Looping kinds and social mechanisms

Jaakko Kuorikoski & Samuli Pöyhönen
University of Helsinki

Penultimate draft, 27.2.2012
Forthcoming in *Sociological Theory*

Abstract

Human behavior is not always independent of the ways in which it is scientifically classified. The existence of the looping effects of human kinds has been used as an argument for the methodological separation of the natural and the human sciences, and for motivating social-constructionist claims about the targets of the latter. We suggest that these arguments rely on false presuppositions about the nature and rationale of scientific classification in general. The discussion concerning the social construction of kinds suffers from a lack of analytic explanatory middle-range framework between atheoretical case studies of specific looping phenomena and grand theorizing about the metaphysical nature of social reality. We suggest that such a framework can be grounded on the concept of *social mechanism*.

1 Introduction

Categorizing people has generally been considered a prerequisite for describing, explaining and ultimately controlling society. However, classifications of people often have unforeseen consequences for the behavior of their targets, and the possibility that classifications in the human sciences could ever be truly-mind independent in the same sense as in the natural sciences has repeatedly been called into question. Categorizations of human beings are special in that the classified objects may become aware of the classification and in consequence change their behavior. Quarks do not mind being labeled as quarks. Similarly, the challenges facing a group of beetles may change drastically if they come to be categorized as pests, but the beetles themselves do not recognize the classification as such. People, on the other hand, do mind if they are labeled as criminals, immigrants, religious, or anorectic. Categorizing people typi-

cally provokes reactions in them, and sometimes the classifications must later be revised to reflect the changed behavior.

The existence of such *looping effects* has been used as an argument for the deep methodological separation of the human and the natural sciences (e.g., Hacking 1999, 108). The interactive nature of human kinds is also a major motivating factor behind sweeping social-constructionist claims: it is a fundamental sociological insight that human kinds are not similarly mind-independent and objective features of the order of the world as categories in the natural sciences are, but instead they are products of our ways of describing and acting with each other (e.g., Burr 1995; Gergen 1996). Yet, such claims about social construction are often highly ambiguous and have connotations that are more likely to lead to conceptual and methodological pseudo-problems than to increased understanding of the social phenomenon under study.

Views about the nature and rationale of scientific classifications and the mode of existence of the classes of things that these classifications capture are linked to wider issues in the epistemology and metaphysics of the social sciences. The existence of "real" kinds in the domain of inquiry is often thought to be a necessary condition for causal or law-like explanations of phenomena as well as a prerequisite for reliable inductive inference. Moreover, it is a common everyday intuition that placing an individual into an appropriate category is by itself sufficient for explaining those properties of the individual that are typical of the category in question. The idea that scientific categories should capture the true mind-independent divisions in reality is thus commonly associated with some form of *essentialism*, according to which individuals possess some of their properties necessarily, in virtue of their nature. No wonder then, that the idea that sociological categories should trace the true divisions in the social realm has often been met with accusations of naïve essentialism. It is simply irrespon-

sible to hold that categories defining ethnicity, class-membership, party-affiliation, or mental illness would correspond to essences that doom their bearers to whatever properties these essences entail. Instead, properties of people are transformed into public classifications and knowledge only by means of discursive articulation, and therefore social categories – both those used by the people studied as well as those deployed by social scientists – are products of norms, power relations, and negotiation. Quite often this sensible resistance to essentialism can take the form of constructionism or pure conventionalism, which imply strong nominalism and, for some, the abandonment of scientific realism. This, in turn, has driven those who find such constructionist stances somehow anti-scientific to propound weaker forms of essentialist realism. Critical realism is a prominent example of this more moderate essentialist approach (e.g., Sayer 1997).

In this paper, we argue that both constructionism and essentialism are too crude views of the nature and rationale of scientific categorization. The shortcomings of these views are exposed in their inability to deal with the very feature of social scientific phenomena motivating the meta-theoretical discussions on constructionism and realism: the *looping effect* of human kinds. Ever since the classics, sociological theorizing has been sensitive to the reflexive, "double-hermeneutic" nature of social scientific knowledge in general (Bourdieu 1992; Giddens 1984), and the effects of labeling on those categorized have been extensively studied in the fields of criminology and deviance, political sociology, mental illness, and sociology of science (cf. Goffman 1963; Lemert 1967; Laub & Sampson 2006; Valentine 2010). The different means of creating, maintaining, transforming and utilizing mutually recognized categorical boundaries between social groups has constituted a fertile area of research. The causal interaction between symbolic boundaries and objectified social boundaries between cate-

gories such as social classes, ethnicities and genders has been argued to be a major factor in sustaining various stubborn inequalities in our society. (Lamont & Molnár 2002; Tilly 1998.)

However, in this paper we approach categories from a slightly different angle. We focus on specific kinds of human classifications, ones that originate in scientific research, and especially on the consequences that the interactivity of human scientific categories has for their *epistemic* functioning. The ideal form of scientific knowledge has traditionally been conceptualized using the concept of *natural kind*, and interactivity seems to call into question the naturalness of human kinds. The nature and implications of looping effects therefore require more systematic study. Our focus on the interactive nature of human kinds suggests a further difference between our approach and earlier research on social categories: Interactivity characteristic of the looping effect is constituted by a two-way causal connection between scientific knowledge and the described behavior. Studies of labeling effects and boundary construction have typically focused only on one direction of influence, the changes in people's behavior caused by the classification, whereas we suggest that the study of looping effects proper requires also examining the properties and complex dynamics of interactive knowledge.

The conception of a specific phenomenon of the looping effect of human kinds only really came to the fore in the 1980s in Ian Hacking's work. The problem is that Hacking's case studies (although groundbreaking and fascinating) only point towards an interesting phenomenon without providing the resources for systematic modeling, analysis and, ultimately, explanation. Much of the ensuing discussion concerning the social construction of human kinds has similarly shied away from any concerted attempt to create a systematic and explanatory framework for looping kinds. This has

enforced the mistaken impression that looping is an interesting but relatively isolated phenomenon, not directly related to the larger issues concerning the possibility of scientific realism in the social sciences.

This article is an attempt to remedy two interrelated shortcomings. First, we believe that the discussion concerning the social construction of kinds and categories suffers from the general malady identified by Robert Merton (1967, Ch. 2): there is no analytic explanatory middle-range framework between atheoretical case studies of specific looping phenomena and grand theorizing about the metaphysical nature of social reality. We suggest that such a framework can be grounded on the concept of *social mechanism*. We align our use of the notion of mechanism with the methodological approach of analytical sociology in that mechanism descriptions have to be more informative than mere labels for observed effects or correlations, or causally non-committal narratives of social processes. A truly explanatory mechanism attribution must enable inferences to salient possibilities over and above what has actually happened (Hedström 2005; Hedström & Bearman 2009; Hedström & Ylikoski 2010).

Second, essentialist and social-constructionist arguments for and against the possibility of realism about human scientific categories typically presuppose wrong-headed views about the nature of scientific classification and about the role that kind-concepts play in explanation and inductive inference. Our aim is to use a recently proposed philosophical view about the nature and function of scientific classifications, the homeostatic property cluster (HPC) theory, as a framework within which to situate looping effects. This will allow us to characterize looping effects within a unifying theory of scientific classification and understand them as a special case of the

causal interaction between social knowledge and social reality.[1] These theoretical tools allow us to see that the interactivity of human scientific classifications does not raise a special metaphysical problem or conceptual curiosity that is characteristic of the human sciences. Nevertheless, a more fine-grained,and explicitly causal-mechanistic understanding of scientific classification is necessary when dealing with social phenomena (e.g., mental disorders) that are produced and sustained by both biological causes and social feedback mechanisms. Without an explicit acknowledgement of the importance of investigating the causal mechanisms behind these kinds of problematic hybrid phenomena, the discussion tends to degenerate into empty disputes about the priority of one kind of mechanism over another – disputes between physiological/genetic determinism and all-encompassing social constructionism.

