Forthcoming in *Philosophy of the Social Sciences*

Social Construction, HPC Kinds, and the Projectability of Human Categories

Jonathan Y. Tsou
Department of Philosophy and Religious Studies, Iowa State University
Email: jtsou@iastate.edu

Abstract:

This paper addresses the question of how human science categories yield projectable inferences by critically examining Ron Mallon's 'social role' account of human kinds. Mallon contends that human categories are projectable when a social role produces a homeostatic property cluster (HPC) kind. On this account, human categories are projectable when various social mechanisms stabilize and entrench those categories. Mallon's analysis obscures a distinction between transitory and robust projectable inferences. I argue that the social kinds discussed by Mallon yield the former, while classifications of biological kinds yield the latter. Classifications from psychiatry ('schizophrenia,' 'hysteria') are discussed as examples.

Keywords: Ron Mallon, projectability of human categories, natural kinds, social constructionism, realism

1. Introduction

In his excellent and provocative book, *The Construction of Human Kinds*, Ron Mallon (2016) attempts to demonstrate how social constructionism regarding human science classifications can be consistent with a broadly naturalist and realist metaphysics. In particular, he aims to show how socially constructed human categories or kinds (e.g., race, gender, homosexuality) can be *real and causally predictive*. Within this project, Mallon proffers an explanation for the projectability of socially constructed human categories, i.e., how human categories that are socially constructed can yield reliable inductive inferences about members of kind. On Mallon's account, socially constructed categories are projectable when a 'social role' creates or produces a homeostatic property cluster (HPC) kind that is stabilized and entrenched by certain social mechanisms.

In this paper, I argue that Mallon's social role account of the projectability of human kind categories obscures and fails to recognize a distinction between transitory and robust projectable inferences, and his account of socially constructed kinds can only explain the former. The more orthodox realist view that human science categories are projectable when such categories accurately individuate natural or biological kinds offers a more compelling explanation for the robust projectable inferences we can draw from certain human science categories (e.g., schizophrenia, bipolar disorder). While the social kinds implied by Mallon's social role account may yield reliable inductive inferences about members of a kind over small time-scales, for a human science classification to yield stable and robust projectable inferences over longer time-scales, the classification must—at least partially—pick out a natural or biological kind.[1] In

---

[1] My argument presupposes the metaphysical assumption that cultural evolution moves faster than biological evolution. For critical discussion of evidence in favor of this assumption, see Richerson and Boyd (2005), Richerson, Boyd, and Henrich (2010), and Perreault (2012).

sections 3 and 4 of the paper, I argue that classifications of natural kinds yield robust and stable

projectable inferences, while classifications of social kinds yield more transitory and unstable

projectable inferences. In this framework, the projectable inferences associated with

classifications of social kinds are contingent, while the inferences associated with classifications

of natural kinds approach the philosophical ideal of necessity. I support this argument with

examples of psychiatric kinds, such as schizophrenia, bipolar disorder, and hysteria. The robust

and ampliative projectable inferences that we can draw from classifications like schizophrenia

(e.g., psychosis can be treated with dopamine antagonist drugs) are explained by the fact that

these classifications individuate natural kinds whose members share similar (intrinsic) biological

properties. By contrast, classifications of purely social kinds, such as hysteria, yield more

transitory projectable inferences. This implies that human and social sciences that aim to produce

predictive human classifications (e.g., medicine, psychology, psychiatry) should classify natural

kinds.[2] The distinction between robust and transitory projectability also suggests that Mallon's

(2016) account of the projectability of human kinds could be fruitfully expanded by considering

how socially constructed categories are sometimes partially constructed by natural properties that

ground them, which is a possibility he is not unsympathetic to (210-13). Nothing in my analysis

is intended to vitiate the idea that human science classifications are socially constructed relative

to particular human interests, decisions, social practices, and historically-situated cultures. What

it does suggest, as a corrective, is that despite such social constraints, some human classifications

---

[2] My analysis is not intended to imply that the social role kinds discussed by Mallon are unworthy of study in the social sciences. What it does imply is that these classifications will not yield the robust and stable projectable inferences that classifications of natural kinds will. For social sciences that are less interested in making predictions and more interested in explaining and understanding human phenomena (e.g., sociology, anthropology), the investigation of social kinds can be extremely fruitful, especially in revealing the mechanisms involved in the stabilization and destabilization of these categories. Indeed, one of the major benefits of Mallon's book is the illumination of such issues.

pick out (more or less) *naturally-occurring classes* (cf. Dupré 1993, 2001), which is ultimately responsible for the capacity to yield robust and stable projectable inferences.

2. Mallon on the Projectability of Human Kinds

Mallon's account of the construction of human categories focuses on 'social roles.' According to Mallon (2016), social roles are structured by *representations*, i.e., "attitudes, theories, narratives, concepts, models pictures, norms, rules, utterances, inscriptions, and texts by which we represent the world and structure our actions" (6), whereas *categories* are what representations are about (e.g., properties or classes in the world), and *human kinds* are projectable categories that support inductive inferences, predictions, and explanation. In this framework, a social role exists if: (1) there are representations that pick out a category of persons and a set of beliefs and evaluations associated with that term, and (2) many of the beliefs and evaluations associated with the role are common knowledge within society. Condition (2) is achieved through the public broadcasting of information concerning the social role, e.g., by scientific institutions, media reports, or other credible sources.

For a social role to become a human kind, it is first necessary for the social role to become entrenched in society and culture. In this account, human kinds are not biological kinds (although they are often thought to be), but social kinds ('social roles') whose reality is constituted by the various representations and human practices that sustain and reinforce these roles.[3] Mallon (2016) writes:

---

[3] Throughout his book, Mallon (2016) makes it clear that the social roles he is discussing are *covert social roles*: social roles (e.g., race or gender) that are regarded as the product of natural facts, rather than human classificatory or social practices. His account maintains that although these categories are thought to be natural kinds, they are in reality social kinds (chs. 2, 3, 8, and 9).

