8-10-2021

Challenges to Causal Classification in Psychiatry

Richard Link

Challenges to Causal Classification in Psychiatry

by

Richard Link

Under the Direction of Daniel Weiskopf, PhD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2021

ABSTRACT

Researchers attempting to explain and treat psychiatric disorders have long discussed the need for new disorder categories. A common proposal is to develop a causal classification system whereby psychiatric disorders are classified according to the causal processes that produce and sustain them. I evaluate the prospects for causal classification of psychiatric disorders. First, I motivate the need for a causal classification system of psychiatric disorders. Second, I introduce one framework for causal classification, the exemplar model. Third, I examine multiple forms of causal complexity that the exemplar model must successfully navigate if it is to be a viable classification model for psychiatry. I argue that the exemplar model can handle some issues of causal complexity but not others. I conclude that psychiatry should utilize two models of classification, the exemplar model and the network model, because each fulfills some, but not all, of the purposes of classification systems.

INDEX WORDS: Classification, Psychiatry, Causal complexity, Causal classification, Exemplar model, Network model

Challenges to Causal Classification in Psychiatry

by

Richard Link

Committee Chair: Daniel Weiskopf

Committee: Eddy Nahmias

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2021

DEDICATION

This thesis is dedicated to Lois Jane Link and Harold Link. Both these individuals helped raise me and helped me become the person I am today, albeit in very different ways. I love them both and they will be sorely missed.

ACKNOWLEDGEMENTS

I wish to express gratitude to Dan Weiskopf for his continued patience, support, and good humor. I have thoroughly enjoyed working on this thesis and finish it with as much interest in questions about psychiatric classification as I began. This is at least in part due to the inspiration of Professor Weiskopf.

TABLE OF CONTENTS

1 INTRODUCTION

Psychiatric nosology is often said to be in a state of crisis. Disorder categories are largely thought to be invalid. As a result, researchers are left with improper targets for research, and clinicians are less prepared to understand and address the problems their patients face. Many experts blame the flawed methodology used for classification by the American Psychiatric Association (APA) to develop the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) (APA, 2013). In response, several authors have attempted to articulate principles for a new classification system. Primarily, these authors desire a classification system that integrates causal information. This thesis will motivate the need for a causal classification system and assess the challenges involved in creating one. In particular, I will focus on Dominic Murphy's exemplar model of disorder classification for psychiatry (Murphy, 2006). The challenges I address primarily relate to causal complexity and heterogeneity. To account for the difficulties faced by the exemplar model, I argue that classification must also rely on network models.

Before motivating the need for a causal classification system, it will help to briefly summarize the APA's approach to classification in the DSM. The DSM is a classification of mental disorders designed for *clinicians* aimed at providing "reliable" and "accurate" diagnoses (APA, 2013, p. xli). The DSM uses diagnostic criteria to assist clinicians with making a diagnosis. These diagnostic criteria include "symptoms, behaviors, cognitive functions, personality traits, physical signs, syndrome combinations, and durations" (ibid., p. 5). Clinicians are supposed to be able to use these diagnostic criteria to determine a reliable diagnosis in the sense that it could be widely agreed upon throughout the clinical community. The DSM does not use information about the causes of disorders as criteria for sorting psychiatric phenomena into

categories. DSM categories merely sort individuals according to their signs and symptoms.[1] Therefore, the utility of DSM categories is that they allow clinicians to communicate about patients' clinical presentation.

The APA's stated motivations for a classification system go beyond merely communicating about patients' clinical presentation, however. These motivations include "guiding treatment recommendations, identifying prevalence rates for mental health service planning, identifying patient groups for clinical and basic research, and documenting important public health information..." (APA, 2013, p. 5). Already we have some tension: while it is explicitly stated that the DSM is for clinical use, DSM categories are also supposed to provide explanatory targets for researchers. While this is not a necessary tension–indeed, researchers and clinicians must ultimately share a common language–the failure of the DSM to provide suitable explanatory targets does undermine the utility of its diagnostic categories for clinicians. When researchers struggle to explain and provide interventions for clinical phenomenon, the ability of clinicians to treat their patients is undermined.

The APA is aware of the frustration felt towards its nosology. Specifically, the APA acknowledges that high comorbidity rates motivate concern regarding the validity of its categories, which it defines as "the degree to which diagnostic criteria reflect the comprehensive manifestation of an underlying psychopathological disorder" (APA, 2013, p. 5). While the DSM's curators likely believe that continued scientific progress will be sufficient for ameliorating these concerns, many believe that the problem is too deep to be rectified without radically rethinking how we approach classification in psychiatry.

[1] Whereas signs are measurable, symptoms are felt but unmeasurable indicators of disorders.

The remainder of this thesis will proceed as follows. Section two will motivate the need for a causal classification system and introduce a few ways psychiatric research has already moved in this direction. In doing so, I will also demonstrate the problems with the DSM approach to classification. Section three will introduce a model for causal classification in psychiatry, namely, Murphy's exemplar model. Section four will present several challenges to causal classification. The challenges have to do with two kinds of causal complexity, namely, multicausality and causal heterogeneity (Ross, forthcoming). I argue that the exemplar model does well at handling multicausality but struggles to address causal heterogeneity. I introduce another model of classification, the network model, which is better at representing causal heterogeneity. Finally, section five will discuss alternative ways of thinking about classification in psychiatry. I argue for a pluralist approach that embraces both the exemplar and network models of classification.

2 MOTIVATIONS FOR A CAUSAL CLASSIFICATION SYSTEM

This section will examine problems with the DSM classification system, specifically focusing on the issues that causal classification systems aim to redress. As such, there are several problems the DSM faces that I will not address. Briefly, these problems include (1) the demarcation between psychiatry and related fields in medicine and the cognitive sciences, (2) the statement of clear principles for determining when something qualifies as a disease or disorder, and (3) issues of transparency and conflicts of interest.[2] While these are important conceptual and procedural issues, adopting a causal classification system does not address them. For this reason, I proceed by examining the issues of comorbidity and heterogeneity. These problems relate to concerns regarding the in/validity of DSM categories. The notion of validity that I'm

[2] The work of Rachel Cooper often focuses on these issues; in particular see Cooper (2005; 2014). Also see the first three chapters of Murphy (2006).

working with corresponds to what Caroline Stone calls "construct legitimacy" (2019).[3] A construct, such as a disorder category, is legitimate to the extent that the theory that posits the construct is justified. Since the theory we are examining is our theory of classification, our disorder categories' legitimacy rises and falls with the justification for our classification principles.[4] After reviewing the issues with DSM constructs, I will present two alternative ways of grouping phenomena suggestive of principles we may use to construct an alternative classification system.