Our discussion in the paper proceeds in three stages. The first stage presents the general philosophical model of scientific classification. We briefly discuss the historical changes in the conception of proper scientific categorization and introduce the notion of *natural kind*[2]*,* proposed to explicate the ontological basis of such categorizations. Building on these conceptual foundations, we introduce the homeostatic property cluster theory of kinds, upon which we build our mechanistic theory of classification. Examining this philosophical background is crucial for understanding debates about scientific classification, because natural kindhood of the classification in question is often the implicit target of social-constructionist criticism, and thus the concept of

[1] This interactivity between social-scientific knowledge and its target is now commonly called performativity. Performativity, like social construction, has become more of a vague slogan than a
[2] 'Kind' here refers to a regularity in the association of properties in the world, whereas by 'classification' and 'kind-concept' we mean representations (either mental or linguistic) referring to kinds.

natural kind needs to be made explicit in order to clarify the debate between realist and constructionist positions.

In the second stage of our argument, we show how our mechanistic framework for understanding classifications can be used to make sense of interactive human kinds and their looping effects, which, after all, are supposed to be especially problematic cases for a realistic understanding of classification in the human sciences. We use the mechanistic theory of natural kinds to provide a preliminary account of different types of feedback mechanisms that are responsible for looping effects. This comprises sections three and four.

We connect the threads in section five, applying the account developed in the paper to the case of autism spectrum disorders. We argue that because these psychiatric illnesses are sustained by a complex set of biological, cognitive and social factors, they manifest several different kinds of feedback mechanisms. Our aim is to show that compared to the alternative views (essentialism, constructionism, and critical realism), our mechanistic theory of looping presents a more analytically fruitful approach to such phenomena. By focusing on the social mechanisms sustaining the classified behavior, we can move beyond a mere semantic characterization of the relationship between classifications and phenomena, towards an explanatory and systematic causal account of interactive human kinds.

2 Scientific categorization and natural kinds

To begin with, we need a better understanding of the epistemic role of classifications in science. We also need to make explicit many of the implicit intuitions that continue to drive the arguments about the nature and role of classifications within the human

sciences. It is often assumed that explanation and inductive inference, perhaps even the very possibility of scientific realism, require that the concepts involved pick out true natural kinds. These assumptions render problematic the class of cases in which the investigated social phenomenon does not satisfy all the intuitions about the naturalness, objectivity, or reality of kinds. Recognizing and correcting these intuitions is thus necessary for bringing clarity to the question of whether interactive human classifications can correspond to "real" scientific kinds.

Although the relationship between the structure of the world and our conceptual schemes is an age-old philosophical topic, modern systematic inquiry into the foundations of scientific classification dates back to the 19th-century debates around biological classification. The formulation of systematic taxonomies of biological organisms was in its heyday; exotic specimens from the new colonies flooded the botanical gardens and zoos of Europe, creating a demand for a theoretical basis for their classification (Hacking 2006, 6). At the same time, the long-held Aristotelian view of the metaphysics of classification was opened for re-examination. According to the Aristotelian view, targets of classification had *essences* according to which they could be grouped into a hierarchical, nested system of species and genera. The rise of more austere forms of empiricism, the mechanization of the scientific world-view (cf. Grene & Depew 2004, Ch. 3), and especially the birth of the theory of natural selection with its emphasis on population-level thinking, dismantled the philosophical basis for the intuitively appealing Aristotelian essentialism: The fact that entities in the natural world seem to fall into well-delineated groups in which all the members share the essential properties of the taxon in question, needed no longer be explained through the metaphysically primitive concept of a species-specific essence. On the contrary, the biological evolutionary mechanisms of inheritance and selection explain

the general phenomenon that populations of inter-breeding organisms exhibit only a limited amount of phenotypic variance.

Discarding Aristotelian essentialism for a nominalist approach to the basis of classification created the problem of how to explain the apparent fact that some classes seem more significant and natural than others. Some classifications appear to pick out pre-existing divisions in reality, whereas others seem to be based on mere convention and linguistic agreement. Classifications of elementary particles appear more fundamental than classification of people according to their political views. The function of the philosophical concept of *natural kind* has been to make the distinction between classifications that merit scientific inquiry and merely conventional classifications, without relying on a metaphysically primitive notion of essence. The central motivation for the tradition of natural kinds (cf. Hacking 1991) has been to make sense of the realist intuition that our scientific classifications and the mind-independent structure of reality should somehow match; classifications ought to carve nature by its joints.[3] This has been considered a prerequisite for the possibility of reliable inductive inference and scientific explanation, for example.

Once essentialism was abandoned, it became common to anchor natural kinds in the laws of nature. That is, a genuine phenomenon worthy of scientific investigation must fall under a law of nature. Laws and natural kinds have largely been seen as two sides of the same coin: laws are regularities linking natural kinds and kinds are natural only if they take part in laws (Fodor 1974; Goodman 1955; Rosenberg 2005). However, the law-based approach to scientific classification has two major problems: Firstly, as much of the 20th-century philosophy of science demonstrates, the concept of law of nature has proven to be at least as difficult to analyze as that of natural kind. Second-

[3] Cf. Plato: Phaedrus 265a–266a.

ly, linking explanation and inductive inference to the concept of law of nature leaves the human sciences (as well as most of the other special sciences) with a serious problem: if generalizations made in the human sciences do not meet the strict criteria for being laws of nature, should the explanations and inductions made within these disciplines be considered somehow defective? Clearly, the answer must be no, and thus the fault must lie with the theory of classification making such demands. In fact, law-based accounts of explanation and causation *are* highly problematic as tools with which to analyze explanation in the human sciences (Hedström and Ylikoski 2010; Woodward 2003, Ch. 5). In general, it could be argued that the classical conception of natural kind is linked to highly problematic views on scientific explanation, induction and the nature of laws and causation.[4]

Recently, the classical conception has been challenged by what we call the *mechanistic theory of natural kinds* (Boyd 1999; Griffiths 1997; Murphy 2006; Machery 2009); As Richard Samuels and Michael Ferreira (2010) observe, there is an emerging consensus among philosophers of science according to which natural kinds should be conceived as clusters of properties sustained by underlying causal mechanisms. Despite its popularity, some misgivings about the mechanistic theory remain. The theory appears to imply moderate classificatory pluralism, or even, according to the critics, conventionalism (cf. Craver 2009; Reydon 2009). However, as an anti-essentialist but causally realist view of natural kinds, it is currently the most viable contender for a foundational theory of scientific classification.

[4] In fact, we suspect that many of these problematic views result from an empirically well-established and pervasive cognitive bias, *psychological essentialism* (cf. Gelman 2003), rather than from substantial methodological or philosophical reasoning.

Kinds and mechanisms

The homeostatic property cluster (HPC) theory of natural kinds originally developed by Richard Boyd (1989, 1991, 1999) moves the focus from the concepts 'essence' and 'law' to the concept of mechanism as the heart of the notion of natural kind. The central idea behind the HPC approach is that classifications should be based on the causal mechanisms that create and sustain the similarities between the entities to be grouped into kinds. Thus, according to Boyd, a scientific kind is composed of (α) a cluster of properties shared by a group of entities, and (β) a homeostatic causal mechanism responsible for the co-occurrence of these properties (Boyd 1999, 67).

Let us illustrate the theory with a simple example. Biological species are perhaps the most intuitively striking (although nowadays highly controversial) examples of natural kinds. According to Boyd, the fact that the members of a species share a set of prototypical physiological and behavioral properties is primarily explained by the mechanisms of reproduction and inheritance (Boyd 1991, 142; 1999, 67). Siberian jays only mate with other Siberian jays, rather than with crows, great tits, or other birds. The circulation of genetic material within a breeding population ensures that most Siberian jays are characterized by a set of species-typical properties. The theory is easily applied also to phenomena in the human sciences (Mallon 2003): for instance, members of an ethnic group are usually subject to social norms that determine the boundaries of proper behavior and appearance, and thus create similarities between the members of the group. These norms are sustained and enforced through a variety of social mechanisms, which thus support the clustering of properties. Furthermore, this clustering of properties guarantees that the corresponding classifications do not appear completely arbitrary.