[A] category (e.g., race or gender or dissociative identity disorder) may not be a biological kind, but it is not nothing either. It can be a real and important kind structured and sustained by the representations of the category, and by the accumulated [causal] effects of such representations. (93)

The idea is that social roles can be *entrenched* by various mechanisms and human practices surrounding representations, which makes that social role a real and causally significant human kind (Mallon 2016: ch. 3). For example, Mallon suggests that looping effects (Hacking 1995) can serve to entrench a social role, when the emergence of a category (e.g., multiple personality) causes people—for strategic or non-strategic reasons—to act in accordance with the stereotypical features associated with that role. On this view, social roles are entrenched by both those who adopt a social role and by the people (e.g., doctors, psychiatrists, the public) who recognize and use the social role term in various social practices.[4] The entrenchment of social roles may also be facilitated by human cognitive mechanisms, such as innate automatic psychological processes (e.g., tendencies towards essentialist thinking or grouping on the basis of stereotypes) or acquired psychological processes (e.g., automatic responses that are shaped and reinforced by culture). On a broader level, social roles can be entrenched by modifications to the social environment or culture (e.g., residential segregation in the United States, formal institutions specifying definitions for some category), which produce feedback to both social role representations and the people represented. All of these mechanisms—through their systematic causal effects—serve to entrench existing social roles, which structure the behavior and expectations for those who recognize and use the social role classification (e.g., psychiatrists, the public) and for those who fall under the classification (e.g., an individual diagnosed with a mental disorder).

---

[4] Whereas Mallon focuses how looping effects stabilize a social role, Hacking (1999; 2007) also emphasizes how looping effects can destabilize human kind classifications (cf. Mallon 2016: ch. 7).

Mallon argues that when social roles become entrenched via various social mechanisms, these human categories are transformed into human kinds that yield projectable inductive inferences. As Mallon (2016) puts it: "I argue that … social roles categories can be *human kinds* … that support induction, prediction, explanation, and intervention in some or another of our inductive projects … by arguing that the kinds amount to *homeostatic property-cluster kinds*" (89, emphasis in original). In Mallon's account, entrenched social roles or human kinds are accurately described by what Richard Boyd (1991, 1999a) calls homeostatic property cluster kinds or HPC kinds. Boyd's account of HPC kinds is a non-essentialist theory of natural kinds, in which members of a kind share a cluster of similar properties, but no property is essential for membership in a kind. According to Boyd (1999a: 143-144), the key characteristics of HPC kinds are the following:

(1) There is a family of properties (F) that are contingently clustered in nature.

(2) Their co-occurrence is the result of what may be described as homeostasis: either the presence of some properties tends to favor the presence of others, or there are underlying mechanisms that tend to maintain the properties in F, or both.

(3) There is a kind term (t) that is applied to things in which the homeostatic clustering of most of the properties in F occurs.

Unlike essentialist natural kinds that are defined by necessary and sufficient conditions (e.g., H_2O, electrons), HPC kinds refer to the messier kinds of classes that are studied in the biological and human sciences (e.g., species, diseases). Boyd's HPC account of natural kinds is motivated in large part to explain the projectability of non-essentialist natural kind classifications. On this account, it is the shared, law-like causal structure—i.e., the homeostatic mechanisms referred to

in (2) that cause properties to cluster in a regular and non-accidental way— that allows for successful inductive inferences about members of a kind (Boyd 1985, 1991, 1999a, 2010).[5]

Mallon finds Boyd's account of HPC kinds apt for characterizing entrenched social roles because it is sufficiently liberal to account for the non-essentialist nature of human social roles or categories. For Mallon, it is particularly significant that the HPC account liberalizes the types of properties that can be included in the property cluster in (1) to include relational properties, in addition to intrinsic ones (cf. Boyd 1991: 140-143), which "opens the door to properties that result from human practices, norms, conventions, and so forth" (Mallon 2016: 92). That is, the HPC account can accommodate clusters of social properties that result from the sorts of social, psychological, and environmental mechanisms that Mallon suggests sustain and entrench social roles. If entrenched social roles (or human kinds) are accurately characterized as HPC kinds, then this could explain the projectability of human kinds. Mallon (2016) writes:

> [C]ausally significant social roles … could figure as the homeostatic mechanism at the center of important property cluster-kinds that structure our social world. Of course, whether a particular type of social role does constitute an explanatory kind is an empirical question, but social roles that do produce and sustain property-cluster kinds may support induction, prediction, and explanation. (92, emphasis added)

Mallon's account of how a social role is supposed to *produce* an HPC kind is somewhat murky, but the basic idea is that human kind terms are projectable insofar as the property clusters associated with these terms are sustained by reiterated representations of a social role and mechanisms (e.g., social feedback mechanisms, psychological mechanisms, environmental

---

[5] Craver (2009) and Khalidi (2013) argue that the HPC theory could dispense with the homeostatic mechanisms requirement, but both agree with the general realist argument advanced in this paper that the kind term needs to *correctly describe the causal structure of the world* to offer projectable inferences (cf. Slater 2015).

mechanisms) that serve to entrench that social role. Part of the unclarity in this account may be due to the fact that Mallon (2016) only offers his account as a "how-possibly" (89) model of how socially constructed human kinds are projectable. That is, Mallon's account is only intended to articulate the general conditions that need to be met for a human kind classification to be projectable (viz., the production of an HPC kind), and his argument is that entrenched social roles can meet these conditions.

In sections 3 and 4 of this paper, I argue that a condition that is neglected in Mallon's account is that some of the properties in the property cluster individuated by an HPC kind term must be *intrinsic or natural (e.g., biological)*. I suggest that this condition must be met for an HPC kind term to yield the sort of non-trivial, robust projectable inferences typically demanded of scientific classifications. Mallon's account of the projectability of human kinds rests on the idea that social roles, when entrenched by various mechanisms, can produce an HPC kind. It is important to notice that the relevant HPC kinds in Mallon's account are *social kinds* and the proper referents for human kinds are *social constructs* (see Mallon 2016, ch. 8; cf. Haslanger 2012, ch. 3). Mallon (2016) briefly contrasts his social role account of human kinds with a realist account that holds that some "human categories are caused by or constituted by or metaphysically grounded in natural kinds—especially biological or medical kinds—that do not necessarily involve human mental states, decisions, culture, or social practices" (207). I contend that this realist account of human kinds, which maintains that human science classifications that refer to natural kinds yield projectable inferences, offers a superior explanation for the projectability of some human science classifications (e.g., 'schizophrenia,' 'depression').

