought to be produced by abnormal processing of monoamine neurotransmitters. In similar fashion, Richard Samuels (2009) argues that delusions should be seen as an HPC kind.

Two points bear emphasis about the HPC account. The first is that kinds are relativized to explanatory and classificatory practices: they are about how categories are picked out within the epistemic and practical concerns of a disciplinary matrix, and a theory of kinds should capture the way that the discipline and the structure of the world accommodate each other. The second is that kinds are defeasibly tied to their cluster of characteristic properties, and while these clusters may be imperfect, they have to have a reasonable degree of statistical integrity. If a property cluster becomes too disunified, it threatens the existence of the kind just as much as the lack of

---

3 It’s widely acknowledged that the HPC account is less well-suited to many other sciences. Magnus (2014) notes, for instance, that it fails to capture the relations among the fundamental inventory of particles posited by the Standard Model in microphysics. Critics have also questioned whether it can account for non-causal, functional, and highly polymorphic kinds, many of which occur within biology (Ereshefsky & Reydon, 2014), and even its applicability to species is not as clear as its proponents have advertised (Reydon, 2006). For an overview of kinds in psychiatry, see Zachar (2014).
an underlying mechanism would. As we will see, this generates much of the eliminative pressure on autism.

3. Hunting for mechanisms

Autism has been defined most prominently by deficits in theory of mind (ToM), or mentalizing and social interaction more broadly. This manifests in a number of ways, including failures on various forms of false belief tasks and difficulty in comprehending ordinary conversational pragmatics. Indeed, the existence of autism was itself considered by some to be a prime piece of evidence for the existence of a ToM module, as would be reasonable if a highly specific deficit was involved. There is disagreement over exactly how the ToM deficit comes about – some posit that it is an early problem with the mechanisms of tracking intentionality through gaze (Baron-Cohen, 1995), others that it is a more general metarepresentational deficit (Perner, 1991). In more contemporary neuroscientific models it has been suggested that autism involves a disorder of the mirror neuron system (Dapretto et al., 2006). Whether couched in cognitive or neural terms, though, the general equation is clear: fundamentally, autism is mindblindness.4

Despite its appealing simplicity, the mindblindness thesis faces difficulties. For one, the disorder is typically more far-reaching than just a problem in understanding mental states. Most notably, there is the behavioral rigidity characteristic of autistic individuals, their often deep attachment to routine and rule-governed behaviors. This goes hand in hand with an almost obsessive and exclusive narrowness of interests, and contributes to the stereotype of the autistic

4 The term “mindblindness” has been used to cover a range of different capacities—see Gallagher (2004) for discussion.
person as a savant. It is hard to see this lack of breadth and fluidity as a direct consequence of a mentalizing disorder (Boucher, 2012).

Another notable characteristic is the perceptual hypersensitivity of many autistic individuals, who often seem to find everyday touches, sights, and sounds intrusive to the point of being utterly intolerable (Donnellan, Hill, & Leary, 2010). These sensory disruptions are sometimes paired with a range of impaired motor functions (Brincker & Torres, 2013). Again, there is no apparent theoretical link between pervasive sensorimotor disruptions and mentalizing, nor is it clear how to link these unpleasant and problematic traits to stereotyped patterns of interest.

Finally, a significant percentage of individuals with autism—55% in one epidemiological study—also suffer from intellectual disabilities (ID), with 16% of these cases being moderate to severe (Charman et al., 2011). While this association was previously assumed to be higher, it is again hard to square with a primary definition of the disorder that appeals to faulty mentalizing.