The mechanistic theory turns the essentialist view of classification on its head: the fact that a kind has a set of apparently essential properties is a thing to be explained, rather than the thing doing the explaining. A homeostatic mechanism explains why a set of entities share a set of properties, rather than a set of essential properties explaining why an individual has a certain observable property. A central principle of the mechanistic theory of classification is that reality, both natural and social, contains causal mechanisms rather than metaphysically primitive Aristotelian essences. From a methodological point of view, the mechanistic theory clarifies the often muddled distinction between mere description and explanation of phenomena: while much of social science often stays at the level of description of the explanandum phenomenon (α), genuinely explanatory theories require inquiry into the causal mechanisms (β) sustaining the phenomena. Knowledge of mechanisms over and above knowledge of patterns and regularities is needed in order to ground inferences to alternative possibilities (what-if-things-had-been-different -scenarios) concerning the object of study. Such counterfactual inferences are necessary for true explanatory import as well as for determining reliable ways of intervening on the target (Woodward 2003) and for extrapolating the findings into new contexts. Stressing the importance of knowledge of mechanisms is a major advantage of the mechanistic approach. The emphasis is important for countering our stubborn tendency to slide back to an essentialist order of explanation (Gelman 2003; Strevens 2000; Prasada & Dillingham 2006).[5]

The mechanistic approach also nicely incorporates the intuition that explanations explain by revealing how something manifest and superficial results from something

[5] As has been suggested by several authors, the essentialist bias of human cognition may have served us well in our evolutionary past, but it certainly has proven to be a hindrance to clear thinking about scientific explanation and theory formation. Scientific explanations should track causal and constitutive dependencies between things in the world (Woodward 2003), and essentialist folk-explanations fail to meet this criterion.

deeper. Category can be used to explain the manifest properties of its members by alluding to a less-visible, 'deeper' feature of the category. This is the intuitively attractive central tenet of critical realism. However, according to the mechanistic approach, these deeper features are not metaphysical kind-essences, but instead the underlying mechanism responsible for the co-occurrence of the observed properties (Hedström & Ylikoski 2010).[6] The mechanistic theory thus circumvents the philosophical muddles related to the charged but ill-understood concepts of essence and essentialism.

Moreover, the mechanistic theory accounts for the intuition that some classifications are more natural than others. Typical examples of natural kinds are characterized by sets of properties sustained by a stable and causally robust mechanism. For example, the characteristic properties of noble metals (e.g., reflectivity, density, malleability) are constituted by the lattice structure of the atoms. This structure is extremely robust with respect to disturbances, and produces the same observable properties in a wide range of background conditions. Mechanisms underlying social phenomena, on the other hand, are often context-sensitive, reflexive, and fragile. The observable common features of members of an ethnic group do not generally depend only on the intrinsic properties of the group members themselves, but instead result from interactions with non-members (boundary work), and are therefore sustained by context-sensitive and extrinsic mechanisms. The implication is that the naturalness of kinds comes in degrees: kinds of social categories and phenomena are often not as robust as natural

[6] Although our approach seems to come close to critical realism, we take issue with the way in which critical realists characterize the central concepts of causality, mechanism, explanation and emergence. Most importantly, we think that the notion of conceptual necessity employed by critical realists is a poor model for the natural necessity of causation. Dealing with these issues would go well beyond the scope of this paper, however.

ones, but they can still be based on causal mechanisms and be conceived as genuine kinds worthy of scientific inquiry.

In the following sections we apply the mechanistic theory of classification to the problematic cases of interactive human kinds. We show how the theory helps to make sense of, and dispels pseudo-problems relating to, socially interactive classifications, especially in cases in which the mechanisms sustaining the property-cluster include both social and biological parts.

3 Classifications in the human sciences

Classifying people has been a central task in the human sciences since their inception. The identification, classification and prevention of "pathological" phenomena such as suicide, prostitution and vagrancy were considered the principal aims of the new sciences of man and society since their establishment. These aims were to be achieved by imitating the basic methodology of the natural sciences: by formulating law-like generalizations that serve as the basis of prediction, explanation and intervention. Scientific ways of classifying people have therefore always been closely linked to the structures of governance and power – a link that has not gone unnoticed by historians and sociologists of science. The sociological study of scientific classifications has made a major contribution to the reflexive self-understanding of social scientists (e.g., Merton 1967; Foucault 1961, 1976–1984; Bowker & Leigh Star 1999; Conrad 2007).

As argued in section two, one's views on the foundations of scientific classification are connected to the understanding one has of the nature of inductive inference and explanation. Classifications in the human sciences, human kinds, exhibit a number of interesting properties. First, their existence and validity are dependent on socio-

cultural practices, and the classifications are therefore limited in space and time. Kinds such as 'members of the conservative party' and 'criminal' are context-dependent in ways that 'carbon atom' or 'Siberian jay' are not. The existence of a cluster of co-occurring properties and the corresponding classification presuppose a socio-cultural niche. This *context-dependency* has been a significant motivating factor behind social-constructionist claims.

Secondly, many medical, psychological and social-scientific classifications straddle scientific theorizing and every-day classification. Human kinds often influence social interaction and governance, and are thus *normatively and politically loaded*. At the same time, they are products and targets of scientific research and therefore seemingly objective and natural. Depression, autism and ad/hd are all classifications created for diagnostic purposes, to facilitate prediction, explanation and intervention. Yet, at the same time they have given rise to new ways of thinking about abnormal behavior and have limited the autonomy and responsibility of those classified.

Thirdly, members of human kinds are *intentional agents*. Whereas the objects of investigation in the natural sciences can mostly be conceptualized as inherently passive things manipulated by the researcher, human subjects can become aware of their status as objects of inquiry and change their behavior accordingly. For this reason, Hacking refers to classifications in the human sciences as *looping kinds*.[7]

The second and third properties combined make classifications in the human sciences potentially normatively and politically loaded. Almost all classifications of humans at least imply a distinction between the normal and the pathological, rather than simply a value-neutral distinction between the typical and the atypical. The inherent normative

[7] In his recent work, Hacking has also employed the phrase 'interactive kind.'

dimension makes human classifications powerful motivators for action. This normativity of classifications is evident in the applications of labeling theory in criminology and sociology of deviance. The normative force of classifications explains processes characterized as *secondary deviance*: Human kinds are kinds to which people wish or do not wish to belong. Being labeled as deviant often sets in motion processes in the subjects themselves as well as alters their interactions with "normals." These processes often result in further deviant behavior (cf. Lemert 1967; Paternoster & Iovanni 1989; Link et al. 1989).

All three features of human kinds mentioned above are in play in the most interesting feature of interactive human kinds, *the looping effect*. When a *new* way of classifying forms of being or behaving emerges, the people classified may change their behavior in a way that makes the original characterization invalid. Thus the classification has to be changed accordingly, which may prompt new changes in targets of classification. This constitutes the looping effect (see Fig 1), which often turns kinds into moving targets.[8] As an example, consider a young black man becoming aware of a hypothesis according to which the genetic make-up of Afro-Americans predisposes them to crime. An understandable reaction might be: *"So I am a born criminal! No point in even trying to stay away from all those things my mom told me not to do."* (Hacking 2004, 298; cf. Singh & Rose 2009). It is equally likely that the reaction will be quite the opposite. A classification can thus function as a self-fulfilling or self-defeating prophecy regardless of the nature, or even reality, of the cluster of properties on which it was originally based. Here it can be seen how looping effects go beyond the basic labeling theory perspective: whereas labeling theorists have been mostly con-

[8] The fact that the target is moving does not necessarily render prediction impossible. Herbert Simon discussed an example of published poll results affecting the voting behavior and proved that the pollster can, in principle (if he or she knows the exact form of the response function of the polled public), create an accurate self-fulfilling prediction of the voting outcome. (Simon 1954.)

cerned about the consequences that classification has for its targets, studies on looping effects must also pay close attention to the consequences that the changing behavior has for the knowledge included in the classification. The two-way interaction between knowledge and behavior often results instability of knowledge and sometimes results in 'making up people', the creation of new social phenomena (Hacking 2002). [Figure 1.]