2.1 Comorbidity

Comorbidity is perhaps one of the most consistently discussed and pervasive issues in psychiatry. A comorbid patient suffers from two or more distinct illnesses. For example, someone suffering from both Major Depressive Disorder (MDD) and Generalized Anxiety Disorder (GAD) is comorbid. From a theoretical standpoint, high comorbidity rates demand an explanation. Why do so many individuals suffer from both MDD and GAD? Rather than try to answer that question directly, many approach high comorbidity rates from a different angle: do high comorbidity rates result from the fact that we do not have valid disease categories? If this latter possibility is the case, then comorbidity is an artefact of our current classification system.

Lecavalier et al. (2019) review studies of comorbidity in children with ASD. They note that comorbidity rates are high but that the exact percentage varies significantly. One study suggests that the number of children with concomitant psychiatric disorders could be as high as 92%, and another argues it is more likely around 41% (p. 58). Of the ASD children with

[3] Stone distinguishes between construct validity, construct legitimacy, and existence claims. Construct validity concerns the relation between measures and the construct. Construct legitimacy concerns the nature of the construct itself. Lastly, existence claims concern the relation between constructs in reality. (Stone notes that realists might not always distinguish between construct legitimacy and existence claims; I remain agnostic as to the question of realism).

[4] In keeping with the majority of the literature, I refer to construct legitimacy in terms of "validity" from here on out.

concomitant psychiatric disorders, 66.1% had two or more comorbidities. The authors comment that evidence of high rates "raises questions about uncertainty of the boundaries between psychiatric diagnoses in children" (p. 58). Furthermore, they suggest some pushback against the DSM-III method of criteria-driven categorical diagnosis and ask whether comorbidities are distinct conditions or "variable phenotypic manifestations of ASD" (p. 62).

It is common in discussions of psychiatric comorbidity for the authors to suggest revisions to the nosology. Implicit in the critiques leveled by Lecavalier et al. is skepticism (at best) of the categories created on the basis of clinical presentation. Supposing that various comorbidities with distinct presentations might be variations of the same disease category (ASD) as the authors suggest, one naturally wonders what exactly unites them under the same category. An article by van Loo et al. (2013) examines some of the literature discussing comorbidity. They argue that much of the work in this area assumes that high comorbidity rates result from our current classification system lacking a causal structure. Therefore, a reclassification of disorder categories based on causal factors might reduce comorbidity rates by eliminating artificial comorbidities.

Examining the work of Andrews et al. (2009), van Loo et al. construct the following argument for how common causes can account for high rates of comorbidity: (2013, p. 749)

1. The fact in need of explanation is that two diseases d1 and d2 occur far more frequently [together] than their separate frequencies suggest, i.e., they have high rates of comorbidity.
2. If there is a common cause C for the two diseases d1 and d2, then their high rates of comorbidity are to be expected.

Therefore,

3. It is plausible that the two diseases d1 and d2 have a common cause C.

Andrews et al. were taking part in a DSM-V initiative to produce a more parsimonious clustering of DSM diagnoses. They took high comorbidity rates within a cluster to be a "validating factor,"

i.e., evidence that the cluster mapped onto real similarities between diagnostic categories rather than superficial ones. Using high comorbidity rates to cluster diagnoses into categories that likely have shared causal mechanisms can be seen as conservative. Others argue that high comorbidity rates indicate that current disease categories cannot possibly be maintained as researchers search for explanatory mechanisms. Van Loo et al. attribute this view to Kendell and Jablensky (2003). I will return to this issue of understanding high comorbidity rates in section five. For now, let us take note that high comorbidity rates raise concerns about the validity of current diagnostic categories and are a consequence of the DSM's failure to integrate causal information into their classification system.

2.2 Heterogeneity

I now turn our focus to the heterogeneity of DSM diagnostic categories. To do this, I will enlist the work of philosopher Kathryn Tabb. Tabb authored a series of papers (2015; 2019) in which she argued that current diagnostic categories are not proper objects of study for psychiatric research. Her arguments rely heavily on the heterogeneity of these categories. Furthermore, she presents and defends the so-called "precision turn" in psychiatric research. Like those regarding comorbidity, issues regarding heterogeneity result from the invalidity of current psychiatric categories and the system that created those categories.

Tabb rejects what she calls "the assumption of diagnostic discrimination," which is the assumption that psychiatry's diagnostic categories are suitable explanatory targets (2015, p. 1049). In support of this thesis, Tabb enlists the work of Olbert, Gala, and Tupler (2014), who share many of Tabb's concerns. Their work is critical of the polythetic diagnostic criteria for categories such as Post-Traumatic Stress Disorder (PTSD) or MDD. The DSM-V includes long lists of signs and symptoms for these disorders, with patients fitting the criteria for a diagnosis if

they exhibit some specified amount of the listed signs and symptoms. As a result, there are 636,120 possible profiles of someone diagnosed with post-traumatic stress disorder, 227 for MDD, and 32,647 for Conduct Disorder (CD) (Olbert, Gala, and Tupler, 2014, p. 457). Two people sharing one of these diagnoses might share few or no symptoms. Additionally, the authors note that, on average, two possible symptom combinations will share fewer than half of their symptoms. For decades, researchers have had relatively little success in discovering the etiologies of these heterogeneous psychiatric disorders. These considerations lead Tabb to conclude that "little of interest from the perspective of biomedicine can be discovered about patients sharing a diagnosis" (2015, p. 1051).

Tabb offers a historical argument against this assumption of diagnostic discrimination, which rests on the observation that the DSM categories have mainly gone unchanged throughout its various iterations. The DSM began as and largely continues to be a nosology based on conventional wisdom from clinical practices (2015, pp. 1050–1051). The observation offered here is that the current psychiatric classification system creates categories according to patient symptomatology. These classification practices have produced the polythetic diagnostic criteria for the categories analyzed by Olbert, Gala, and Tupler. We should consider these classification principles and diagnostic criteria as responsible for these categories' wild heterogeneity. Consequently, we should welcome efforts to develop new classification principles. A causal classification system promises to systematically sort psychiatric phenomena into categories that provide researchers with valid targets for scientific explanation.

2.3 Research Domain Criteria (RDoC)

In response to growing concerns about the validity of current diagnostic categories, the National Institute of Mental Health (NIMH) designed the Research Domain Criteria (RDoC).