The history of attempts to explain autism is littered with the wrecks of failed theories. Aside from proposals centering on a damaged or absent ToM system, other accounts propose a more general cognitive dysfunction as the root of the problem. Some of these focus on perceptual processes, such as the so-called “intense world hypothesis” (H. Markram, Rinaldi, & Markram, 2007; K. Markram & Markram, 2010), while others target central processes, such as weak central coherence (Frith & Happé, 1994) or executive dysfunction (Pennington & Ozonoff, 1996). But these general dysfunction models face two problems: first, not all persons with autism display the relevant dysfunction, and second, the dysfunctions typically cannot explain all of the behavioral symptoms (Rajendran & Mitchell, 2007).
Consider, for example, the weak central coherence model, which proposes that cognition in autism involves a focus on local elements that are processed independently or piecemeal, rather than global or large-scale patterns that are processed by being holistically integrated. Yet weak central coherence can be dissociated from other core traits such as impaired joint attention and verbal ability, each of which make distinct contributions towards classification (Morgan, Maybery, & Durkin, 2003). One review notes that only a subgroup of persons with autism may display weak coherence, and that it probably has a different etiology than, and therefore cannot account for, ToM and other social shortcomings (Happé & Frith, 2006).

Beyond these psychoneural, behavioral, and social disruptions, there is also a host of poorly understood physiological and metabolic changes to be explained. This point is important, since it reminds us that autism is not a purely cognitive disorder but a more pervasive medical condition. For instance, 42% of ASD patients in one large group survey had either gastrointestinal (GI) disorders, seizures, or sleep disorders, 28% had two of these conditions, and 25% had none (Aldinger, Lane, Veenstra-VanderWeele, & Levitt, 2015). A meta-analysis covering 15 studies over 32 years found that persons with autism suffered from general GI complaints, constipation, diarrhea, and abdominal pain at significantly higher rates than non-affected patients (Mannion & Leader, 2013). GI disorders might be part of the etiology of some neurocognitive symptoms, or they might be produced by those symptoms, or there might be a more complicated form of feedback at play.

While it is assumed that autism must involve some underlying brain pathology, attempts to uncover it have mostly turned up inconclusive. The largest study to date draws on the Autism Brain Imaging Data Exchange, which collects images from 1000 participants aged 6-65 (Haar, Berman, Behrmann, & Dinstein, 2016). Ventricular volume was slightly greater in ASD patients,
and the central corpus callosum was slightly smaller, and some parts of occipital and temporal cortex were slightly thicker; beyond that, no significant anatomical differences were found. Overall, the anatomical profile of the ASD patients is “mostly indistinguishable” from that of normal controls—though it should be noted that these were all high-functioning (mean IQ 106) patients, and therefore the results may not generalize to all ASD individuals. This puts pressure on several cognitive theories of ASD, including the so-called “amygdala theory” (Baron-Cohen et al., 2000).

The picture when it comes to functional connectivity is equally complex and conflicting (Di Martino et al., 2014; Kana, Uddin, Kenet, Chugani, & Müller, 2014; Maximo, Cadena, & Kana, 2014). Early accounts proposed that the brains of persons with autism displayed patterns of long-distance underconnectivity (Just, Keller, & Kana, 2013). Recently, however, a series of studies has found evidence of widespread overconnectivity instead (Keown et al., 2013; Supekar et al., 2013). Making things more complicated, it may be that whether there is functional over- or underconnectivity is not a stable characteristic, but instead one that varies with the region or network being investigated, age and developmental level (Uddin, Supekar, & Menon, 2013), and overall task demands (Schipul & Just, 2016). While some of these connectivity patterns have suggestive links with the observed deficits of autism, there has been no systematic attempt to map them onto the entire complex cognitive phenotype.

Attempts to find a genetic basis for the disorder have led into the weeds in similar fashion. ASD is broadly heritable (82% concordance among monozygotic twins). However, it is also wildly genetically heterogeneous. One survey, for instance, identified over 100 known genetic disorders that are associated with ASD diagnosis, suggesting that it might constitute a “final common pathway” for a host of independent genetic disorders (Betancur, 2011, p. 62).
Still, an identifiable genetic alteration is present in only 10-20% of patients (Devlin & Scherer, 2012) and recent genome-wide association studies conclude that there is no individual single nucleotide polymorphism that predicts ASD, although many common variants, *en masse*, may have an effect (A. B. Lee, Mahajan, Manaa, & Pawitan, 2014). While a recent review speculates that ASD genes may converge on a developmental pathway involving synaptogenesis, given that there may be up to 1000-1500 genes involved in the disorder this is tentative at best (Ziats & Rennert, 2016). In other words, autism may be like schizophrenia, bipolar disorder, and many other psychiatric conditions in lacking a single unifying genetic mechanism (S. H. Lee et al., 2013).