3. Biological Kinds Yield Robust Projectable Inferences

In arguing that classifications of natural kinds yield robust and stable projectable inferences, I adopt a realist position about natural kinds. Mallon (2016) contrasts his social constructionist position for understanding human categories to an alternative realist view that maintains that some human classifications are constituted by natural kinds, especially biological kinds (207). Compared to the more qualified version of realism ('basic realism') endorsed by Mallon (2016, ch. 6), this more orthodox realist position—which maintains that constructed classifications that individuate natural kinds are projectable—provides a superior explanation for the robust and ampliative projectable inferences yielded by some human science classifications. On this view, human science classifications of natural kinds yield robust and stable projectable because they individuate classes underwritten by a stable set of natural (e.g., biological) properties shared by members of a class. In terms of Mallon's discussion of HPC kinds, this amounts to the claim that at least some of the mechanisms that maintain HPC kinds are intrinsic or natural (e.g., biological) properties of the kind in question.

While Mallon's social role account implies that purely social kinds (e.g., gender) constituted exclusively by relational properties can yield reliable projectable inferences, there are reasons for thinking that there is a significant difference between the projectable inferences produced by classifications of natural kinds versus social kinds. First, notice that we normally do not expect classifications of *purely* social kinds (e.g., 'teachers,' 'liberals,' 'widows') to yield any ampliative projectable inferences in the long term. These categories are recognized to be purely conventional classes, and we do not expect them to yield any scientifically interesting kinds of inferences. Most of the reliable inferences that can be drawn from these classifications are implied, directly or indirectly, by the definition of the classification (e.g., teachers will be

educated, widows are women), and hence, such inferences are non-ampliative. While social kinds can undoubtedly sometimes produce relatively stable ampliative predictions (e.g., teachers are less likely to be incarcerated), the social mechanisms that underwrite these predictions are subject to change as the social world evolves. By contrast, we expect human science classifications of natural kinds (e.g., 'PKU,' 'schizophrenia') to yield ampliative and stable projectable inferences. Classifications of natural kinds yield robust and stable projectable inferences *because these classes are assumed to be underwritten by stable natural (e.g., biological) mechanisms.*[6] Classifications of purely social kinds, by contrast, are assumed to be underwritten by much less stable social mechanisms (e.g., broadcasting of social roles, adoption of social roles, imitation of stereotypes). Due to this difference regarding the mechanisms that underwrite these classifications, classifications of purely social kinds yield *transitory* projectable inferences relative to classifications of natural kinds.

The account of natural kinds outlined above draws on a venerable philosophical tradition that regards members of a natural kind as sharing some natural (intrinsic) properties (Hacking 1991). On this traditional understanding, natural kinds possess the following features (Bird and Tobin, 2018):

1) Members of a natural kind should have some natural properties in common.

2) Natural kinds should permit inductive inferences.

3) Natural kinds should participate in laws of nature.

---

[6] This view articulated herein draws on an argument I have made elsewhere (Tsou 2013, 2016) that some psychiatric disorders (e.g., schizophrenia, depression) are aptly described as what Kendler, Zachar, and Craver (2011) call mechanistic property cluster (MPC) kinds: kinds constituted by biological mechanisms at multiple levels (e.g., molecular, developmental, neurobiological) that interact to produce the key features of the kind. The MPC kind account can explain the projectability of kind terms by appeal to the regular and non-accidental (law-like) intrinsic properties shared by members of kind. For a broader and discussion of the projectability of psychological and psychiatric categories in connection with natural kinds, see Fodor (1974), Kim (1992), Block (1997), Griffiths, (1997, 1999, 2001), Millikan (1999), Antony (2003), Godman (2013), Khalidi (2013), and Zachar (2008).

Essentialist accounts maintain that natural kinds yield projectable inferences (2) because members of a natural kind share some intrinsic (natural) properties in common (1) that participate in laws in nature (3). Unlike essentialist accounts of natural kinds, which identify the natural properties in (1) with essential properties, Boyd's HPC account maintains that HPC kinds yield projectable inferences (2) because members of HPC kinds share a family of properties (1) that cluster in a regular and non-accidental (law-like) manner because of homeostatic mechanisms (3). Mallon's strategy for explaining the projectability of human kinds is to maintain that entrenched social roles (e.g., gender) can produce HPC kinds, and entrenched social roles share a family of *social properties* (1), which cluster in a regular and non-accidental manner because of homeostatic (e.g., social and environmental) mechanisms (3). Can these relational (social) properties ground the projectability of HPC kinds in the same way that intrinsic (natural) properties do? Ian Hacking has offered an influential argument for answering this question negatively. Hacking (1991, 1995) argues that the kinds ('human kinds') that are studied in the social sciences are fundamentally different from the traditional conception of natural kinds insofar as they are constituted by social properties. Compared to the stable natural properties that constitute natural kinds, Hacking (1995, 1999, 2007) argues that the social properties that constitute human kinds are *inherently unstable* because of the looping effects of human kinds. Looping effects are social feedback effects distinctive to the social sciences, wherein the meaning of a human science classification (e.g., 'multiple personality') *changes* the experiences and behaviors of classified people in unpredictable ways, thereby requiring revisions to the classification. Because looping effects constantly change the social properties that are utilized to identify members of a human kind, Hacking argues that the objects of human science classifications are 'moving targets' and, hence, their classifications will not yield stable

11

projectable inferences (Hacking 1999: 108; Hacking 2007). While I have argued against the generality of Hacking's conclusions elsewhere,[7] I agree with his claim that socially determined properties are not sufficiently stable to ground robust projectable inferences. The idea that the projectability of kind terms stems from stable properties is also amenable to James Woodward's (2000) argument that generalizations that are explanatory in the special sciences are moderately *invariant* under various interventions. For Woodward, a generalization is invariant (or stable) insofar as it would continue hold as other variables or background conditions change. Within Woodward's framework, robust projectable inferences can be understood as generalizations that remain invariant under a wide domain of interventions and background (e.g., social or historical) conditions. Below, I argue that Mallon's social kinds do not yield projectable generalizations of this sort.