The looping effect is a typical feature of human kinds and is often related to ways of being and behaving that are considered socially problematic. Classifications (and the corresponding phenomena) such as Asperger's syndrome, AD/HD, and bulimia all exhibit significant looping. Sometimes the cluster of properties targeted is partially constituted by biological factors. In these cases, the looping effect makes it hard to empirically and conceptually distinguish this "innate" biological component from the social, "constructed", component. This difficulty has invoked politically and morally concerned social scientists to make sweeping claims about the social construction of entire phenomena (e.g., Burr 1995; Gergen 1996), which are in reality sustained both by entangled webs of social influence and robust physiological factors.

Ultimately, the phenomena of feedback and homeostasis are ubiquitous in the social realm: Classifications are linguistic entities and therefore social institutions. Social reality is constituted by these institutions, which are maintained by feedback connections. The centrality of this dialectic feedback becomes evident when one considers social reality as a giant self-fulfilling prophecy, sustained by *bootstrapping induction*, in Barry Barnes' (1983) terminology. The stability and the spontaneous order of social reality is founded upon shared and self-reinforcing expectations concerning appropriate behavior: Shared expectations give rise to convergent behavior, which feeds back to the expectations thus creating a feedback loop between the description of so-

cial behavior and the behavior itself. Human scientific knowledge is an essential ingredient in the formation of these expectations. New conceptions of the nature of man and society alter our expectations concerning the likely behavior of others, and these expectations influence our behavior. This, in turn, enforces the newly informed expectations of others. Social-scientific knowledge is part of the social reality investigated in the social sciences and new knowledge can therefore always potentially alter its target (cf. Giddens 1991; Osborne & Rose 1999; Callon 1998). The Barnesian viewpoint is useful because it highlights the fact that feedback often actually stabilizes the cluster of behaviors on which the classifications are based, in contrast to the interesting unstable and ephemeral cases in Hacking's studies, in which the feedback makes the classified phenomenon a moving target.[9]

Although in the paradigmatic cases looping is caused by conscious reactions of agents, it is obvious that also many things totally lacking awareness of anything are in causal interaction with the scientific knowledge about them. Livestock, pets, houseplants, and strains of bacteria are all causally dependent on our ways of conceptualizing them. If a beetle is classified as a pest, the behavior of the beetles will change (or cease to exist), due to changes in their environment brought about by the change in the classification. The ubiquity of feedback in general has been one reason why Hacking limits his in-depth investigations of looping kinds mostly to cases in which the feedback mechanisms go *through the awareness* of the classified individuals. The phenomenon changes because the targets of the classification become aware of its content and change their behavior in consequence.[10] Without this limitation, almost

[9] Here it must be borne in mind that although the stability and homeostasis of phenomena is typically maintained by negative feedback mechanisms, not all social feedback processes must be looping mechanisms involving scientific re-description.

[10] Hacking briefly discusses groups such as hyperactive children and women refugees as cases of the looping effect without explicit acknowledgment of the classification on the side of the classified.

all objects of scientific research would be interactive in some sense. We argue, however, that a satisfactory systematic account of the looping effect cannot be formulated without relaxing this constraint. First, with regard to the mechanistic theory of classification, the constraint is quite arbitrary. Second, as we argue in the next section, it rules out many sociologically interesting forms of the looping effect.

4 Feedback-mechanisms as the basis of the theory of looping kinds

In this section, we use a set of examples of different kinds of feedback exhibited by the social-scientific re-description of people in order to build a general typology of feedback mechanisms of looping. Hacking's treatment of looping kinds consists of interesting case studies, with a few generalizable remarks and isolated stabs at theorizing. This is completely intentional – Hacking is explicitly skeptical about the possibility of generating a substantial and generalizable theory of the looping effect. He thought that forms of looping are too varied and always highly dependent on specific socio-cultural contexts. Nevertheless, we are optimistic in that something more can be said. Simply pointing out the dialectic relationship between the classification and the targets classified does not provide grounds for making inferences beyond what has actually happened and is therefore not explanatory in itself (cf. Merton 1948). More generally, a purely "semantic" approach to classifications and the behavior that they alter is not satisfactory: in order to understand how classifications change behavior, it is not sufficient just to note that there is a two-way relationship between the content of scientific classifications and the reactions of the classified. Instead, the causal processes behind the propagation of scientific representations and psychological and sociological reactions to them ought to be examined in such detail so as to enable rea-

---

However, his focus, and the emphasis of the subsequent discussions, is firmly on cases in which awareness plays a crucial role.

sonable inferences to possible alternative courses of events and even prediction. In order to gain further understanding of interaction between classification and behavior, it is not sufficient only to label the phenomenon "looping mechanism." Therefore we suggest that Hacking's "theory" of looping kinds could be seen as a general sketch to which more substantial cognitive and social mechanisms can be added according to the case at hand. Our mechanistic theory of classification can be used as a general framework for these middle-range theories of interactive kinds in the human sciences. This is the central claim of our paper:

> *In theorizing about looping kinds we should aim at middle-range theories of kinds by describing and analyzing the social and cognitive mechanisms that mediate the looping effect.*

We emphasize the importance of explanatory middle-range theories in order to steer the discussion away from the philosophical debate concerning the conceptual and ontological nature of social reality. The looping effect of human kinds is an empirical causal phenomenon within society, not a conceptual curiosity of the social realm. As such, looping cannot be used as an argument against a realistic understanding of the human sciences. The remainder of this article is an attempt to sketch a framework for different types of feedback mechanisms, and to draw up some tentative hypotheses about their connection to different forms of looping effects. Our aim is, to a certain extent, similar to that of Charles Tilly's taxonomy of mechanisms of social boundary formation and change (Tilly 1998; 2004). Also Tilly's account of the mechanisms of social boundary construction and change aims at providing causal understanding of the creation, maintenance and effects of social categorizations. However, our specific focus on the feedback between behavior and scientific knowledge leads us to delineate the relevant mechanisms differently. Moreover, many of the mechanisms dis-

cussed by Tilly (such as "conversation", "imposition" and "relocation") appear more as descriptive labels for effects or achievements, rather than descriptions of underlying mechanisms that bring about such social phenomena.[11] Although also our own schemas remain tentative, we hope that they will serve as an inspiration for further explanatory theorizing about looping kinds.

Different kinds of feedback mechanisms

If we accept that all human scientific knowledge is, in principle, potentially interactive, we can view Hacking's case studies as exemplifying a special form of the looping effect in which awareness of the classification by those classified leads to instability in the cluster of properties picked out by the classification. However, a general theory of looping effects should also accommodate cases in which the awareness of the classified is not an essential element in the feedback mechanism, and in which the feedback stabilizes rather than destabilizes the corresponding property cluster. Feedback between scientific knowledge and its objects in general is widespread and has been studied from various sociological perspectives. In our age of neuroscience and genetics, the impact of the emergence of new forms of biomedical knowledge and technologies to our conception of the self as well to our social relations has been an object of intense study (e.g., Clarke et al. 2003; Conrad 2007; Rose 2007; Shostak et al. 2007; Singh & Rose 2009). However, only a part of such scientific knowledge is truly looping in the sense that its validity may be compromised by its causal effects on those that the knowledge is suppose to apply to.