4. Phenotypic variability and failures of idealization

Early research into autism focused on emotional and social deficits, later generalized to other forms of abnormal mentalizing, as being definitive of the disorder. This encouraged the hope that it would have a unified clinical presentation as well as a single underlying causal explanation. But as this review suggests, the contemporary picture is one on which autism is disunified at several levels of analysis, frustrating attempts to find mechanisms that could explain the core properties of the disorder. Making matters worse, even what *counts* as a “core property” of autism is up for grabs. Looking back at the history of attempts to stabilize its definitive behavioral characteristics reinforces the notion that it has always been an unruly condition. As Bonnie Evans observes, “[a]utism has always referred to the most severe pathology in infantile thought and changes in its meaning reflect broader historical changes in what is considered abnormal in the thought and behaviour of all infants and children” (Evans, 2013, p. 26).
This instability has been present since the beginning of autism studies. It’s widely recognized, for instance, that the case histories of Kanner and Asperger trace cognitive and behavioral profiles that are strikingly different in their emphasis (Draaisma, 2009). Kanner’s children were described in terms of *emotional* isolation, while today social and communicative challenges send to be seen as more central. Their speech was absent or limited to echolalia. Asperger’s children, by contrast, were not generally communicative but could talk elaborately about their specialized interests (the “little professors”). His cases also display the savant-like abilities at memory and calculation that have become so central to the popular (though inaccurate) image. These differences in stereotypical presentation reflect different understandings of what it means to be a person with autism.

These multiple presentations of the disorder have only proliferated today. For instance, a canonical trait of many children with autism is severely impoverished imagination (Leevers & Harris, 1998), accompanied by restricted or absent pretend play (Jarrold, 2003) and limited recourse to pretense. Yet this is hard to square with the creativity displayed by autistic individuals like Temple Grandin (2009), visual artist Stephen Wiltshire, or any number of technical and scientific innovators retrospectively judged to be “on the spectrum” whose work involves enormous insight and creativity by any measure (Mills, 2008). Which of these is closer to the representation of the “true” autistic individual? Which characteristics ought to be incorporated into our exemplar of the category and made the primary focus of our explanatory and therapeutic efforts?

Another long-puzzling phenotypic variation is between males and females. Autism is predominantly diagnosed among males, but whether this accurately reflects its distribution or is an artifact of poorly tuned screening is unknown (Timimi et al., 2011, Ch. 9). Recent studies
have begun to map how the presentation of the disorder differs in females versus males, and have found a number of striking variations in number of comorbid disorders, patterns of neural growth and development, and cognitive traits such as executive function and visuospatial processing, among others (Howe et al., 2015; Lai, Lombardo, Auyeung, Chakrabarti, & Baron-Cohen, 2015). This has inspired Simon Baron-Cohen’s highly controversial proposal that that autism is an “extreme male brain” syndrome (Baron-Cohen, 2010). Males, the theory claims, tend towards a “systematizing” intellectual style, while females tend towards “empathizing”; thus in autism, we allegedly find a pattern of adherence to abstract structure at the expense of emotional connection, consistent with placement far out on one end of this continuum. While this theory has few adherents, it is one of the only models to directly address this sharp phenotypic divide.

Finally, the move to treat autism as essentially including somatic complaints has taken hold strongly among parents and pediatricians. As noted above, GI symptoms as well as inflammatory, immune system, and sleep disorders are sufficiently frequent that they are now a focus of specific therapeutic attention. Chloe Silverman quotes a pediatric neurologist who emphasizes the importance of treating children with autism first and foremost as “sick kids”:

These complaints, once one learns to ask about them, turn out to be so common that it has become impossible for me to ignore them or assume that they are less important than the behavioral features. These children cannot be assumed to have nothing more than brain and behavior problems, since so many of them are also physically ill (2012, p. 191). Yet among cognitively oriented researchers one often searches in vain to find mention of such concerns, which form part of the daily struggles of parents and patients alike.