The framework outlined above suggests that there is a meaningful epistemological distinction between the projectable inferences yielded by classifications of social kinds compared to classifications of natural kinds. Classifications of social kinds yield transitory and unstable projectable inferences, while classifications of natural kinds yield robust and stable projectable inferences. This difference can be summarized by stating that the projectable inferences drawn from classifications of social kinds will be contingent, while the projectable inferences drawn from classifications of natural kinds will at least approach the philosophical ideal of necessity. While I present this distinction as a difference in kind, I regard it as a difference of degree, based on the assumption that social and cultural evolution moves at a much faster pace than biological evolution (see note 1). From this perspective, consider Mallon's paradigm examples of social

---

[7] Against Hacking, I argue that the targets of some human science classifications (e.g., schizophrenia, depression) are stable *despite of the social feedback associated with their classifications* (Tsou 2007, 2013, 2016). In accordance with the argument articulated in this paper, I maintain that the targets of human science classifications are stable (and yield projectable inferences) precisely when they are natural kinds constituted by intrinsic biological properties.

kinds: race and gender. What kinds of ampliative projectable inferences can be drawn from these categories? Mallon is especially interested in how these socially constructed categories can yield projectable inferences related to oppression. In that light, one could argue that there are reliable and ampliative inferences related to oppression that we can currently draw from these categories. For instance, African Americans are more likely to be pulled over by the police in the United States, or women are more likely to spend time on household activities than men. The analysis of this paper suggests that these projectable inferences, if they are in fact predictively valid, are contingent. They are contingent insofar as the social forces involved in stabilizing these generalizations (e.g., racism, patriarchy) are historically contingent and subject to change. Moreover, these generalizations will not be invariant (or will have a limited domain of invariance) in Woodward's sense insofar as they will not remain stable in different social contexts (e.g., in equitable societies). For philosophers interested in social change, it should be regarded as a good thing that these sorts of projectable inferences are neither necessary nor stable. The analysis articulated herein suggests that even deeply entrenched social mechanisms can be subject to change. By contrast, projectable inferences associated with classifications of natural kinds (e.g., 'PKU,' 'introversion,' 'schizophrenia') are necessary, or at least, approaching the philosophical ideal of necessity insofar as the biological mechanisms underwriting these categories remain stable over relatively long evolutionary time scales. The distinction between transitory and robust projectable inferences articulated above is not intended to be controversial. It is meant to point to a fundamental (perhaps trivial) distinction that is presupposed in some human and social sciences (e.g., medicine, psychology, psychiatry) that assume that the proper objects of scientific classification are natural kinds, and not social kinds.[8]

---

[8] In presenting this work, discussants have either regarded the distinction between transitory and robust projectability to be either uncontroversial and trivially correct, or highly problematic. For those who oppose this

My analysis suggests that while Mallon may have shown that classifications of social

kinds or social constructs can yield transitory projectable inferences, he has not demonstrated

that they can yield the robust and stable projectable inferences that *are demanded of predictive*

scientific classifications. In order to yield robust and stable projectable inferences, a human

classification needs to—at least partially—individuate a natural kind underwritten by stable

biological mechanisms. In terms of Mallon's discussion of HPC kinds, this amounts to the claim

that some of the mechanisms that maintain HPC kinds are natural or intrinsic (e.g., biological)

properties of the kind in question. This position speaks to an outstanding issue in the HPC kinds

literature regarding whether the causal mechanisms underwriting HPC kinds are intrinsic or

relational mechanisms. Philosophers of biology who have discussed HPC kinds in connection

with species have emphasized the importance of relational mechanisms (e.g., interbreeding with

conspecifics, phylogenetic relations, exposure to similar environments) that maintain the stability

of property clusters associated with a species (Boyd 1999a, Millikan 1999, Wilson 1999, Wilson,

Barker, and Brigandt 2007, Ereshefsky 2017). Some (e.g., Griffiths 1999, Okasha 2002, LaPorte

2004) suggest that the cluster of properties associated with species can be explained *exclusively*

in terms of relational properties (e.g., being a descendant of an ancestor, interbreeding with

conspecifics), while others (e.g., Boyd 1999a, Wilson 1999, Devitt 2008, 2010) insist that some

---

distinction, I would challenge them to provide an example of a *purely socially constituted human classification* that yields ampliative and robust projectable inferences. In conversation, Stephen Turner has suggested that classifications from the Caste system might be a counterexample. I do not regard this as a counterexample to my distinction, but an example of a causally significant social classification system that is deeply entrenched within Indian culture. Projectable inferences that can be drawn from Caste level distinctions (e.g., higher Caste members tend to be wealthier than lower Caste members) are ultimately contingent; but they have remained stable over time because there are formal and informal social rules and norms (e.g., job discrimination against lower Caste levels) that support the stability of these inferences. My position is that a relevant change in the social environment would render these inferences invalid. By contrast, robust projectable inferences that are underwritten by stable biological mechanisms will remain invariant or stable—in a Woodwardian sense (Woodward 2000)—*across* different and changing social environments.

intrinsic properties (e.g., genetic properties) are necessary.[9] The argument of this paper suggests that in order to yield robust projectable inferences, a classification needs to individuate some intrinsic properties. It is these intrinsic (biological) mechanisms that are shared by members of a kind, which explain the projectability of scientific classifications. In this regard, it is important to clarify two distinctive roles that causal mechanisms play in the theory of HPC kinds:

1) From a metaphysical standpoint, they fix the stability of natural classes and explain the unity ('homeostasis') of the properties that are used to identify kinds and its members.

2) From an epistemological standpoint, they ground the stability or invariance of projectable generalizations made about such kinds.