RDoC encourages researchers to investigate a particular construct relating to a functional domain along a specific unit of analysis (Insel et al., 2010). Examples of constructs include negative or positive emotionality, cognition, arousal, etc. Units of analysis refer to whether the constructs are being investigated at the genetic, cellular, neural circuitry level, etc. RDoC's website specifically lists heterogeneity and comorbidity of the sort discussed here as among the reasons for developing the new framework (About RDoC, n.d.). Proceeding from the observation that decades of research for the biological mechanisms of DSM diagnostic categories have borne little success, the framers of RDoC hypothesize that grouping phenomena differently will prove more successful. Rather than searching for genetic underpinnings of Bipolar disorder, for example, a researcher might search for genetic underpinnings of finer-grained phenomena such as negative valence or low arousal. Participants for such a study would be expected to span multiple DSM diagnostic categories, perhaps including Bipolar disorder, MDD, and GAD.

Thomas Insel, the director of NIMH at the time, recognizes that RDoC is not in itself a framework for classification: "RDoC is not a diagnostic system, it's merely a framework for organizing research" (Insel, 2014, p. 396). The goal is "to create a framework for research on pathophysiology, especially for genomics and neuroscience, which ultimately will inform future classification schemes" (Insel et al., 2010, p. 748). A new classification system, in turn, ought to foster improved patient outcomes because, according to the NIMH, diagnostic categories ought to pick out groups of people for which we can devise strategies for treatment and predict their response to such treatment. If RDoC is successful in increasing our knowledge of the pathophysiology of mental distresses, then presumably, we should be much better situated with regards to identifying disorder categories homogenous enough to make such inferences.

2.4 Endophenotypes

Beauchaine and Constantino (2017) argue that identifying and researching endophenotypes will be a fruitful endeavor for psychiatrists. Endophenotypes are physical markers at a level of description between genotype and phenotype. While endophenotypes were initially defined as disorder-specific (Gottesman and Gould, 2003), Beauchaine and Constantino argue that we should drop this requirement. Their reasoning relates to the unreliability of current diagnostic categories and to the (related) possibility that divergent phenotypes share underlying vulnerabilities. Thus defined, an endophenotype has the following properties (taken from Beauchaine and Constantino, 2017, pp. 770–771):

- They segregate with illness in the general population
- They are heritable
- They are state-independent
- They cosegregate with the disorder within families
- They present at a higher rate within affected families than in the general population
- They are capable of being measured reliably and are specific to the illness of interest

State-independence refers to when a specific sign is present regardless of whether the individual currently presents symptoms. Low brain-derived neurotrophic factor (BDNF) in bipolar individuals cannot serve as an endophenotype since low BDNF levels are only observed during manic episodes and thus are not state-independent (ibid., pp. 771–772). Though endophenotypes are heritable, they cosegregate with the disorder, meaning they will be present in family members suffering from psychiatric disorder and absent in those without that psychiatric disorder. Endophenotypes should not be conflated with pathognomonic signs, which are physical markers necessary and sufficient for demonstrating the presence of a particular disorder, or biomarkers, which are not state-independent (ibid., pp. 770, 772).

To demonstrate the potential usefulness of the endophenotype concept, Beauchaine and Constantino use the example of deficits in affect dopamine neurotransmission (DA). Noting that

some genes related to DA are expressed heavily in the striatum, they link DA deficits and deficits in striatal responses during decision-making. These deficits can be observed across diagnoses, including ADHD, MDD, and CD (2017, p. 773). Identifying DA deficits as an endophenotype serves to provide a suitably fine-grained target for research. This includes research investigating the genetic correlates of DA deficits as well as further research into the role of DA deficits in mechanisms contributing to psychiatric illness. Understanding the genetics of DA deficits could prove more achievable than understanding the genetics of a complex and heterogeneous disorder such as ADHD and could provide suitable targets for medical intervention. As with RDoC, identifying and researching endophenotypes does not offer us new principles for disease classification. Endophenotypes might be crucial for understanding abnormal mental events and processes, but how they relate to the way we classify people into disease categories is yet to be worked out. We will return to this question in section five.

3 THE EXEMPLAR MODEL OF CLASSIFICATION

The previous section motivated the need for a causal classification system by highlighting the failures and deficiencies of the current symptom-based approach. This section will explicate how the exemplar approach handles causal classification amidst causal complexity. I begin with a general exposition of the model before demonstrating how it departs from the DSM's methods.

3.1 Exemplar Model of Classification

Dominic Murphy develops and defends the exemplar model of disorder classification for psychiatry. He defines an exemplar as "a representation of the clinical features and typical course of a disorder, abstracted away from the detail of individual variation" (Murphy, 2006, p. 202). Exemplars are idealizations meant to strike a balance between representing the shared processes that cause and sustain a disorder and the diverse processes that can lead to different

manifestations of those shared mechanisms. As a classification system, the exemplar model uses exemplars to represent categories of disorder that can be differentiated on the basis of shared causal processes. Exemplars will represent etiological and pathological processes shared across instances of the disorder category but will abstract away from differences in etiology and pathology.[5] Regarding the representation of symptom profiles, an exemplar would include "all the symptoms that a patient might show" (ibid., p. 206). Using the exemplar model involves matching an individual's symptom profile to some subset of those represented by the exemplar and then determining whether the individual's disorder is caused by the etiological and pathological processes that define the category.[6]

Explanation in psychiatry involves explaining the processes causing and sustaining a disorder and explaining the relationship between these causal processes and the various manifestations that a patient may present.[7] It is in this way that exemplars have causal information built into them. The etiology of a disorder, i.e., the processes that cause the disorder, is explicitly modeled as explaining a patient's pathology, which explains a patient's symptomatology. Of course, these relationships, as represented by the exemplar, are idealizations. Causal etiological factors can only ever roughly explain a patient's pathology. A complete explanation would have to take into account all aspects of a patient's biological history. This sort of explanation would be impossible and applicable to only a single individual. For this

[5] Patients rarely, if ever, share exactly the same etiology or pathology. Exemplars abstract away from individual differences while representing shared etiological and pathological processes. This is related to explanation in the following paragraph.

[6] Often, an individual's symptom profile will match some subset of many disorders. This is precisely why symptomatology is a poor way to classify psychiatric disorders. With the exemplar model, determining the disorder a patient suffers from will come down to an understanding of the etiological and pathological processes causing and sustaining the disorder.

[7] I am intentionally vague here about what exactly these explanations might look like and what would count as an explanation. I remain open to a diverse array of explanatory models e.g., mechanistic, computational, topological explanations etc.

reason, we depend upon idealizations that allow us to make generalizations across individuals and focus on the causal difference makers. These causal difference-makers mark important ways in which a disordered individual deviates from a healthy one and are the building blocks of etiology.