Recognition of these differences in presentation has motivated the creation of the Autism Phenome Project, an initiative founded at UC Davis in 2006 by David Amaral with goal of
mapping and defining the major clinical autism subtypes. To date the project has recruited over 300 families, gathering immune and genetic profiles, medical records, behavioral observations, and brain MRIs from each participant. These aspirations aside, though, even the brief examples reviewed here suggest that, in Silverman’s terms, “[a]utism has proven almost infinitely mutable” (2012, p. 12). And this instability in how autism is represented is problematic since psychiatric taxonomy in the HPC mold depends on the existence of a single coherent explanandum. It is therefore ill-suited to capture the kind of phenotypic heterogeneity at work in autism.

Graphically speaking, if we depict the possible phenotypes of autism within a multidimensional space whose dimensions correspond to ways that cognitive, behavioral, and somatic traits can vary in the disorder, the HPC account presumes that most instances will be grouped relatively near to each other, forming a single peak in the landscape, with variation distributed around the peak being minimal. This convergence in traits makes up the property cluster that characterizes the kind. What we actually find in autism, however, is a landscape of qualities in which there is no such dominant mass, but rather a bumpy terrain, a set of peaks joined by ridges representing local clusters of traits that can potentially be singled out as objects of investigation. These scattered peaks correspond to the exemplars that are picked out using the criteria employed by different investigative communities. In this sort of multifocal trait landscape, where no single property cluster dominates, the HPC model loses traction.

Autism, then, meets neither of the two main criteria for being an HPC psychiatric kind: it has neither a unitary property cluster nor a singular underlying mechanism. These results are fuel for the eliminativists’ fire. If every person with autism has it in their own way such that few predictions can be made about the qualities that they will have in common, the inferential utility
of placing anyone within the category decreases sharply. The range of therapies, supportive environments, and caregiver interventions similarly shrinks, making the category less practically significant as well.

I will briefly consider three objections before turning to my own proposal. One is that this picture ignores the role of idealization in constructing psychiatric categories. As Dominic Murphy (2006) emphasizes, normal explanatory practices, not only in psychiatry but in the sciences more generally, imply that much of the variation in actual patients’ profiles and experience is ignored as we shift our focus to the pattern that makes up the exemplary case. Idealization involves simplification of the messy real world by representing it in its purified form, capturing only the properties thought to be most theoretically significant within an investigative context, even if they are never actually exemplified.

However, while such idealization is often indispensable, it does not seem available given the extreme heterogeneity shown by autism. There is no good way to abstract a single core exemplary case out of the pattern of instances without doing violence to the whole, since there are many equally possible, equally good idealizations available, and—importantly—there is no objective basis for favoring one pathway to idealization over another. Idealizing along any particular direction discards significant variation, and the different idealized cases that result correspond to substantially different explananda.

A second reply is to split the disorder into two or more separate kinds. Kind-splitting involves separating a group into multiple fine-grained ones on the basis of their having independent mechanistic underpinnings (Craver, 2004). However, the fault lines in autism do not support such a revision. If they did, we would expect to find that the many etiological and constitutive mechanisms discovered so far could be gathered into reasonably coherent bundles.
each of which underlay a distinct clump of symptoms or focal exemplar. If all cases of a particular exemplary type were produced via a common causal pathway, or all language disorders in autism could be mapped onto a common neural dysfunction, autism might be ripe for being split along these lines.

But there is so far no evidence for this; the fissures in the underlying causal structure do not conform to the pattern of several mechanisms producing superficially similar but ultimately distinct disorders. Instead, we find incomplete, partial resemblance among conditions that share some but not all of their mechanisms, yielding a structure resembling a multilevel, misaligned patchwork. I elaborate more on this causal and statistical patchwork in the next section.

A third approach is to retreat to more modest notions of natural kindhood. But the problems canvassed so far generalize to many other popular accounts of kinds. For example, Matthew Slater (2015) proposes that kinds are stable property clusters, meaning that they tend to be “cliquish”: having any one of the properties in a cluster strongly indicates that the others will also be present. Further, these statistical connections are robust, meaning that they hold under a fairly wide range of possible counterfactuals and manipulations.