[11] There are also other examples of "thicker" mechanistic theorizing about categories. For instance, Bruce Link's modified labeling theory aims at constructing substantial causal models of secondary deviance. (Link et al. 1989).

Another societally and theoretically important example of true looping is the encroachment of economic conceptualizations into various aspects of social life previously thought of as distinct and independent of markets. Below we give examples of different kinds of feedback mechanisms related to an economics-inspired way of re-describing and conceptualizing people – cases of making up homo economicus (cf. Ferraro et al. 2005). Although the following are not paradigmatic cases of classification in that the re-described population is not partitioned into different kinds, they are examples of situations in which there exist empirically well-supported and detailed accounts of the mediating feedback mechanisms. These accounts are conducive to our purposes in their explicit focus on the social and cognitive mechanisms mediating the looping effect. Some of the following examples draw on experimental work in social psychology and some on field studies of natural experiments. Such quasi-experimental evidence is highly valuable in building theories of causal mechanisms. We will then generalize these special cases to a schema, which we believe to be applicable to similar feedback mechanisms in action in more paradigmatic cases of describing people with kind-terms proper.

Example 1: Motivation crowding out

One example of a socio-politically important form of feedback mechanism, which does not necessarily incorporate awareness of the classification, is the phenomenon of *motivation crowding out*. David Hume once proclaimed that when social institutions are designed, people should be treated as self-regarding knaves in order to make the institutions immune to abuse (Hume 1964/1741, 117–119). This argument has found a home in the economics-based tradition of institutional design. The guiding principle is

to conceptualize people as rational agents competently pursuing their interests and then to formulate incentives in a way that would steer these rational agents to act in accordance with some socially beneficial outcome. Usually such incentives are monetary, and free-riding as well as principal-agent problems (e.g., moral hazard) are controlled by means of sanctions and bonus schemes.

This kind of incentive-based social engineering may have adverse long-term consequences. In some cases it has already produced counterproductive results, largely because such applications of rational-choice theorizing do not take into account the important feedback mechanisms of social-scientific concepts. When the decision situation of agents is framed in terms of explicit monetary incentives, the way the agent conceptualizes her situation, and herself in that situation, changes accordingly. Conceptualizing the consequences of decision alternatives in terms of expected gains and costs diminishes the intrinsic motivational force of such things as civil virtues, expressive rationality, and the intrinsic value of the possible actions. When everything is given a price, obligations, commitments, rights, and the common good are classified as goods to be consumed within the limits of the budget constraint.

Explicit incentives that render the intended manipulation transparent to the agent constitute another mechanism leading to the crowding out of intrinsic virtues. This diminishes the agent's sense of autonomy and also usually leads to more cynical, calculative behavior. The conceptualization of people as rational economic agents can be a powerful self-fulfilling prophecy.

The crowding out of intrinsic motivations may lead to deterioration of the desired behavior when incentives are introduced. What happened in an Israeli day-care center is an example of this phenomenon: the rate of parental lateness in picking up their

children increased when a monetary penalty was imposed. The hypothesis is that this was due to the fact that the parents no longer conceptualized being on time as a responsibility, but they now saw extra time as a good that could be purchased. Likewise, when the Swiss government decided to give a monetary reward to any community willing to accept nuclear waste, public willingness to accept the waste declined. A motivation-crowding hypothesis is that this was because people in a given community no longer conceptualized the acceptance of nuclear waste as a civic duty for the greater good of the nation, but saw it as a service, the price of which should be calculated against the possible losses resulting from a loss of image and possible local environmental effects. (Frey & Jegen 2001; Ostrom 2000.)

The conception of man as a rational economic actor may thus be a self-fulfilling prophecy due to the feedback mechanism of the crowding out of intrinsic motivations. Crowding out is the socio-psychological mechanism that sustains the cluster of properties characteristic of the newly created homo economica.

However, it is likely that neither the Israeli parents nor the Swiss villagers were really aware of this change in the concepts used to describe and classify them. In a sense, the space of possible actions changed with the introduction of the incentives. Buying child-free time and selling storage space were not really possible until the relevant goods had a price, but this change was not due to any novel ways in which the agents could conceptualize themselves. *Motivation crowding out works through alterations in the practical reasoning of the classified agents, but it does not require awareness of the classification.*

Example 2: The norm of self-interest

Let us now compare the previous form of homo-economicus feedback to a norm-mediated mechanism related to the concept of the rational economic agent. According to the social psychologist Dale Miller (1999), the pursuit of self-regarding interests is not a deep feature of human nature, but a cultural norm bound to specific societal and scientific conditions, and thus part of the self-fulfilling structure of the social reality. The *norm of self-interest* is based on and sustained by the widely accepted pseudo-scientific folk theory about the driving forces of human behavior, and is tied to the widely publicized view of "the human nature" found in popularized expositions of economics and evolutionary psychology.

The norm of self-interest influences behavior via two mechanisms. First, it is a shared and pervasive assumption about the factors that motivate other people's behavior, and, second, it dictates the proper attribution of responsibility for rationalizing one's own actions. Miller's empirical results suggest that people commonly overestimate the importance of generally self-regarding and especially monetary preferences in motivating action. These shared biased expectations are a self-fulfilling prophecy in that they lead to increasingly self-regarding and calculative action – people do not want to be fools exploited by their self-regarding and calculative fellows.

The other mechanism works through changes in the (communally) shared resources for rationalizing action. In groups in which the norm of self-interest is prevalent, altruistic acts, such as monetary donations, have to be rationalized (to oneself and to others) with reference to a warm-glow feeling, an expected increase in social status, or the need to alleviate feelings of guilt. The norm of self-interest need not (primarily) alter the motivating factors of action (which are often not within the reach of introspection), but it can influence the expectations concerning the kind of behavior that is considered appropriate. Of course, these mutual expectations, in turn, steer the social

behavior of people, and thus alter the way in which social reality is created through bootstrapping induction. The importance of this feedback channel is revealed in experiments in which giving the subjects more conceptual or other resources for rationalizing their other-regarding behavior (such as giving symbolic tokens in exchange for donations) reduces self-regarding and calculative action (Ratner & Miller 2001). The picture discernible in popular science that treats man as a self-regarding maximizer can thus function a self-fulfilling prophecy by altering the set of actions deemed appropriate, or even inherently understandable. Biased expectations of strategic, self-regarding behavior and changes in the norms allocating the burden of rationalization one's actions are two more socio-psychological mechanisms creating and sustaining the cluster of behaviors characteristic of the homo economicus.

Example 3: Selection and filtering mechanisms

In principle, the feedback mechanisms of human scientific knowledge need not involve any changes in the behavior of the individual agents being classified. Certain ways of describing and classifying may lead to changes in the environment of the people concerned in a way that crowds out some individuals. For example, ever since the new-public-management movement began to take hold, numerous populations of, say, university staff have undergone shifts in composition due to the weeding out of "inefficient" personnel unable to adapt to constantly changing administrative structures and production targets (for a general review of how economics inspired management practices can become a self-fulfilling prophecy, see Ferraro et al. 2005). Changes in the behavior of a group do not have to be caused by changes in any of the individuals; it is enough that the set of individuals comprising the group has changed in some systematic manner. Such selection mechanisms operate on the population level. Purely institutional changes can also alter the behavior of a group through the

screening of new candidate entrants, rather than weeding out already selected members. As in biological cases, such selection mechanisms may create the illusion of a kind-essence, intrinsic to each and every individual of that population, but in reality, it is the population level mechanisms that sustain the clustering of properties characteristic of membership in that population.

Feedback mechanisms and looping effects

We are now in a position to suggest a few dimensions along which feedback mechanisms of human scientific classifications may differ, and thus result in different kinds of looping effects.

(1) Feedback can either stabilize or destabilize the property cluster on which the classification is based.

(2) The primary causal link of the feedback mechanism can either consist of changes in the cognition of the classified agents, or of changes in their environment (institutional changes). Both channels can, of course, operate at the same time.