The exemplar model relies on causal discrimination in at least two ways. First, the exemplar model requires us to have the means, empirically, to discriminate between causal processes. Therefore, if we desire a causal classification system, we need to be capable of determining the causes of clinical phenomena. Second, we must have the means, conceptually, to determine when we can support a distinction between two disorders on the basis of differential causal processes. The complications related to this second conceptual dimension of causal discrimination will be presented in section four and relate to causal complexity. In brief, what might be thought to be distinct disorders could plausibly share several causal processes and individual instances of what is classified as the same disorder might not always share all the same causal processes as other instances of that disorder. We will need to be precise in articulating conditions for when we should distinguish between separate disorders versus when to not distinguish disorders despite some differences in the causal processes producing them.

Murphy is relatively optimistic about our ability, empirically, to discriminate differential causal processes. Addressing the objection that a causal classification system is impractical for psychiatry because we have so little an understanding of the causes of psychiatric disorders, Murphy points out that causal understanding is not necessary for causal discrimination. For causal discrimination, all we need to know is that the causal processes underlying and sustaining a disorder are different from the causal processes underlying and sustaining other disorders. We do not need to have a complete (or even partial) understanding of the causal processes; we only

need to know they are distinct from others. Some methods useful for inferring differential causal processes include double dissociations and differential reactions to pharmaceutical interventions (Murphy, 2006, p. 322). Murphy emphasizes the importance of predictive validity in this context. If we have correctly isolated a distinct disorder, then we ought to be able to make some predictive inferences about how the disorder will unfold.

Before discussing these examples, it is important to clarify that we classify diseases, not people. The idea is not to examine a patient and decide upon the single best diagnosis for them. Rather, we attempt to detect specific causal processes occurring within the patient and classify whatever disorder(s) they are suffering from. As Murphy notes, a patient may fit several exemplars (2006, p. 346). Fitting several exemplars would mean that several distinct causal processes are relevant for explaining the patient's symptoms and the course the disorder will take.

To demonstrate how causal discrimination can be used to distinguish disorder categories, Murphy works out his theory using the examples of pediatric autoimmune neuropsychiatric disorders associated with streptococcus (PANDAS), OCD, and Sydenham's chorea. Sydenham's chorea is characterized by a very particular "dance" and results from rare complications to untreated *streptococcus* infection. After noticing that many patients with Sydenham's chorea also presented with symptoms of OCD, researchers investigated whether children with OCD also showed signs of a *streptococcus* infection. The result was the discovery of a new disorder: PANDAS. What distinguishes PANDAS from childhood OCD is the etiological agent–the *streptococcus* infection– present in PANDAS but absent in OCD. This case thus offers an example of how a causal taxonomy can distinguish between two disorders based on etiology. On the other hand, PANDAS and Sydenham's chorea share parts of their etiology but have

important differences in pathology (Murphy, 2006, 352–353). Pathology too, then, can be used to support distinctions in a causal classification system.

Having outlined the exemplar model of disease classification, I now wish to elucidate the relationship between an exemplar and an individual. We should be careful to note that there is never a perfect fit between the disease represented by the exemplar and an instantiation of that disease in any given patient. Murphy states that diagnosis involves "fitting a patient to a portion of the exemplar" (2006, p. 206). Since an exemplar contains a description of all the symptoms a patient with the condition might show, patients will only ever conform to parts of the exemplar. When relating exemplars to individuals, we should also be careful to remember that exemplars are idealizations of processes that occur within individuals. Since an exemplar involves abstraction from detail, an exemplar will never contain all the relevant information relating a patient to their condition. What is important for diagnosis is that by matching an individual to a portion of an exemplar, we can gain an understanding of the causal processes underlying the patient's condition. Doing so would then raise the possibility of investigating strategies for therapeutic intervention.

3.2 The Exemplar Model vs. DSM

As was mentioned earlier, the DSM does not utilize causal information. Still, one might wonder how radical a revision the exemplar model would prove to be. Might many DSM categories remain? Murphy acknowledges that some categories may remain but for different reasons. The categories that survive the shift to exemplar categorization are those that already pick out entities that are relatively causally homogenous. That the DSM may pick out some such categories is to be expected. Observing differences in presentation, after all, is one way of detecting different etiological and pathological processes. The problem is that this method is not

precise. Similar presentations are often causally heterogeneous and divergent presentations can sometimes share important causal processes. I explore these possibilities in section four.

Even though some categories may survive the shift to exemplar classification, those categories will exist as exemplars for very different reasons than in DSM classification. Remember that the DSM creates categories based merely on clinical presentation. This is because the many goals of the DSM are held together by the requirement of inter-rater agreement. Clinicians should be able to agree about what category or categories of disorder a patient is suffering from. While this goal has many pragmatic benefits, the heterogeneity of DSM categories, which I discussed in section two, undermines many of them. The goal of exemplar classification, on the other hand, is to create categories on the basis of shared causal processes. The hope is that doing so will group phenomena together in more productive ways.

4 CHALLENGES FOR A CAUSAL CLASSIFICATION SYSTEM

The previous section presented the Exemplar Model of classification for psychiatry. This section will introduce several challenges brought by causal classification. These challenges involve the causal complexity of mental health conditions. Section four begins with a conceptual distinction that will help us understand these causal complexities. After introducing this distinction, I will introduce empirical work from developmental psychopathology and psychiatric genetics that provide examples of causal complexity. Along the way I will evaluate how the exemplar model handles these cases.

4.1 Multicausality and Heterogeneity

Lauren Ross (forthcoming) provides a valuable discussion of what is meant by “causal complexity.” She notes that causal complexity can be assessed both at the type-level and the token-level, where a type is a particular disease or diagnosis. A token is a particular instance of

that disease (in an individual). Causal complexity at the type-level occurs when different instances (tokens) of the disease arise from different causes, meaning there is not a single cause defining the type. Ross calls this *causal heterogeneity*. Causal complexity at the token-level occurs when multiple distinct causes contribute to the manifestation of the disease in any given individual. Ross calls this *multicausality*.

To help distinguish between complexity at the type and token-levels, keep in mind that multicausality at the token level can be homogenous at the type-level. In such an instance, though there is causal complexity in the sense that multiple distinct causes contribute to the manifestation of the disease in an individual, these causes are shared across all instances of the disease. Conversely, there could be a causally heterogeneous disease — meaning that the causes of the disease are different across tokens— but always monocausal— meaning there is always a single distinct cause in a given individual. Causal complexity at the type-level does not determine causal complexity at the token-level and vice versa. Many psychiatric phenomena exhibit both causal complexity and causal heterogeneity. Section 4.2 explores instances of multicausality and causal heterogeneity as researched in developmental psychopathology.