From what has been said, however, we can see that neither the cliquishness nor the robustness conditions hold for autism, since there are many ASD subpopulations in which these connections are disrupted. Happé and Ronald (2008) argue that even the core triad of diagnostic traits can be dissociated: in a study of 3000 monozygotic twin pairs ages 7-9 years old, “degree of social difficulty, communicative impairment, and rigid/repetitive behavior are only modestly related,” with correlations ranging from 0.2 to 0.4. Having single core traits was not uncommon (59% showing social impairments had only that trait), and having one of the core traits only
generated slight risk of showing one of the others (e.g., only 32% of children with social impairments showed communication difficulties).

Neither will it help to liberalize the sorts of causal relationships that underlie these groupings. One proposal along these lines is developed by Kendler, Zachar, and Craver (2010), who introduce the notion of mechanistic property kinds as a generalization of the notion of an HPC kind. While they are open to the prospect of many mechanisms operating simultaneously to sustain a single disease entity, they nevertheless hold that “the identity of the disease across time and across cultures is grounded in the similarity of the complex, mutually reinforcing network of causal mechanisms in each case” (p. 1148). That is, for any two instances of a disease, there must be either the same mechanisms present, or the same symptom cluster present, or both. But within autism we find cases being co-classified despite both of these being false.5

5. A network model of autism

Autism exemplifies a challenge that faces theories of kindhood and taxonomy more broadly, namely the problem of heterogeneous kinds. This term has an aura of paradox: the purpose of gathering entities into kinds is precisely to render them homogeneous with respect to their theoretically significant qualities—their characteristic properties, their causal mechanisms, or both. Psychiatric disorders such as autism, however, illustrate one way that categories can be theoretically and causally structured without being homogeneous. This puts pressure not only on the HPC account but on all philosophical theories of kinds that emphasize their inductive and predictive unity.

5 The same sorts of considerations can plausibly be deployed against an even more modest thesis, namely Muhammad Ali Khalidi’s causal theory of kinds, on which all that is necessary for kindhood is something’s being a node in a causal graph (Khalidi, 2013). Autism does not correspond to any single such node, as the model developed in the next section illustrates.
I suggest that what unifies autism is its network structure. In building a network to represent a disorder, we begin with the heterogeneous set of actual cases: particular patients, with their unique histories, biology, experience, and patterns of strengths and deficits. From these cases we can, via abstraction and idealization, generate the set of idealized exemplars that represent, given our present interests, the explanatorily important core case profiles. Each focal exemplar in the set is connected to at least one other in virtue of their sharing some theoretically significant property. This might be a common genetic etiology or developmental pathway, a somatic biomarker such as chronic inflammatory response or GI troubles, a shared neuroanatomical alteration or pattern of cognitive traits, or a set of behavioral proclivities. This chaining of exemplars allows the creation of a networked category in which there is no set of properties that all or most members have, but where each member is nevertheless connected to another by at least one theoretically significant property.

Taking a psychiatric category to consist of the set of networked idealized exemplars (plus the variation in cases that surround them), this model gives a way of organizing these members according to the properties that they share. Each link between members stands for some featural overlap, so linked pathways are equivalent to inferential routes showing how properties can be projected throughout the disorder as a whole. The numerical weight of a link in the chain represents the degree to which the property is shared between cases falling under each exemplar, and therefore the confidence with which the property can be inferred from one exemplar cluster to another.
Importantly, since many psychiatric disorders involve properties spanning many levels of analysis, they are not characterized by a single network but by a set of them. Exemplars belong to genetic, neurobiological, immunological, behavioral, cognitive, and many other networks—as many as there are sets of theoretically relevant properties for grouping cases. And given the complexities of interlevel relationships, there is no guarantee that these networks will coincide perfectly, meaning that cases that are chained by a common genetic etiology may not share a common neurobiological dysfunction, similar neural structures may produce dissimilar cognitive profiles depending on the sort of environment the patient inhabits, and similar behaviors may be underwritten by distinct types of cognitive profile.