If feedback involves changes in individual cognition, then:

(3) The agents' behavior can change due to (i) awareness of the classification, (ii) changes in the surrounding concepts indirectly brought about by the classification, or the (iii) unconscious adaptation of the behavior of classified agents caused by the introduction of the classification.

Hacking's case studies concern a subgroup formed by a particular combination of the aspects above: He has focused on destabilizing looping effects caused by cognitive changes in the individuals, requiring explicit awareness of the classification. On the other hand, motivation crowding out results from institutional changes, which lead to

conceptual changes, but it does not require awareness of the re-description and is, in the long run, most likely to be stabilizing. Furthermore, the norm of self-interest leads directly to changes in concepts and in beliefs, requires at least some level of change in self-conceptualization (but not necessarily explicit awareness of the re-description), and is also mostly stabilizing. Finally, selection mechanisms work through structural changes and do not require cognitive changes in individual agents.

[Table 1]

It is clear that the above schema is still very coarse and cannot be used as such to provide explanatory middle-range theories (such as the theory of motivation crowding out). Nevertheless, we believe that it has analytic value, and that it can be used to bring desperately needed clarity to many social-constructionist debates. Moreover, we think that our mechanistic theory of classification is preferable to its alternatives, essentialist approaches and social constructionism. In contrast to forms of essentialism (including critical realism), our approach gets the order of explanation right: in the three cases described above, the classified people end up behaving in ways characteristic of homo economicus, but the typical properties of the economic actor are not explanatory in themselves. Instead, they are parts of the cluster of properties maintained by an underlying causal mechanism. The distinction is important, because in order to understand the differences between the three apparently similar cases of homo economicus behavior, the properties of mechanisms have to be investigated. Locating a specific looping classification within our schema of different kinds of mechanisms will suggest hypotheses about what kind of interventions might be effective, and which ones likely ineffective, in changing the behavior thus classified.

Our theory is also superior to constructivist approaches because it does not commit us to the superficial distinction between natural and constructed phenomena, but instead it can accommodate homeostatic mechanisms that consist of both kinds of sub-mechanisms. In order to demonstrate the analytic value of our framework we will now apply it to a particularly problematic case of the looping effect, one in which the classified phenomenon is clearly a result of both neurophysiological etiology and social feedback. This interaction renders a purely social-constructionist perspective at best insufficient, and probably misleading. Although most researchers' approaches in the social sciences fall between the radical extremes, naïve essentialism and strong social constructionism, these mid-way positions are seldom made explicit. Our theory is an attempt to explicate one such plausible position.

5 The social construction of autism

Autism is a disorder of neural development characterized by problems in reciprocal social interaction and communication, and by rigidity of behavior. Contemporary classifications of mental disorders (ICD-10 and DSM-IV) recognize several different autistic conditions, which are together said to comprise the autism spectrum. Probably the most well known conditions on the spectrum are childhood autism and Asperger's syndrome. In addition to the distinctions between the different autism spectrum disorders (ASDs), severe cases of autism are also often distinguished from milder ones, i.e., high-functioning autists.

The classification of ASDs is controversial due to the fact that there are still few solid facts about the etiology or the neurological basis of autism. However, the different disorders have quite similar diagnostic criteria. ASDs are typically diagnosed in early childhood, and the difficulties they cause remain throughout life. People with ASDs

usually have an impaired social sense, and are at least partially unable to respond to social signals. Shared affect and the interpretation of facial expressions are often problematic for autistic individuals. Problems of communication are also among the core features of these disorders: in severe cases of autism, speech perception and the use of language are often limited or delayed. The verbal intelligence of people with Asperger's syndrome is usually high, but their language use is often atypical. In addition to experiencing problems in social interaction and communication, autists often show preference for repetitive activities and have narrow interests. (Roth 2010, Ch.2; Bailey et al. 1996.)

Regardless of whether the contemporary classification of ASDs[12] is, in some sense, correct, and whether the disorders have the same genetic or neurological basis, autism spectrum disorders and their classification are an interesting case for the theory of feedback mechanisms introduced in this article. We are realists about the spectrum in the sense that we believe that ASDs are real developmental disorders, and thus partly biological phenomena. Yet autism is as much a social phenomenon as it is a medical one (cf. Eyal et al. 2010; Valentine 2010), and different kinds of feedback and looping are at work in different types of autism disorders.

In severe cases of childhood autism, when the child suffers from major problems of social interaction and communication, the principal channel of feedback between knowledge about autism and the behavior of the child goes through the child's family and caretakers, not the afflicted individual herself. That being said, the behavioral patterns of the autistic individual are still likely to adapt to the changed environment, at least to a certain degree (through behavioral therapy, for example). The changes

[12] For information on the proposed changes to DSM-V on autistic disorders see http://www.dsm5.org/ProposedRevisions/Pages/proposedrevision.aspx?rid=94

could result in further modification of the environment based on knowledge concerning the behavior of the patient, and thus modest looping could occur. In these cases, however, it seems improbable that the social-feedback mechanism would lead to drastic changes in the nature of the condition.[13]

One historical example of this relatively weak feedback connection is the development of knowledge concerning the causes of autism. From the early diagnoses in the 1940s to the 1960s, it was mistakenly thought that the condition was caused by insufficient attachment between the child and the mother. So-called emotionally cold "refrigerator mothers" were thought be responsible (Hacking 1995, 376–377). Although this erroneous hypothesis indubitably traumatized a group of parents of autistic children, it appears not to have had a significant influence on the behavior of the children themselves.[14]

On the other hand, there are cases of ASDs in which the feedback between knowledge and behavior is channeled through the conscious reactions of the afflicted individual. Asperger's syndrome could show this kind of looping: although people with Asperger's tend to have limited social skills, they are not confined outside the social reality. They ("aspies") are aware of their own condition and the scientific knowledge concerning it.[15] Therefore it is understandable that the Asperger's diagnosis is a great relief to many afflicted individuals. Understanding one's condition as a consequence of a known disorder can relieve psychological suffering. On the other hand, for some

[13] Our description of the looping effects is based on the current scientific characterization of ASDs. We ourselves do not aim to suggest anything substantial about the true abilities or capabilities of autistic individuals. Furthermore, we recognize that medical and scientific accounts of the condition fall short of capturing the lived experience of autistic individuals, and should therefore be treated with caution.
[14] Another example of a weak looping effect in autism between the 1960s and 1990s is the disappearance of self-stimulating and self-mutilating behaviors as diagnostic criteria for the disorder (Eyal et al. 2010, 205–211).
[15] Asperger's syndrome often correlates with high intelligence. Afflicted people often work in environments such as universities, where impairments in social sense are not necessarily an obstacle to work performance (cf. Baron-Cohen et al. 2001).

people the diagnosis can trigger a psychological crisis. Different kinds of reactions to the scientific knowledge lead to different kinds of behavior. In some cases, understanding oneself as a person with Asperger's syndrome may make the person avoid social situations even more strongly than before, whereas in others he or she might consciously adopt new effective learning strategies to compensate for the limitations of social cognition.[16] There are also other possible reactions to the classification. Recently, the neurodiversity movement has promoted a more radical approach to Asperger's syndrome by arguing against the idea that Asperger's is a pathology to begin with. According to this self-advocacy movement, the condition should not be considered a disorder, but people with Asperger's are simply behaviorally and neurophysically different from the rest of the population (Eyal et al. 2010, 226–233). It is obvious that these different reactions make a big difference to the development of the behavioral profile of Asperger's syndrome.