4.2 Developmental Psychopathology

Developmental psychopathology studies the development of psychiatric disorders from childhood into adulthood, focusing on neurodevelopment (Beauchaine, Constantino, and Hayden, 2018). Central to the practice of developmental psychopathology are *equifinality* and *multifinality*. Equifinality describes instances where different starting points and processes reach the same outcome. Multifinality occurs when shared or similar causal factors reach differential outcomes (Cicchetti and Rogosch, 1996; Beauchaine, Constantino, and Hayden, 2018). In other words, multifinality describes variations in how some entity functions in relation to the

environment in which that entity operates. The methods and commitments of developmental psychopathology are helpful for understanding how we might make sense of causal complexity and heterogeneity. One particular area where developmental psychopathology has made progress is in understanding and predicting the development of externalizing disorders through childhood, adolescence, and adulthood. I will now turn to this literature.

To highlight the interdisciplinarity of developmental psychopathology approaches and to illustrate their utility for psychiatry, Beauchaine, Constantino, and Hayden (2018) utilize research into the risk-conferring genetic and environmental influences of attention deficit hyperactivity disorder (ADHD) and conduct disorder (CD). Their neurodevelopmental model of externalizing behavior centers around disrupted striatal activity. It considers genetic risk factors, epigenetic regulation of neural function, neurohormonal changes, environmental factors such as *in utero* hypoxic events, and social factors such as wealth and crime levels. Genetic risk factors, epigenetic influences, and neurohormonal deficits together account for an individual's liability to developing some trait, in this case, blunted striatal function.[8] Abnormal striatal functioning results in irritable, anhedonic psychological states, resulting in impulsive, reward-seeking behaviors (147).

The move from similar or shared genetic and neural risk factors to multifinal outcomes in Beauchaine, Constantino, and Hayden's model is accounted for by differential neurodevelopmental maturation and social-environmental conditions. In particular, they argue that "[a]lthough specific neural networks differ, both neuromaturation of the PFC and efficiency of sub-cortical connectivity become increasingly compromised across development in

[8] "Liability" refers to an individual's total risk for showing a disease phenotype. Individual risk factors, e.g. a genetic mutation, increase one's liability by a certain degree. Some risk factors will confer only a small increase in one's liability, while others will confer a larger risk.

internalizing and externalizing disorders" (149). This is important because it pinpoints the locus of the distinction between ADHD and more progressed externalizing disorders, such as CD, oppositional defiant disorder (ODD), or even substance abuse. More progressed externalizing disorders are associated with disrupted PFC maturation and subcortical connectivity beyond that seen in ADHD. The question then becomes focused on the causal factors that give rise to the differential outcomes, i.e., those with "just" ADHD and those with more progressed externalizing disorders. The authors' model emphasizes social-environmental factors such as parenting style and socio-economic factors such as poverty, neighborhood crime rates, and educational opportunities. The model refers to such factors as *risk mediators* (149).

The model developed by Beauchaine and Constantino allows us the opportunity to understand the challenges involved in crafting a causal classification system. Their model is multicausal: A variety of genetic factors combine with social-environmental risk mediators to determine one's liability to a phenotype. Faced with multicausality, the challenge for a causal classification system becomes determining which causes will demarcate one disease category from another. Some questions are whether and to what extent might two disease categories share causes? Do we always demarcate disease categories according to their cause(s) at, say, the genetic "level," or might some categories be distinguished based on more proximal causes? If we put aside the complexities with the genetic, neurohormonal, and (nonsocial) environmental factors, the externalizing disorders might seem relatively homogenous until the onset of neurodevelopmental delays in the PFC.

The causal complexity of Beauchaine and Constantino's developmental model of the externalizing disorders affords us an opportunity to disentangle how exemplars abstract away from the details of individual variation and how they distinguish the importance of various

causes for classification. While the genetic and neurohormonal causes of the externalizing disorders are similar and shared across individuals, pointing to any single or small subset of causes contributing to these disorders would entail a certain level of abstraction. Those causes picked out would have to be shared across all instances but would not fully explain any one individual's disorder, since each individual will exhibit multicausality and there will be some degree of causal heterogeneity across individuals. Such causal complexities are the very thing the exemplar model allows us to abstract away from. By focusing on a single or small subset of causes shared across individuals, we can make some sense of phenomenon that might otherwise be impossibly complex. Doing so also has the benefit of providing researchers with a more specified target for explanation. If this target can be understood, then clinicians may be able to intervene and relieve patient suffering.

The exemplar model thus offers the conceptual tools to support a distinction between ADHD and more progressed externalizing disorders because it allows us to focus on disrupted PFC maturation as a causal factor that is shared amongst those that depart from the ADHD exemplar while abstracting away from the details of individual variation. The basis for this distinction relates to differences in the disorder's course and outcome. Since individuals with delayed PFC maturation have notably different experiences and outcomes from others in the exemplar, we can establish delayed PFC maturation as the locus of a distinct sub-group within the larger category of externalizing disorders. Other causes, such as social-environmental factors might also provide plausible loci for distinguishing sub-groups if classifying them as such allows us to develop better targets for explanation and intervention.

4.3 Genetic Risk for ASD

That ASD is highly heritable, and thus demands at least a partial genetic explanation, is evidenced by familial clustering of ASD (Devlin and Scherer, 2012). Estimates of the risk that genetics confer on ASD manifestation range from ASD contributing between 60% and 90% of total liability (Gaugler et al., 2014; Constantino, 2018). Studying the genetics of autism spectrum disorder (ASD) provides an opportunity to explore both multicausality and causal heterogeneity. Additionally, the messiness of understanding the genetics of ASD affords an appreciation of the difficulty in relating genetic factors to disease phenotypes. Much of the genetic risk for ASD is shared with other conditions. To help understand these complexities, I begin this section with some background on two competing theories in genetics research: the common disease common variants hypothesis and the common disease rare variants hypothesis. I introduce the concept of missing heritability before reviewing research on the genetics of autism.

The search for risk-conferring genetic variants of complex diseases has historically been driven by adherents to one of two hypotheses.[9] The first is the common disease common variants (CDCV) hypothesis; the second is the common disease rare variants (CDRV) hypothesis. The CDCV hypothesis predicts that the risk-conferring genetic variants for complex diseases will consist of many common variations with low-penetrance acting additively to manifest the disease phenotype.[10] The CDRV hypothesis predicts that rare variations with high penetrance are primarily responsible for disease manifestation (Schork et al., 2009).[11] While it is accepted that

[9] "Complex" or "multifactorial" diseases are those that are caused by a number of contributing factors, spanning from genes to the environment. In this way, multifactorial diseases already conform to Ross's notion of multicausal.
[10] "Penetrance" is determined by the relation between the presence of a gene variant and the manifestation of a phenotype. A 1:1 correspondence is fully penetrant.
[11] "Common" variants are those that occur at a rate greater than 1% of the population. "Rare" variants are those that occur at a rate lower than 1% of the population.

both common and rare variants contribute to one's overall genetic liability, debate still surrounds the relative role each plays (ibid.).