Whether a property can be projected from one member to another, then, depends not only on network connectivity but also on which network level is being considered. Sharing properties at one level may imply little about whether they are shared at others. Finally, in a chained category, there may be no way to make inductive inferences from an arbitrary member to another; whether or not such inferences hold for any property and any pair of members will depend on the local structure of the network. Network categories may, but do not intrinsically, support general kind-based inductions.

The motivation for co-classifying different-seeming cases, then, is that they are members of the same network. However, if these cases are separated by a distance of several links, there may be relatively little resemblance between them, and few causal mechanisms that they share. One patient’s case may, for example, involve a preterm birth followed by consistently lagging development, moderate or severe ID, and poor language use, while another might involve a period of normal development without ID followed by regression in social contact and

---

6 Perhaps not all psychiatric disorders work like this. Some might be much more straightforward. The present claim is only that heterogeneous network-based disorders, of which autism is a paradigm, constitute an important and interesting class that have been largely neglected. See Section 6 for further caveats and limits on the view.
hypersensitivity to auditory stimuli. This is, of course, exactly what lies behind the complaint made by eliminativists that these individuals may have radically divergent capacities, experiences, and life prospects. All of which is true enough, but no objection to the network model, since the basis for classification here is not any particular shared syndrome, no matter how idealized.

One virtue of such a network model is that it captures the fact that, causally speaking, autism is best regarded as a psychiatric “perfect storm”. Perfect storms arise when a set of mostly uncorrelated conditions make individually small causal contributions to the production of a single large outcome. While perfect storms are often thought of as infrequent or unpredictable events, there may be regions of genetic, neurodevelopmental, and behavioral space where they are far more likely. Happé, Ronald, and Ploman (2006) make essentially this suggestion: “heterogeneity in ASD, on our account, is not simply due to noise or the complex unfolding of development, but is an unavoidable consequence of variation along at least three largely independent (although of course interacting) dimensions of impairment”. That is, autism might be a region of cognitive-behavioral space where a number of imperfectly interconnected deficits overlap or co-occur with some regularity, and where this co-occurrence can be achieved through a variety of separate neural and genetic pathways. On this view, a lot of things have to go awry to produce the characteristics associated with the family of autism phenotypes, but these cannot be collected and systematized into a unified mechanism of the disease.

Chain-based network models face challenges in establishing the boundaries of psychiatric categories, particularly since there may be individuals who share qualities with people who have a disorder but nevertheless do not have it themselves. If being chained to a category member by itself were sufficient for membership, disorders would have “leaky” boundaries. Many disorders
display this sort of blurring or overlap—thus it is not uncommon for patients to be diagnosed with multiple conditions, and many separate disorders share etiology and symptoms. The notion of the “broader autistic phenotype” is meant to capture the fact that family members of persons with autism may display characteristics that can be seen as paler, more mild versions of some of the disorder’s core traits, yet not have ASD itself.

This illustrates the fact that chaining alone is not what determines the presence of a disorder: rather, its boundaries are defined by its ideal exemplars plus the pattern of underlying networks that they share. It is nearness to a sufficiently focal exemplar and membership in sufficiently many networks that determines a disorder’s extension; similarly, there are two separate ways to fail to have a disorder, either by not falling under any of the idealized phenotypes captured by the exemplar cluster, or by failing to participate in any of the underlying networks. As the numerous hedging clauses indicate, these criteria are fuzzy, but no one, I think, expects sharp boundaries within psychiatric taxonomy.

A further virtue of the network model is that it accurately captures the historical dynamics of classification schemes for autism. Some might think this unimportant, but these historical facts partially explain why the category has the structure that it does (Tsou, 2011). Consider that psychological disorders are often first recognized as clinically important phenomena through the description and dissemination of cases. As statistical information about these cases and theories of how they are to be explained evolve, new cases are assessed for their “fit” with the emerging category structure. There can be many grounds for amending an existing category structure, but as new exemplars are grafted onto and pruned from this structure, its
branches and their endpoints may drift sufficiently far from the original case and from one another that any two arbitrarily chosen points may no longer strongly resemble one another.  