While relational mechanisms can address (1), they will be insufficient to address (2). This is because relational mechanisms, such as interbreeding with conspecifics or phylogenetic relationships, are *too general* to provide specific inferences about members of a particular species (cf. Devitt 2008: 352-353). For projectable inferences about members of a species to be drawn, a kind term needs to individuate some intrinsic properties of the kind in question (e.g., genetic properties that explain the physiological and morphological features of a particular species). I support this general argument in the following section, where I aim to show that robust and ampliative projectable inferences yielded by some psychiatric classifications (e.g.,

---

[9] In responding to an argument by Millikan (1999) that species should be regarded as 'historical kinds' maintained by relational mechanisms (e.g., a copying process), Boyd (1999b: 81) emphasizes that species are HPC kinds underwritten by *both* intrinsic mechanisms (e.g., genetic properties, phenotypic traits) and relational ('external') mechanisms (e,g., historical relations). Devitt (2008, 2010) notes that—like his own essentialist account—HPC accounts of species that require some intrinsic mechanisms (e.g., Boyd 1999a, Wilson 1999) imply that species have 'partly intrinsic essences.'

schizophrenia, bipolar disorder) are explained by the intrinsic (biological) mechanisms shared by members of a kind.

As a qualification, it should be stated there is a sense in which the analysis of this paper and Mallon's analysis are working at cross-purposes. My primary interest is to explain the (robust) projectability of human science categories, while Mallon's primary interest is to explain how socially constructed categories can become causally significant in the sense that there are real and systematic causal differences associated with being a member of a constructed category. Throughout his discussion, Mallon frames his project as demonstrating how socially constructed categories can be *causally significant* (e.g., see Mallon 2016: 4, 8-10, 68-69); it might be the case that all that Mallon needs to do to establish this point is to show that these categories can yield contingent and transitory projectable inferences. If this is the case, then I would grant Mallon this point, but emphasize that there is a significant epistemological difference between the projectable inferences yielded by classifications of social kinds (transitory) and natural kinds (robust), respectively.

4. Robust vs. Transitory Projectability: Some Examples

The assumption motivating the argument in the previous section is that classifications of natural kinds yield stable ampliative projectable inferences because these classifications (to some extent) accurately individuate classes associated with a stable biological causal structure, i.e., a set of biological mechanisms that interact to cause various properties to cluster in a regular and non-accidental manner. These projectable inferences are more robust and stable than projectable inferences drawn by classifications that individuate purely social kinds or the social constructs

16

discussed by Mallon. To support this argument, I present some examples of projectable human categories (e.g., PKU, schizophrenia, bipolar disorder) in the following section. These examples are intended to show that the robust ampliative projectable inferences that we can draw from these kind terms is due precisely to the fact that they individuate a natural kind underwritten by intrinsic biological mechanisms. Hence, the realist account offers a straightforward explanation of the projectability of some human science classifications: robust projectable generalizations are *causal consequences* of the common intrinsic (biological) properties shared by members of a kind. To support the idea that purely socially constituted kinds, like the social roles discussed by Mallon, cannot yield robust projectable inferences, but only transitory inferences, I discuss the example of the hysteria classification, which meets the general conditions specified by Mallon's social role account, but does not yield any stable projectable inferences. These examples support the general conclusion that human science classifications that yield robust projectable inferences are classifications that individuate natural kinds.

Consider the example of phenylketonuria (PKU), which is a rare recessive-gene disorder. The characteristic signs of PKU include a cluster of symptoms including: small head, prominent cheek and upper jaw bones with widely spaced teeth, poor development of tooth enamel, decreased body growth, and a mousy odor of the urine (Kring et al. 2007: 481-482). The symptoms of PKU are caused by an inherited lack of a liver enzyme, phenylalanine hydroxylase, which is needed to convert phenylalanine (an amino acid of food protein) to tyrosine (an amino acid that synthesizes proteins). When this enzyme is deficient, excessive phenylalanine builds up in the blood, which interferes with the myelination of neurons in the brain. When an infant with PKU eats foods containing phenylalanine, the brain fails to develop normally, resulting in severe mental retardation. Unlike other congenital disorders (e.g., Down syndrome), PKU is treatable.

Treatment of infants with PKU consists of placing them on a strict low-phenylalanine diet, which keeps phenylalanine levels in the blood low, allowing myelination to take place normally. Once myelination is complete, dietary restrictions can be removed since phenylalanine no longer threatens brain development. I maintain that the PKU classification refers to a biological kind insofar as the cluster of properties associated with PKU are constituted and maintained a set of stable (intrinsic) biological mechanisms and processes (e.g., genetic mechanisms, developmental mechanisms). In this example, notice that the PKU classification yields ampliative and scientifically useful projectable inferences (e.g., placing PKU infants on a low-phenylalanine diet will allow brain development to occur normally) precisely because we have knowledge about the biological mechanisms underwriting this category. Hence, the PKU classification yields stable projectable inferences because the classification accurately individuates a natural kind.

One might object to my analysis by contending that PKU simply is a medical kind and that Mallon's social role account is intended to capture more socially-robust 'kinds of people'—to borrow Ian Hacking's (2007) phrase—that are constituted by folk-psychological social roles that are common knowledge in societies. To offer some examples that more obviously involve social roles of this sort, I discuss schizophrenia and bipolar disorder below, on the assumption that the 'schizophrenic' and the 'manic-depressive' are putative examples of entrenched social roles or kinds of people.

Schizophrenia is currently defined by a cluster of symptoms including: (1) delusions (e.g., persecutory, grandiose), (2) hallucinations (typically auditory), (3) disorganized thinking and speech, (4) disorganized or catatonic behavior, and (5) diminished emotional expression or low levels of motivation (American Psychiatric Association 2013: 87-88, 99). Symptoms (1) and (2) are referred to as the 'positive symptoms' of schizophrenia, while (3) and (4) are referred to

18

as 'disorganized symptoms,' and (5) are referred to as 'negative symptoms.' While research on the causes of schizophrenia is complex and controversial (e.g., see Kendler and Schaffner 2011), there is compelling evidence that the symptoms of schizophrenia are closely related to abnormal neurobiological processes (see Tsou 2012, 2017). According to the 'dopamine theory of schizophrenia,' excessive dopamine activity in the mesolimbic pathway—which projects from the midbrain in the ventral tegmentum to the nucleus accumbens—causes the positive symptoms of schizophrenia; while deficient dopamine activity in the mesocortical pathway— which projects from the ventral tegmental area to the prefrontal cortex—causes the negative symptoms associated with schizophrenia (Kring et al. 2007: 363-365). While this neurobiological theory is oversimplified and glosses over complexities (e.g., the role of glutamate), there is robust evidence that the cluster of symptoms associated with the schizophrenia classification are constituted by a set of stable biological mechanisms. For the purposes of this paper, it is important to notice that robust ampliative projectable inferences (e.g., inferences regarding treatment) can be drawn from the schizophrenia classification precisely because the cluster of properties associated with the schizophrenia classification are underwritten by a set of intrinsic neurobiological mechanisms. For example, we can reliably predict that individuals exhibiting the positive (or psychotic) symptoms of schizophrenia will benefit from dopamine antagonist drugs that decrease dopamine activity in the mesolimbic pathway on the assumption that the schizophrenia classification accurately picks out a biological kind.