[Table 2]

The rise in public awareness of Asperger's syndrome since the 1980s has also created another interesting looping effect. Although diagnosis is hard even for medical experts, it has become a common folk-psychological practice to describe oneself or one's family member as an "aspie." Thanks to these lay diagnoses, many normal but perhaps somewhat asocial people have begun to understand themselves through the Asperger profile, and this has likely had a significant effect on their identity, regardless of the truthfulness of the diagnosis. One could perhaps say that, in these cases, the Asperger diagnosis has spilled over into the surrounding population, as it were,

[16] See Angermeyer & Matschinger 2005; Jorm & Griffiths 2008 and Read, Haslam & Davies 2009 for a recent discussion on the stigmatizing effects of mental illness diagnosis.

and created novel feedback mechanisms influencing people's identities and their behavior.

As mentioned above, there is still little knowledge of the physiological foundations of different autism disorders. Some researchers argue that the different autism-spectrum conditions are manifestations of the same neurological abnormality, whereas others believe that several different brain pathologies are responsible for the different conditions. It should be noted, however, that the existence of physiological differences between ASDs is not necessary for the different conditions to be significantly different kinds of phenomena: for example, the different kinds of feedback mechanisms associated with severe autism and Asperger's syndrome produce and amplify differences between the symptoms of the two conditions. Different looping effects cause the property clusters constituting the symptom profiles to diverge. Most importantly, the design of efficient treatments (and decisions of whether "treatment" is necessary in the first place) requires knowledge of the feedback mechanisms sustaining the observable profile of the phenomena. Although it is quite plausible to assume that different ASDs share a common neurological core, sensible interventions might target the sociocultural sustaining factors responsible for the differences in observed conditions.

However, it is necessary to distinguish between different feedback mechanisms in order to assess the extent to which autism could be called a socially constructed phenomenon. The fact that the whole autism phenomenon was unknown until the 1940s is a *prima facie* argument in favor of the constructionist position. It was only after Leo Kanner and Hans Asperger made their diagnoses that has there been a group of people described in the language of autism. These people were earlier thought of as simply retarded or subnormal. In Hacking's terms, scientific research "made up a new

kind of people", the autists (Hacking 1995, 375–379). The feedback mechanisms described above also give reason to call many ASDs socially constructed phenomena. The properties of the disorders are not independent of our practices, but instead are connected to scientific research on autism on the one hand, and to the values of society on the other.

Social-constructionist claims should be taken with a grain of salt, however. As Hacking (1999) points out, constructionist views often have little in common in terms of content, but often work as consciousness-raising devices that are used to highlight societally relevant dependencies between the "constructed" phenomenon and its social context. What makes these claims problematic is that they often sound like a form of idealism: they imply that social reality magically emerges from people's thoughts and can therefore be changed simply by changing these thoughts. It is for this reason that we believe social constructionism to be of little use in emancipatory research. The feedback mechanisms of human scientific knowledge are not in the eye of the beholder, nor are they confined inside people's heads. The looping effects are an empirical causal phenomenon that ought to be studied with empirical methods (for the case of autism, see especially Liu et al. 2010). This should be borne in mind in the search for forms of effective intervention on allegedly constructed phenomena.

Recognizing the causal nature of looping is especially important in cases such as ASDs, in which the causal mechanisms behind the phenomenon consist of both social and biological factors. Straightforward talk of the social construction of the phenomenon is at best misleading in such cases, and often simply false. It is very likely that ASDs are characterized by genetically determined deviations in the ontogenetic process. As suggested above, treating severe cases of childhood autism as socially constructed phenomena would even be ethically suspect in that it could lead to inefficient

recommendations for medical intervention and needless blame on the parents of autistic children.

6 Conclusions

Classifications in the human sciences do not pick out differences between people that flow from kind-essences or the laws of nature. Furthermore, as looping effects show, the properties and behavior of the people classified are not independent of classificatory practices. However, this does not mean that theory and concept formation in the human sciences and the natural sciences should be radically different: the classical view of natural kinds does not apply to much of natural science, either. Moreover, controversial classifications in the human sciences often capture phenomena that are very real in the sense that they have substantial psychological and social consequences, contrary to what constructionist approaches often appear to imply. We have argued that intuitions concerning natural kinds, essences, and the one correct taxonomy of the reality are mainly just theoretical reflections of inherent biases in human cognition, not scientifically justifiable principles or aims of scientific practice.

The mechanistic theory of classification provides an account that fits the practices of classification in both the natural and the human sciences. The emphasis on mechanisms is not only compatible with the concept of looping kind, it also steers the discussion away from philosophical puzzles concerning the fundamental ontology of the social world, and towards explanatory middle-range theories. The formulation of such mechanistic middle-range theories would be especially important when the classified phenomenon is constituted of both biological causes and social feedback. In such cases, replacing hopelessly vague theorizing about the social construction of the phenomenon with explanatory causal accounts of the feedback-mechanism would in-

crease our understanding and improve our ability to influence the phenomenon. The mechanistic theory also circumvents the philosophical pitfalls related to essentialism. While most social scientists are likely to avoid both strong essentialist as well as constructionist positions, our aim has been to make explicit the reasonable mid-way option in the form of our mechanistic theory. We have thus attempted to open up conceptual space for explanatory research, which could, in addition to current case study methodology, also utilize other methods, such as statistical analysis, formal models, and possibly agent-based simulation (cf. Hedström 2005).

In principle, all human scientific knowledge is potentially interactive. Social reality is largely constituted of mutually convergent beliefs, self-fulfilling prophecies, and the normative practices supporting this convergence. Human scientific knowledge is an integral part of this shared social reality. The ubiquity of feedback does not imply an argument against methodological naturalism, however. As our mechanistic theory of classification suggests, the mechanisms responsible for the feedback are causal, "objective" social and cognitive phenomena. They are dependent on the conceptualizations of social scientists only via other feedback-mechanisms, not because of "internal" connections between the meanings of concepts.

References

Angermeyer MC, and H Matschinger. 2003. "The stigma of mental illness: effects of labelling on public attitudes towards people with mental disorder." *Acta Psychiatrica Scandinavica* 103, 304–309.

Bailey, Anthony, Wendy Phillips, Michael Rutter. 1996. "Autism: Towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives." *Journal of Child Psychology and Psychiatry* 37: 89–126.

Barnes, Barry. 1983. "Social Life as Bootsrapped Induction." *Sociology* 17: 524-545.

Baron-Cohen, Simon, Sally Wheelwright, Richard Skinner, Joanne Martin, Emma Clubley. 2001. "The autism-spectrum quotient (AQ): Evidence from aspeger syndrome/high-functioning autism, males and females, scientists and mathematicians." *Journal of Autism and Developmental Disorders* 31/1: 5-17.

Bourdieu, Pierre 1992. *Invitation to reflexive sosiology*, Chicago and London: University of Chicago Press

Boyd, Richard. 1989. "What Realism Implies and what it does Not." *Dialectica* 43: 5–29.

— 1991. "Realism, anti-foundationalism and the enthusiasm for natural kinds." *Philosophical Studies* 61: 127–148.

— 1999. "Kinds as the 'Workmanship of Men'." In *Rationalität, Realismus, Revision,* edited by Nida-Rümelin, Julian. Berlin & New York: Walter de Gruyter.

Bowker, Geoffrey, and Susan Leigh Star. 1999. *Sorting Things Out: Classification and its consequences*. Cambridge: MIT Press.

Burr, Vivien 1995. *An Introduction to Social Constructionism*. London: Routledge.

Callon, Michel. 1998. Introduction: The Embeddedness of Economic Markets in Economics. In *The Laws of the Market*. edited by Callon, Michel. Oxford: Blackwell: 1-50.

Clarke, Adele E., Janet K. Shim, Laura Mamo, Jennifer Ruth Fosket, and Jennifer R. Fishman. 2003. "Biomedicalization: Technoscientific Transformations of Health, Illness and U.S. Medicine." *American Sociological Review* 68: 161-194.

Conrad, Peter. 2007. *Medicalization of Society: On the Transformation of Human Conditions into Treatable Disorders*. Baltimore: Johns Hopkins University Press.