Notably, the dissemination of cases throughout a field such as psychiatry is not a process of passive diffusion, but one that requires active argument by clinicians, researchers, patients, and their advocates. This is part of the process by which these categories become (or fail to become) stabilized (Sullivan, 2014). Separate research communities may have their own criteria for inclusion or exclusion, which is part of what gives rise to the existence of multiple partially correlated networks: diagnostic arguments based on genetic criteria draw on different evidence and are addressed to different audiences than those based on, for example, behavioral signs in infancy. In Boyd’s terms, the accommodation demands of psychiatric genetics may produce one way of classifying a group of focal exemplars that differs from how developmental psychology classifies them.

The rules governing these dynamics are not strictly epistemic, but are also based on pragmatic, sociopolitical, and ethical factors. 8 Ian Hacking (2014) has even provocatively suggested that these latter factors have been the main drivers of reclassification in autism, but while this is an exaggeration, neither should they simply be dismissed. Heterogeneity is as much a product of our practices as it is of the world. As the exemplars we focus on vary, so do the properties that need to be explained about autism—both their content and their theoretical weighting. The disorder becomes a territory that interested parties compete over in an attempt to

---

7 While the category structure described here exists as a sociological phenomenon rather than as something represented within the minds of individuals, a number of studies show that people can readily learn and use categories that involve this sort of exemplar-driven growth, and that languages sometimes encode them (Heit, 1992; Xu, Regier, & Malt, 2016).

8 Simon Cushing (2012) has charted several of these revisions to the diagnostic criteria for autism in recent decades, and argues not only that these shifts have not been motivated by advances in scientific understanding, but also that intrusions of value into the classification undermine any hopes of interpreting it realistically. As the next section makes clear, I disagree with this pessimistic assessment. Bernd Verhoeff (2014) offers a more detailed look at the history of attempts to stabilize autism across disciplines, reaching a similar but more tentative eliminativist conclusion.
reclaim what it means. Verhoeff (2013) argues in this vein that “[d]iagnostic and classification practices cannot avoid depending on often implicit socio-cultural norms related to normal child development” (p. 6), and these therefore “play, in autism at least, a necessary role in what becomes a recognizable cluster of symptoms in the first place” (p. 7).

Variation in social norms and expectations about appropriate behavior may lead to disagreement about classifying cases. Where there is value-laden competition over the exemplars by which a disorder is represented, the relevant explanatory targets shift as well. In a heterogeneous disorder like autism, this means that what counts as an explanation of the phenotype will depend on which exemplar is taken as focal in a context. There is no way to settle such problems of kind stabilization in a value-free, bottom-up way, and this helps to account both for the shape of the autism network as it currently exists and the challenges to be reckoned with in refining it.

6. Realism and partial unification

Autism, on this account, is a category that links patterns of phenotypic variation by reference to a plurality of underlying theoretical connections. The eliminativist argument is driven by the notion that disorders must be unified by appeal to underlying causal-mechanistic basis. Since network models deny that there must be any such basis, they face the challenge of satisfying the broadly realist impulse that lies behind this argument.

I’d emphasize that this isn’t being advanced as a model for handling all psychiatric disorders. Some may be less heterogeneous, and hence have a more unified or simpler structure. But when viewed from a network perspective, disorders are real to the extent that the multilevel networks they are associated with show significant overlap. Overlap occurs where exemplars that
are members of a network at one level also tend to be connected within many others. This doesn’t require that the overlap regions be identical—i.e., that the network of cases that are chained to each other should align precisely—only that among these exemplars taken as a group there is a relative density of shared properties. If these networks were graphically superimposed on one another, the region corresponding to the disorder would be one that is thick with interconnections, even if not the same connections at each level.

The density of connections among exemplars at many levels of analysis can be thought of as a kind of real pattern, in the sense of Dennett (1991). It is not a pattern defined by sharing of any particular common properties or mechanisms, but one defined by the fact that at all levels of analysis, all members of the network family share properties with some others (such that there are no “stranded” members that connect to no others) and that this pattern of shared properties is relatively exclusive to members of the family. The idealized exemplars stand out from their surroundings by being the rough boundaries of a region of high density multilevel property overlap.