State and Levitt (2011) explore the CDCV and CDRV hypotheses as they relate to ASD. Starting from the observation that studies have historically overestimated the effects of specific alleles, State and Levitt introduce the concept of *missing heritability*. Missing heritability occurs when the identified genetic risk factors only account for a portion of one's liability. The authors note several potential explanations for missing heritability, all of which appear to have some accuracy to them. First, it is possible to have overestimated the role of genetic variations in producing the phenotype. To the extent that this is correct, we should emphasize the search for environmental and developmental factors that interact with genetic risks to produce a disease phenotype. Second, the CDCV hypothesis could be correct, and we could have, as of now, failed to identify most of the genetic risk factors (all of small effect size). Third, the CDRV hypothesis could be correct, and we could have, as of this study, failed to identify the relevant rare variations.

Taking these possibilities for explaining missing heritability and applying them to the case of ASD, State and Levitt review what we know about contributions from rare and common genetic variations. Genome-wide association studies (GWAS) have so far failed to identify promising replicable results regarding common variants. More promising results have come from studying rare *de novo* copy number variants. However, studies have still been unable to identify variants that can explain more than a small subset of individuals. The authors note two studies that independently identify roughly 300 *de novo* CNVs related to ASD (p. 1502).

In summary, to the extent that psychiatric disorders are similar to ASD, we should expect several things. First, we should *not* expect to find fully penetrant genetic variations. This will

prevent us from identifying necessary and sufficient genetic causes of psychiatric disorders. We *should* expect what Ross refers to as causal heterogeneity: we should expect that the genetic factors contributing to the manifestation of psychiatric disorders are not homogenous across instances of the same disorder. Thus, even though genetic variations are causes of psychiatric disorders, it will be challenging to create a causal classification system that relies too heavily on demarcating disorders based on their genetic causes.

Before moving on, I need to respond to an important objection. The reader may object that the complexities I have been exploring are artefacts of research directed at an unsuitable target. After all, earlier I pointed to significant problems with categories such as ASD. If we had properly defined disease categories, could we expect better results and more homogeneity? I believe this objection to be misinformed. While a better classification system might lead to a clearer understanding of the genetics of psychiatric disorders, I believe it is likely that causal complexity at the genetic level will always be present. In a recent paper, Craver et al. argue that the relative lack of success in using genome-wide association studies (GWAS) to identify genes of interest in psychiatric disorders results from poor targets *and* the inherent complexity of psychiatric disorders (Craver et al., 2020). The inherent complexity of psychiatric disorders is explained to be a result of the "soft construction" of the cognitive capacities involved in psychiatric disorder. Soft construction refers to developmental processes involving decentralized mechanisms, resulting in a multitude of ways for an organism to reach a functional end state (Ibid., pp. 3–5).

For reasons pointed to by Daniel Weiskopf (2017), the exemplar model does not have the resources to make sense of disorders that exhibit large amounts of causal heterogeneity. To make sense of causal complexity, the exemplar model relies on selection (of certain causes) and

abstraction. The causes we choose to select and use to define a category must be present across individuals of the category and also capable of explaining a large amount of those individuals' disorder processes and symptoms. While this is compatible with some multicausality and causal heterogeneity amongst the various other causes producing and sustaining the disorder across individuals, this minimum level of homogeneity will be necessary. Weiskopf's contention is that, in the case of ASD, finding even this minimum amount of homogeneity is unachievable. On Weiskopf's reading, there are no causal mechanisms that can explain any meaningful subset of ASD patients, and thus no causal basis for productive distinctions within ASD (Weiskopf, 2017, pp. 181–182).

As an alternative, Weiskopf (2017) suggests that ASD can be understood as a network structure. The network model of psychiatric disorders takes a disorder category to consist of many loosely connected patient profiles. Different profiles are linked by shared properties at some level of analysis (i.e., genetic, molecular, neural, cognitive levels, etc.). For example, two patients with ASD might share some, but not all, of the same genetic vulnerabilities. One of those two patients might share some, but not all, of their genetic vulnerabilities with a third patient, but that third patient might not share any of the same genetic vulnerabilities with the other of the original two patients. The disorder category would be represented by the set of networks representing these linked profiles at each level of analysis.

Network models and exemplars are not unrelated. Any possible patient profile should, in principle, be derivable from a completed network representation. An exemplar model would represent the particular subset of causal processes, at each level of analysis, responsible for producing and sustaining that patient's disorder as well as the relations between causal processes at various levels. Any one of these particular patient profiles could be the basis of distinguishing

a particular exemplar. The utility of exemplar categories, however, rely on the ability to choose some subset of causal processes capable of explaining how a substantial subset of individuals diverge from the larger category either in terms of etiology or pathology.[12] An exemplar that explains only a single or few individuals will not be worth distinguishing for classification purposes. Network approaches do not require any such homogeneity. Individuals may have more or less similar profiles across levels of analysis. The network model, therefore, accords more closely with the soft construction developmental model Craver et al. use to explain the inherent heterogeneity of psychiatric disorders.

5 THE UTILITY OF CLASSIFICATION MODELS

In the previous sections, I motivated the need for a causal classification system for psychiatry, explored two approaches to causal classification, and explored challenges involved in causal classification. Namely, the motivation is that current classification system, which sorts phenomena according to clinical presentation, produces invalid categories. High comorbidity and heterogeneity are evidence of the invalidity of these categories. Invalid disorder categories have consequences for our abilities to explain and treat mental disorders. A causal classification system would provide better explanatory targets and allow clinicians to develop more precise and effective interventions. The challenges to constructing a causal classification system have to do with causal complexity. Mental disorders exhibit both multicausality and causal heterogeneity. In this section, I will argue that classification systems are a tool we have for

[12] Whether the number of individuals a possible exemplar applies to represents a "substantial subset of individuals" has to do with the proportion of individuals the exemplar can apply to in relation to the size of the population exhibiting the broader disorder category. If some exemplar can reliably distinguish, say, 10% of the population exhibiting some broadly defined disorder category, then it might be a worthwhile distinction to make. There is no exact number or percentage that makes a subset substantial, however.

understanding the world and evaluate the utility of the exemplar and network models in satisfying three purposes of classification.