Bipolar disorder—of which 'bipolar I disorder' represents the modern classification of the manic-depressive classification advanced in the nineteenth century—is defined by a cluster of symptoms including the presence of at least one manic episode (defined by inflated self-esteem, decreased need for sleep, flight of ideas, distractibility, increase in goal-directed activity,

and excessive involvement in risky behaviors) and typically involves the cycling between manic and depressive episodes (American Psychiatric Association 2013: 123-129). Bipolar disorder is a particularly interesting case since research on the biological causes of bipolar disorder are equivocal, despite the fact that there is a known effective pharmacological treatment, viz., lithium treatment. While the biological causes of bipolar disorder are unclear, research suggests that bipolar disorder is best understood as a family of similar but subtly different conditions, whose symptoms are maintained by the dysfunction of interconnected brain networks. Biological processes implicated in bipolar disorder include loss of volume in brain areas known to be involved in mood regulation (i.e., the amygdala, ventrolateral prefrontal cortex, and ventral anterior cingulate cortex) and cognitive control (i.e., the dorsal anterior cingulate cortex, dorsomedial prefrontal cortex, dorsolateral prefrontal cortex), dysregulation of glial-neuronal interactions at the cellular level (e.g., overactive microglia), and increased inflammation in the periphery of the body (Maletic and Raison 2014). The details of the biological mechanisms involved in bipolar disorder are less important than the assumption that there are underlying biological mechanisms that interact to regularly produce the cluster of symptoms associated with the bipolar classification. If this assumption is correct, then the category of bipolar disorder yields stable projectable inferences (e.g., bipolar patients can be treated with lithium) because of the intrinsic biological mechanisms that maintain its cluster of symptoms.

The examples of PKU, schizophrenia, and bipolar disorder are intended to illustrate that the projectability of human science classifications can quite naturally be explained by the realist account that some human classifications (to some degree) accurately individuate biological kinds. In these examples, the classifications yield projectable inferences *because the cluster of properties individuated by these classifications are caused by a set of stable natural (i.e.,*

biological) mechanisms shared by members of a kind. While Mallon's preferred social role account of human kinds offers a plausible 'how possibly' model of how human kinds—conceived of as social kinds—are projectable, the realist (or biological) account of human kinds provides a more compelling and parsimonious account of the projectability of human categories. Mallon's social role account might explain how human kind classifications like schizophrenia or bipolar disorder yield non-ampliative projectable inferences (e.g., schizophrenics will report hearing voices, manic-depressives will experience rapid shifts of mood); however, it fails to offer a clear explanation of how these human categories can yield robust ampliative projectable inferences (e.g., the psychotic episodes of schizophrenics can be treated with dopamine antagonist drugs). On the realist account, these inferences are possible precisely because the category accurately picks out intrinsic (biological) properties that ground the category.

To see why Mallon's social role account is insufficient to explain the projectability of socially constructed human kinds, consider the classification of hysteria as it emerged in the late nineteenth century. Hysteria offers an example that fits nicely with the general account of social construction and social roles articulated by Mallon, yet it is a classification that did not yield any robust projectable inferences in the long term. Since the classical antiquity period, hysteria was characterized by symptoms such as convulsions, spasmodic seizures, and feelings of strangulation; by the nineteenth century, other symptoms were added, including fainting and swooning, paralysis of the limbs, anaesthesias, and trancelike states (Goldstein 1989: 323-324). Around the 1880s in France, there was a dramatic increase in recorded cases of hysteria, which gradually declined during the twentieth century to its virtual disappearance by the 1980s (Micale 1993). During its heyday in late nineteenth-century France, hysteria satisfied the general conditions specified by Mallon's account of human kinds. There were many representations of

21

hysteria (e.g., theories, models, narrative, pictures), which reinforced a common stereotype about hysteria (e.g., it is a neurological condition, typically afflicting females, caused by past trauma). Moreover, the public broadcasting about this social role—e.g., Charcot's lectures at the Salpêtrière that featured the treatment of hysterical patients with hypnosis and popular depictions of hysteria in paintings and photographs—rendered the hysteria role common knowledge. There are multiple social mechanisms that maintained and entrenched this social role. For example, looping effects inevitably played a role in structuring the expectations and behavior of both those labelled as hysteric and the users of that label. Moreover, the kinds of implicit incentives (e.g., patients being reinforced for complying with psychiatrists' expectations) discussed by labelling theorists (e.g., Scheff 1966) also functioned to entrench this social role. There were also environmental mechanisms (e.g., the institutional recognition of hysteria by doctors, the authoritative role of medicine and neurology, the presence of hospitals for mentally-disturbed women) that entrenched the hysteria role. Despite its entrenchment, the hysteria classification failed to yield long-term robust projectable inferences, which led to the demise of this category when it eventually split into several separate categories (e.g., 'conversion disorder,' 'somatization disorder,' 'histrionic personality disorder'). On the analysis suggested herein, the reason why the hysteria classification failed to yield robust projectable inferences is because the category failed to accurately individuate a biological kind.