A similar conception of psychiatric disorders as “robust patterns” has been developed by Kathryn Tabb and Kenneth Schaffner (2017). Robust patterns, in their sense, are those that emerge from the simultaneous comparison of classifications of patients according to many theoretical perspectives: “When diverse decompositions overlap significantly they create a robust psychiatric kind, and we have reason to be optimistic that some of the interactional complexity between theoretical frameworks can be resolved in ways that will allow for prediction and manipulation”.

Tabb and Schaffner’s robust patterns analysis largely converges with the network model presented here, with the proviso that these decompositions can be diverse in two ways: not only
by involving many levels corresponding to the accommodation demands of the different fields that take the disorder as an object of study, but also by involving many subnetworks within a single level. These give rise to the “patchwork” organization alluded to above. Where disorders are intractably heterogeneous, the first stage in explanation is to map the contours of their variation. Networks do this by displaying the degree to which what can be said about one exemplar can be generalized to other, possibly quite different, ones, and they thereby deliver relevant information about the underlying relationships between different phenotypes.

Finally, network categories need not be thought of as capturing a single disease or disorder. Networks might be usefully identified with mental disorders for some purposes, but this isn’t obligatory. Moreover, even the question of whether there is one disorder here or several strikes me as unimportant: the network taken as a whole is the unit of analysis. Once we recognize this large-scale relational structure, the issue of how to count disorders is decidedly secondary.

It may be objected, still, that if autism has the structure of a network, it does not constitute a kind. As Happé and Ronald (2008) note, this doesn’t automatically follow: “it is quite compatible to assert that ASD results when a number of independent impairments co-occur, and to assert that the resulting mix has a special quality, distinct prognosis and response to intervention, and is therefore worthy of a distinct diagnostic label.” Yet even if there are no generalizations that cover all cases in the autism network, but only a patchwork of criss-crossing inferences among different subgroups of exemplars, this doesn’t show that our taxonomy needs revision. Rather, it shows that we are gaining a more fine-grained and accurate understanding of the real heterogeneous structure of the domain. The philosophical error, bequeathed to us by the

---

9 This marks one of several ways in which the network model proposed here differs from that of (Cramer, Waldorp, van der Maas, & Borsboom, 2010), who straightforwardly take disorders to be networks of comorbid symptoms.
tradition of natural kinds, is to perseverate in insisting on a model of scientific categories as homogeneous even when the phenomena repeatedly say otherwise. The category of autism will be worth keeping so long as the heterogeneity that we discover within it comes with a rich and systematic set of interconnections to explore that are grounded in the objective structure of the domain and hold out the promise of eventual interventions and practical therapies.

7. Conclusions

When faced with the Gordian knot of heterogeneity, eliminativism offers an Alexandrian solution. Yet once the knot has been cut, we are still faced with the problem of how to reunite the severed loose ends. Removing the label of “autism” does not in itself suggest any positive way forward, and so eliminativism, despite its radical sensibilities, leaves everything on the taxonomic scene fundamentally as it is.

The network model presented here, by contrast, has two overarching goals: first, to be a descriptively accurate account of the structure of the family of autisms, given our current knowledge; and second, to show how classifying specific problems and profiles as belonging to this family can bring out their commonalities and, ultimately, contribute towards better understanding and treating them. These networks reflect one sort of structure in the world that accommodates these demands tolerably well; better, I have argued, than does the eliminativist’s prospective erasure of any such category.

References


Cushing, S. (2012). Has Autism Changed? In J. L. Anderson & S. Cushing (Eds.), The
Philosophy of Autism (pp. 75–94). Plymouth, UK: Rowman and Littlefield.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., &


Opinion in Genetics and Development, 22, 229–237.

autism brain imaging data exchange: Towards large-scale evaluation of the intrinsic brain

Sensory And Movement Differences. Disability Studies Quarterly, 30.


986.

Evans, B. (2013). How autism became autism: The radical transformation of a central concept of


Samuels, R. (2009). Delusion as a natural kind. In M. Broome & L. Bortolotti (Eds.), *Psychiatry*
as Cognitive Neuroscience: Philosophical Perspectives (pp. 49–82). Oxford: Oxford University Press.