Craver, Carl. 2009. "Mechanisms and natural kinds." *Philosophical Psychology*, 22, 5, 575-596.

Eyal, Gil et al. 2010. *The Autism Matrix. The social origins of the autism epidemic*. Cambridge: Polity Press.

Ferraro, Fabrizio, Jeffrey Pfeffer and Robert Sutton. 2005. "Economics Language and Assumptions: How Theories can Become Self-Fulfilling." *Academy of Management Review* 30(1): 8-24.

Fodor, Jerry. 1974. "Special sciences (Or: The disunity of science as a working hypothesis)." *Synthese* 28, 97–115.

Foucault, Michel. 1961. *Folie et déraison. Historie de la folie à l'age classique.* Paris: Plon.

— 1976–1984. *Histoire de la sexualité.* Vol 1–3. Paris: Gallimard.

Frey, Bruno & Reto Jegen. 2001. "Motivation Crowding Theory." *Journal of Economic Surveys* 15: 589-611.

Gelman, Susan. 2003. *The Essential Child*. New York: Oxford University Press.

Gergen, Kenneth. 1996. "Social Psychology as Social Construction: The Emerging Vision." in *The Message of Social Psychology: Perspectives on Mind in Society*. edited by C. McGarty and A.Haslam. Oxford: Blackwell

Giddens, Anthony. 1984. *The Constitution of Society*. Berkeley: University of California Press.

Giddens, Anthony. 1991. *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Stanford: Stanford University Press.

Goffman, Erving. 1963. *Stigma. Notes on the management of spoiled identity.* London: Penguin.

Goodman, Nelson (1983/1955). *Fact, Fiction and Forecast*. 4th ed.. Cambridge & London: Harvard University Press.

Grene, Marjorie, and David Depew. 2004. *The Philosophy of Biology. An episodic history*. Cambridge: Cambridge University Press.

Griffiths, Paul. 1997. *What Emotions Really Are.* Chicago: University of Chicago Press.

Hacking, Ian. 2006. "Des classifications naturelles." Lectures at College de France, http://www.college-de-france.fr/media/historique/UPL17441_ihackingcours0506.pdf, 422–430.

— 2004. “Between Michel Foucault and Erving Goffman: Between Discourse in the Abstract and Face-to-Face Interaction.” *Economy and Society* 33/3: 277–302.

— 2002. *Historical Ontology*. Cambridge & London: Harvard University Press.

— 1999. *The Social Construction of What*. Cambridge: Harvard University Press.

— 1995. “The looping effects of human kinds.” in Sperber & Premack & Premack 1995. *Causal Cognition. A multidisciplinary debate.* Oxford: Clarendon Press.

— 1991. “A tradition of natural kinds.” *Philosophical Studies* 61: 109–126.

Hedström, Peter. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge University Press.

Hedströn, Peter and Peter Bearman (eds.). 2009. *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press.

Hedström, Peter and Petri Ylikoski 2010. “Causal Mechanisms in the Social Sciences.” *Annual Review of Sociology* 36: 49-67

Hume, David. 1964/1741. *The Philosophical Works* Vol. 3. Ed. Green, Thomas Hill, and Thomas Hodge Grose. London: Scientia Verlag Aalen.

Jorm A, and K Griffiths. 2008 “The public’s stigmatizing attitudes towards people with mental disorders: how important are biomedical conceptualizations?” *Acta Psychiatrica Scandinavica* 118, 315–321.

Lamont, Michèle and Virág Molnár. 2002. “The Study of Boundaries in the Social Sciences.” *Annual Review of Sociology* 28, 167-195.

Liu, Ka-Yuet, Marissa King, and Peter S. Bearman 2010. "Social Influence and the Autism Epidemic." *American Journal of Sociology* 115 (5): 1387–1434.

Machery, Edouard. 2009. *Doing Without Concepts*. New York: Oxford University Press.

Mallon, Ron. 2003. "Social Construction, Social Roles and Stability." In *Socializing Metaphysics: The Nature of Social Reality*. edited by Schmitt, Frederick. Lanham, MD: Rowman and Littlefield: 327–353.

Merton, Robert K. 1967. *On Theoretical Sociology*. New York: The Free Press.

—— 1948. "The self-fulfilling prophecy." Antioch Review 8: 193-210.

Miller, Dale. 1999. "The Norm of Self-Interest." *American Psychologist* 54: 1053-1060.

Murphy, Dominic. 2006. *Psychiatry in the Scientific Image*. Cambridge: MIT Press.

Osborne, Thomas, and Nikolas Rose. 1999." Do the social sciences create phenomena?: the example of public opinion research." *British Journal of Sociology*, 50, 367–396.

Ostrom, Elinor. 2000. "Crowding Out Citizenship." *Scandinavian Political Studies* 23: 3-16.

Paternoster, Raymond, and Leeann Iovanni. 1989. "The labeling perspective and delinquency: An elaboration of the theory and an assessment of the evidence." *Justice Quarterly* 6:359–93.

Prasada, Sandeep, and Elaine Dillingham. 2006. "Principled and statistical connections in common sense conception." *Cognition* 99: 73–112.

Ratner, Rebecca and Dale Miller. 2001. "The norm of self-interest and its effects on social action." *Journal of Personality and Social Psychology* 81:5-16.

Reydon, Thomas 2009. "Do the Life Sciences Need Natural Kinds?" *Croatian Journal of Philosophy* 9, 167–190.

Rose, Nikolas 2007. *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century*. Princeton, N.J.: Princeton University Press.

Rosenberg, Alex. 2005. "Lessons from Biology for Philosophy of the Human Sciences." *Philosophy of the Social Sciences* 35: 3-19.

Read, John, Nick Haslam, and Emma Davies. 2009. "The need to rely on evidence not ideology in stigma research." *Acta Psychiatrica Scandinavica*, 119(5), 412-3.

Roth, Ilona et al. 2010. *The Autism Spectrum in the 21st Century: Exploring Psychology, Biology and Practice.* London: Jessica Kingsley Publishers.

Laub, John, and Robert Sampson. 2006. *Shared Beginnings, Divergent Lifes*. Harvard University Press.

Samuels, Richard, and Michael Ferreira. 2010. "Why *don'*t concepts constitute a natural kind?" *Behavioral and Brain Sciences 33,* 222–223

Sayer, Andrew. 1997. "Essentialism, social constructivism, and beyond." *The Sociological Review* 45: 453-487.

Simon, Herbert. 1954. "Bandwagon and Underdog Effects and the Possibility of Election Predictions." *Public Opinion Quarterly* 18, 245-253.

Singh, Ilina, and Nikolas Rose. 2009. "Biomarkers in psychiatry." *Nature* 460, 202-207.

Shostak, Sara, Peter Conrad and Allan V. Horwitz 2007. “Sequencing and Its Consequences: Path Dependence and the Relationships between Genetics and Medicalization.” *American Journal of Sociology* 114: 287-316.,

Strevens, Michael. 2000. “The essentialist aspect of naïve theories.” *Cognition* 74: 149–175.

Tilly, Charles. 1998. *Durable Inequality*. Berkeley and Los Angeles: University of California Press.

Tilly, Charles. 2004. “Social Boundary Mechanisms.” *Philosophy of the Social Sciences* 34, 211-236.

Valentine, Kylie. 2010. “A consideration of medicalisation: Choice, engagement and other responsibilities of parents of children with autism spectrum disorder.” *Social Science & Medicine* 71, 950–957,

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford and New York: Oxford University Press.

Tables

Table 1.

Primary causal path Feedback requires	**Individual**	**Institutional**
Awareness of classif.	Hacking-cases	
No awareness	The norm of selfishness	Motivation crowding out

Table 2.

Primary causal path Feedback requires	**Individual**	**Institutional**
Awareness of classif.	Hacking-cases Asperger	ad/hd High-functioning autists
No awareness	The norm of selfishness	Motivation crowding out Serious autism

Figures

Figure 1.