What is missing from Mallon's account is an explicit position regarding the extent that human kinds must be grounded in natural (e.g., biological) properties. Mallon is officially neutral on this issue. In contrasting his account to Ásta's 'conferralist' account of social construction (Sveinsdóttir 2013)—which suggests that category membership is solely determined by being classified or conferred as a member of that category in a certain context—Mallon (2016) states

that his own account maintains that "membership in a category … consists in possession of some sufficient cluster of properties that *may include putatively natural properties* … , being classified by practices … , or the consequences of such ascriptive practices" (212, emphasis added). Mallon (2016) elaborates on this neutral position:

> [M]y account is neutral with regard to the constituting or metaphysical grounding properties for a category … I hold that the relevant homeostatic property-cluster … which constitutes membership in a particular human category is to be determined a posteriori by investigation into the features that explain whatever phenomena ground [the] use of the term. In this vein, as I consider the construction of social categories like race and gender, it seems important to me not to exclude in advance the possibility of *partial construction*, that is, *the possibility that some actually natural properties figure alongside social roles and their consequences in grounding some categories*. (212-213, emphasis added)

Mallon's (2016) explicit consideration of 'partial construction' (cf. Gannett 2010) and inclusion of natural (or intrinsic) properties as possible grounding properties for human categories conflicts somewhat with the way that he articulates his account in other parts of his book (e.g., see chaps. 3, 8). Regardless, if Mallon is to offer a satisfactory account of the projectability of constructed human categories, he cannot remain neutral on this issue. As I have argued, the robust projectability of human kinds is entirely dependent on such categories accurately individuating natural (intrinsic) properties shared by members of a kind. Without appeal to such natural properties, Mallon's account cannot persuasively explain how human kinds yield *stable and ampliative* projectable inferences. At the very least, the realist account of human kinds offers a superior explanation compared to the social role account.

5. Conclusion

In this paper, I argued that Mallon's explanation of the projectability of socially constructed human categories needs to be qualified insofar as it fails to distinguish between transitory and robust projectable inferences. While the social kinds articulated in Mallon's account may yield transitory projectable inferences, they will not yield the robust and stable ampliative inferences demanded of social science classifications that aim to be predictive. The more orthodox realist view that human classifications yield projectable inferences when they individuate natural or biological kinds offers a more persuasive explanation for the robust projectability of some classifications. One way of expressing this argument is that Mallon's reconciliation of social constructionism with naturalism is not realist enough insofar as it fails to sufficiently acknowledge a (realist) distinction between the projectable inferences yielded by classifications (e.g., 'hysteria') that are exclusively constructed by human representations and those (e.g., 'bipolar disorder') that correspond to a set of biological properties shared by members of that class.

Although I presented Mallon's account in terms of social roles producing social kinds and the realist account that I favor in terms of categories individuating biological kinds, this simplified distinction was articulated for purposes of clarity and convenience. Most of the interesting projectable categories and kinds posited in the human and social sciences are classes inevitably constituted by both biological and social mechanisms. To offer a more compelling explanation of the projectability of human categories, Mallon's social role analysis needs to take into account more directly the (intrinsic) biological properties that can both ground various

human kinds and explain the similarity relations among members of that kind. The partial construction account of social construction entertained by Mallon offers a promising path towards a more comprehensive account of the projectability of human categories. The psychiatric categories discussed in this paper can be instructive in this regard (cf. Kleinman 1988: chs. 2-3; Tsou 2007: 339). The analysis of this paper suggests that the biological mechanisms individuated by a classification (e.g., 'schizophrenia') are responsible for producing a core cluster of characteristic signs (e.g., delusions, hallucinations). If a human category accurately individuates a biological kind, we can expect this core set of properties—and projectable inferences yielded by the general category—to remain stable over time. Social mechanisms (e.g., broadcasting of social roles, looping effects, role adoption) that function to fix criteria for a classification (e.g., 'hysteria') can be understood as social forces that can potentially stabilize a more local or culturally-sanctioned expression of these core signs (e.g., depression is expressed more in terms of somatic symptoms in China compared to Western countries). While these social mechanisms can undoubtedly contribute to establishing local and transitory projectable inferences, it is the accurate identification of biological mechanisms and kinds that allows for robust projectable inferences. Among investigations into the mechanisms that contribute to the property clusters associated with human categories, those that distinguish the contributions of biological mechanisms from social mechanisms—and explore how these mechanisms interact—merit further attention.

Acknowledgments

References

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

Antony, Louise. 2003. "Who's Afraid of Disjunctive Properties?" *Noûs* 37 (Supplement: *Philosophical Issues* 13): 1–21.

Block, Ned. 1997. "Anti-Reductionism Slaps Back." *Noûs* 31 (Supplement: *Philosophical Perspectives* 11): 107–132.

Bird, Alexander, and Emma Tobin. 2018. "Natural Kinds." *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2018/entries/natural-kinds/

Boyd, Richard. 1985. "Observations, Explanatory Power, and Simplicity: Toward a Non – Humean Account." In *Observation, Experiment, and Hypothesis in Modern Physical Science*, edited by Peter Achinstein and Owen Hannaway, 47–94. Cambridge, MA: MIT Press.

Boyd, Richard. 1991. "Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds." *Philosophical Studies* 61 (1–2): 127–148.

Boyd, Richard. 1999a. "Homeostasis, Species, and Higher Taxa." In *Species: New Interdisciplinary Essays*, edited by Robert A. Wilson, 141–186. Cambridge, MA: MIT Press.

Boyd, Richard. 1999b. "Kinds, Complexity, and Multiple Realization: Comments on Millikan's 'Historical Kinds and the Special Sciences.'" *Philosophical Studies* 95 (1–2): 67–98.

Boyd, Richard. 2010. "Realism, Natural Kinds, and Philosophical Methods." In *The Semantics and Metaphysics of Natural Kinds*, edited by Helen Beebee and Nigel Sabbarton-Leary, 212–234. New York: Routledge.

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

Devitt, Michael. 2008. "Resurrecting Biological Essentialism." *Philosophy of Science* 75 (3): 344–382.

Devitt, Michael. 2010. "Species Have (Partly) Intrinsic Essences." *Philosophy of Science* 77 (5): 648–661.

Dupré, John. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge MA: Harvard University Press.

Dupré, John. 2001. "In Defence of Classification." *Studies in History and Philosophy of Biology and the Biomedical Sciences* 32 (2): 203–219.

Ereshefsky, Marc. 2017. "Species." *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2017/entries/species/

Fodor, Jerry A. 1974. "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." *Synthese* 28 (2): 97–115.