5.1 Classification Systems as Tools

Classification systems are not typically thought of as tools, but I argue that they are. We can think of the principles of a classification system as analogous to the shape of a tool. The shape, or structure, of a tool determines its function. Classification systems function to sort phenomena in particular ways.[13] We use classification systems to produce categories according to the system's principles of classification. These products have practical utility. By dividing up the world in a certain way, we are able to do a number of things. Depending on how the system divides up the world, categories may be used to make inferences, make predictions, serve as explananda, or allow us to better visualize the phenomena we are working with.

The benefit of thinking about classification systems as tools comes once we realize that tools are utilized in relation to some purpose. Different tools function differently, and particular tools are well-suited for certain tasks and poorly suited for others. The same is true of classification systems. Certain ways of categorizing phenomena in psychiatry will be helpful for some purposes more than others. Given the variety of goals within psychiatry, e.g., constructing explananda, predicting disease course and outcome, predicting response to treatment, understanding disease processes, etc., we might expect that some classification systems are helpful for some of these endeavors and less so for others. The rest of this section will be spent demonstrating that this is the case: some classification systems are good for some purposes and bad for others. Therefore, I argue that we should be pluralists about classification in psychiatry.

[13] Of course, not all tools function the way they are meant to. It's plausible that some tools are designed with the intention of having them perform some function that they then perform poorly. This would be a poorly designed tool. The DSM could be said to be a poorly designed tool.

5.2 Variety, Purpose, and Value

There are a number of potential reasons to classify phenomena into categories. The chief aim of psychiatric research and clinical practice is, and ought to be, treatment. Of course, there are a number of possible routes to developing better treatment outcomes. Any route, however, will involve utilizing a classification system that allows us to represent psychiatric phenomena in a way productive for the goals of treatment. The following is a list of desiderata that classification systems might satisfy as a means of representing categories productively for the purposes of psychiatry:

- Categories might make it easier for us to comprehend the relations between complex phenomena
- Categories may serve as explanatory targets
- Categories may be useful for communicative purposes such that by telling someone that John can be categorized as *x*, I pass along certain information about John

The first two desiderata are particularly important for the goals of treatment. Understanding the relations between complex phenomena situates researchers to develop intervention strategies that focus on shared vulnerabilities or, alternatively, might help us understand why some treatment strategies only help some patients with a given disorder. Similarly, good explanatory targets are productive for the goals of treatments because they help facilitate the development of explanations that may provide the basis for new intervention strategies. The final desideratum is particularly important in clinical contexts when deciding upon particular intervention strategies and generating predictions about patient course and outcome. I will argue that the exemplar and network models serve these interests in different ways. In some cases, one may be superior to the other in relation to one of these purposes and/or perform poorly in relation to another. In other cases, the two models may serve one particular purpose in different ways without being superior

to another on the whole. The remainder of this section examines how the exemplar and network models serve each of these purposes.

As a model of causal classification, the exemplar model has merits. The exemplar model allows us to classify disorders with attention to causal processes without the need to consider the ways these causal processes might manifest differently. This has the benefit of producing more homogenous categories than DSM categories provide. Of course, the homogeneity of the categories may not be reflected at the symptoms level, since the same or similar causal processes may present differently across individuals. Even so, obtaining homogeneity of causal processes is a marked improvement over DSM categories, which fail to be homogenous in any meaningful sense. Plausibly, individuals sharing a DSM diagnosis often fail to share symptoms or causal processes.

Categorization by exemplar also has the benefit of eliminating artificial comorbidities. Artificial comorbidities occur when a classification system produces two or more categories that may manifest as a result of the same or similar causal processes. As a result, individuals whose condition is caused by some distinct causal process may manifest symptoms that fit multiple diagnoses. For example, if the DSM categories MDD and GAD can both involve some causal process(es) *x*, then some individual whose condition is caused by *x* may receive both the MDD and GAD diagnoses. If causal process(es) *x* is used to distinguish a specific disorder, on the other hand, then artificial comorbidities such as these would not be a problem.

Together, these benefits make the exemplar model beneficial for the purposes of providing research targets and for communicative purposes. As research targets, exemplars provide clearly demarcated explananda that are also causally homogenous to some degree. This makes them well-suited for mechanistic explanations and pharmaceutical interventions in

particular. As a communicative tool, the exemplar model produces categories that communicate more about individuals fitting those categories as we begin explaining and understanding the causal processes that distinguish the exemplar. Even an unexplained exemplar carries information about the typical course of a disorder and the possible symptoms a patient may suffer from. As we increase our understanding of the causal processes distinguishing an exemplar, the category will provide even richer communicative information, such as what intervention strategies patients might respond to or new methods for confirming a diagnosis.

The exemplar model proves less useful for the first of the three purposes. As discussed in section four, a drawback of the exemplar model is that it has a more difficult time distinguishing and representing causal heterogeneity. In certain cases, extreme causal heterogeneity makes it particularly difficult to identify and understand links between causal processes at various levels of explanation and patient problems. Distinguishing exemplars within such a context may not be feasible. Even when we can distinguish various causal processes and typical courses of a disorder, relating these to one another in the ways proposed by the exemplar model requires a level of regularity and consistency between causal processes and outcomes that is simply not present with some psychiatric phenomena. In these instances, there may be relations between phenomena that are worthwhile for exploring but that are not represented by the exemplar model.

In comparison, network models, do well to represent the diverse ways broadly defined disorder categories can relate to one another at various levels of analysis. The network model is also helpful for representing mental disorders in ways that are productive for investigating different kinds of interventions than naturally suit the exemplar model. For example, one thing we might learn from network models is whether various cognitive therapies targeting commonly linked symptom clusters could be helpful for ameliorating a variety of patient problems. The

relevant information for these kinds of intervention strategies is represented by network models as statistical relations, rather than as causal processes. Craver et al. (2020) have a similar idea of how we might pursue intervention strategies in contexts of genetic complexity. They describe the prospect of a *minimum s-t cut*, which describes the minimum number of links within a network that would need to be "severed" in order to alter the output value (the disordered state of the patient) (p. 9). For example, if network models help us to understand that some subsets of genetic vulnerabilities are responsible for some significant proportion of one's liability for a disorder, then those vulnerabilities become the intervention targets.

Since the exemplar and network models are differentially well-suited for various kinds of phenomena and explanations, neither can be said to be better, on the whole, than the other at providing explanatory targets. Both systems sort psychiatric phenomena in productive ways for the purposes of understanding and intervening upon psychiatric disorders and different approaches to explanation and treatment should be embraced. The drawback of the network model is its utility for communicating information about individuals fitting its categories. Since it represents disorders as networks of loosely connected symptom profiles and causal processes, knowing that some individual has disorder *x*, where disorder *x* is represented as a network, tells us relatively little definitive information about how the patient presents clinically, what processes are sustaining their illness, or about possible intervention strategies without additional investigation.

In this section I've argued that the exemplar and network models of classification are both useful in relation to three desiderata. I have, however, been largely focusing on how classification can be a useful tool for research purposes. While research is obviously an important component of psychiatric science, the other component is practicing psychiatry.

Psychiatric practice takes place in the context of a clinic. Classification in clinical settings presents distinct challenges. Clinicians have limited information upon which to base a decision about how to classify their patients. That same complexities that make the DSM approach to classification a poor one also makes it difficult for clinicians to classify their patients using causal models. Namely, most of the information clinicians have relates to the symptoms their patient is suffering with. The problem, as has been stressed throughout this thesis, is that knowledge regarding symptoms alone very rarely tells us anything distinctive about the nature of the patient's illness or how to treat it.

In addition to having limited information, clinicians also have pragmatic concerns that do not come up in research contexts. For one, each patient is different and, in many ways, unique. Since classification always involves abstracting away from the details of individual variation, patients will always have more complex presentations than can be represented by a classification system. Also at issue is that clinicians have to classify patients for billing and records purposes. This means that clinicians have to classify their patients despite often only having limited information about their patients' problems. As such, the classifications patients receive in the clinic do not always communicate or need to communicate precise information about the processes causing and sustaining their problems or how their problems relate to various other causal processes and symptom profiles. Primarily, clinicians need to classify their patients according to a basic description of the problems the patient presents with. Additional detail is usually detailed in the patients "chart."

Despite the inevitable challenges that clinicians will continue to face when classifying what their patients suffer from, improved classificatory practices can still benefit psychiatric practice. There is good reason to believe that new principles for classification will do well to

improve our understanding of psychiatric disorders. Better explanatory targets and more clearly represented disorder categories should help facilitate more productive research. Explaining mental disorders and devising successful intervention strategies could finally provide the therapeutic relief for which patients are long overdue.

REFERENCES

About RDoC. (n.d.). National Institute of Mental Health. https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/about-rdoc.shtml

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Andrews, G, Goldberg, D. P., Krueger, R. F., Carpenter Jr., W. T., Hyman, S. E., Sachdev, P., Pine, D. S. (2009). Exploring the feasibility of a meta-structure for DSM-V and ICD-11: could it improve utility and validity? *Psychological Medicine, 39,* 1993–2000.

Beauchaine, T. P. and Constantino, J. N. (2017). Redefining the endophenotype concept to accommodate transdiagnostic vulnerabilities and etiological complexity. *Biomarkers in Medicine*, *11*(9), 769–780.

Beauchaine, T. P., Constantino, J. N. and Hayden, E.P. (2018). Psychiatry and developmental psychopathology: Unifying themes and future directions. *Comprehensive Psychiatry, 87,* 143–152.

Cicchetti, D. and Rogosch F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology, 8,* 597–60.

Constantino, J. (2018). Deconstructing autism: from unitary syndrome to contributory developmental endophenotypes. *International Review of Psychiatry, 30*(1), 18–24.

Cooper, R. (2005). *Classifying Madness: A Philosophical Examination of the Diagnostic and Statistical Manual of Mental Disorders*. Springer.

Cooper, R. (2014). *Diagnosing the Diagnostic and Statistical Manual of Mental Disorders*. Karnac.

Craver, C. F., Dozmorov, M., Reimers, M., and Kendler, K. S. (2020). Gloomy Prospects and Roller Coasters: Finding Coherence in Genome Wide Association Studies. *Philosophy of Science, 87*, 1–12.

Devlin, B. and Scherer, S. W. (2012). Genetic architecture in autism spectrum disorder. *Current Opinion in Genetics and Development, 22*, 229–237.

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J., Ripke, S., Sandin, S., Sklar, P., Svantesson, O., Reichenberg, A., Hultman, C. M., Devlin, B., Roeder, K., and Buxbaum, J.D. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics, 46*(8), 881–885.

Gottesman, I. I. & Gould, T. R. (2003). The endophenotype concept in psychiatry: Etymology and strategic interventions. *American Journal of Psychiatry, 160,* 636–645.

Insel, T. R. (2014). The NIMH Research Domain Criteria (RDoC) Project: Precision Medicine for Psychiatry. *American Journal of Psychiatry, 171,* 395–397.

Insel, T. R., Cuthbert, B. N., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., and Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry, 167*(7):748–751.

Kendell, R. & Jablensky, A. C. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry, 160,* 4–12.

Lecavalier, L., McCracken, C. E., Aman, M. G., McDougle, C. J., McCracken, J. T., Tierney, E., Smith, T., Johnson, C., King, B., Handen, B., Swiezy, N. B., Arnold, L. E., Bearss, K., Vitiello, B., Scahill, L. (2019). An exploration of concomitant psychiatric disorders in children with autism spectrum disorder. *Comprehensive Psychiatry, 88,* 57–64.

Murphy, D. (2006). *Psychiatry in the Scientific Image*. MIT Press.

Olbert, C. M., Gala, G. J., Tuppler, L. A. (2014). Quantifying Heterogeneity Attributable to Polythetic Diagnostic Criteria: Theoretical Framework and Empirical Application. *Journal of Abnormal Psychiatry, 123*(2), 452–462.

Ross, L. N. (forthcoming). Explanations in Contexts of Causal Complexity: Lessons from Psychiatric Genetics. *Minnesota Studies in Philosophy of Science*.

Schorck, N. J., Murray, S. S., Frazer, K. A., and Topol, E. J. (2009). Common vs. Rare Allele Hypotheses for Complex Diseases. *Current Opinion in Genetics and Development, 19*(3): 212–219.

State, M. W. and Levitt, P. (2011). The conundrums of understanding genetic risks for autism spectrum disorders. *Nature Neuroscience, 14*(12), 1499–1506.

Stone, C. (2019). A Defense and Definition of Construct Validity in Psychology. *Philosophy of Science, 86,* 1250–1261.

Tabb, K. (2015). Psychiatric Progress and the Assumption of Diagnostic Progress. *Philosophy of Science, 82,* 1047–1058.

Tabb, K. (2019). Philosophy of psychiatry after diagnostic kinds. *Synthese, 196,* 2177–2195.

van Loo, H. M., Romeijn, J. W., de Jonge, P., Schoevers, R. A. (2013). Psychiatric comorbidity and causal disease models. *Preventive Medicine, 57,* 748–752.

Weiskopf, D. A. (2017). An ideal disorder? Autism as a psychiatric kind. *Philosophical Explorations, 20*(2), 175–190.