Gannett, Lisa. 2010. "Questions Asked and Unasked: How by Worrying Less about the '*Really* Real' Philosophers of Science Might Better Contribute to Debates about Genetics and Race." *Synthese* 177 (3): 363–385.

Goldstein, Jan E. (1989). *To Console and Classify: The French Psychiatric Profession in the Nineteenth Century*. Chicago: University of Chicago Press.

Godman, Marion. 2013. "Psychiatric Disorders Qua Natural Kinds: The Case of the 'Apathetic Children'." *Biological Theory* 7 (2): 144–152.

Griffiths, Paul E. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Griffiths, Paul E. 1999. "Squaring the Circle: Natural Kinds with Historical Essences." In *Species: New Interdisciplinary Essays*, edited by Robert A. Wilson, 208–228. Cambridge, MA: MIT Press.

Griffiths, Paul E. 2001. "Emotion and the Problem of Psychological Categories", In *Emotions, Qualia, and* Consciousness, edited by Alfred W. Kaszniak, 28–41. Singapore: World Scientific.

Hacking, Ian. 1991. "A Tradition of Natural Kinds." *Philosophical Studies* 61 (1–2): 109–126.

Hacking, Ian. 1995. "The Looping Effects of Human Kinds." In *Causal Cognition: A Multidisciplinary Debate*, edited by Dan Sperber, David Premack, and Ann J. Premack, 351–383. Oxford: Clarendon Press.

Hacking, Ian. 1999. *The Social Construction of What?* Harvard, MA: Harvard University Press.

Hacking, Ian. 2007. "Kinds of People: Moving Targets." *Proceedings of the British Academy* 151: 285–318.

Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.

Kendler, Kenneth S., and Kenneth F. Schaffner. 2011. "The Dopamine Hypothesis of Schizophrenia: An Historical and Philosophical Analysis. *Philosophy, Psychiatry, & Psychology* 18 (1): 41–63.

Kendler Kenneth S., Peter Zachar, and Carl Craver. 2011. "What Kinds of Things Are Psychiatric Disorders?" *Psychological Medicine* 41 (6): 1143–1150.

Khalidi, Muhammad Ali. 2013. *Natural Categories and Human Kinds: Classification in the*

Natural and Social Sciences. Cambridge: Cambridge University Press.

Kim, Jaegwon. 1992. "Multiple Realization and the Metaphysics of Reduction." *Philosophy and Phenomenological Research* 52 (1): 1–26.

Kleinman, Arthur. 1988. *Rethinking Psychiatry: From Cultural Category to Personal Experience*. New York: The Free Press.

Kring, Ann M., Gerald C. Davison, John M. Neal, and Sheri L. Johnson. 2007. *Abnormal Psychology*, 10th ed. Hoboken, NJ: John Wiley & Sons.

LaPorte, Joseph. 2004. *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press.

Maletic, Vladimir, and Charles Raison. 2014. "Integrated Neurobiology of Bipolar Disorder." *Frontiers in Psychiatry* 5: 98. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4142322/

Mallon, Ron. 2016. *The Construction of Human Kinds*. Oxford: Oxford University Press.

Micale, Mark S. 1993. "On the 'Disappearance' of Hysteria: A Study in the Clinical Deconstruction of a Disorder." *Isis* 84 (3): 496–526.

Millikan, Ruth G. 1999. "Historical Kinds and the 'Special Sciences.'" *Philosophical Studies* 95 (1–2): 45–65.

Okasha, Samir. 2002. "Darwinian Metaphysics: Species and the Question of Essentialism." *Synthese* 131 (2): 191–213.

Perreault, Charles. 2012. "The Pace of Cultural Evolution." *PLoS One* 7(9): e45150. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3443207/

Richerson, Peter J., and Robert Boyd. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Richerson, Peter J., Robert Boyd, and Joseph Henrich. 2010. "Gene-Culture Coevolution in the Age of Genomics." *Proceedings of the National Academy of Sciences of the United States of America 107* (Suppl 2): 8985–8992.

Scheff, Thomas J. 1966. "The Role of the Mentally Ill and the Dynamics of Mental Disorder: A Research Framework." *Sociometry* 26 (4): 436–453.

Slater, Matthew H. 2015. "Natural Kindness." *British Journal for the Philosophy of Science* 66 (2): 375-422.

Sveinsdóttir, Ásta K. 2013. "The Social Construction of Human Kinds." *Hypatia* 28 (4): 716–732.

Tsou, Jonathan Y. 2007. "Hacking on the Looping Effects of Psychiatric Classifications: What Is an Interactive and Indifferent Kind?" *International Studies in the Philosophy of Science* 21 (3): 329–344.

Tsou, Jonathan Y. 2012. "Intervention, Causal Reasoning, and the Neurobiology of Mental Disorders: Pharmacological Drugs as Experimental Instruments." *Studies in History and Philosophy of the Biological and Biomedical Science* 43 (2): 542–551.

Tsou, Jonathan Y. 2013. "Depression and Suicide are Natural Kinds: Implications for Physician-Assisted Suicide." *International Journal of Law and Psychiatry* 36 (5–6): 461–470.

Tsou, Jonathan Y. 2016. "Natural Kinds, Psychiatric Classification, and the History of the *DSM*." *History of Psychiatry* 27 (4): 406–424.

Tsou, Jonathan Y. 2017. "Pharmacological Interventions and the Neurobiology of Mental Disorders." In *The Physics of the Mind and Brain Disorders: Integrated Neural Circuits Supporting the Emergence of the Mind*, edited by Ioan Opris and Manuel F. Casanova, 327–367. Cham: Springer.

Wilson, Robert A. 1999. "Realism, Essence, and Kind: Resuscitating Species Essentialism?" In

Species: New Interdisciplinary Essays, edited by Robert A. Wilson, 187–207. Cambridge, MA: MIT Press.

Wilson, Robert A., Matthew J. Barker, and Ingo Brigandt. 2007. "When Traditional Essentialism Fails: Biological Natural Kinds." *Philosophical Topics* 35 (1–2): 189–215.

Zachar, Peter. 2008. "Real Kinds but No True Taxonomy: An Essay in Psychiatric Systematics." In *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology*, edited by Kenneth S. Kendler and Joseph Parnas, 327–367. Baltimore, MD: John Hopkins Press.