Natural kinds with extended mechanisms

(rough draft, 14.5.2010)

Samuli Pöyhönen
(samuli.poyhonen@helsinki.fi)
Philosophy of Science Group / TINT
University of Helsinki

Abstract

The objects of scientific inquiry are natural kinds. In the human sciences, there are phenomena that seem to challenge this idea, such as culture-bound mental illnesses, extended cognitive systems, and homo economicus – the ideal rational actor. What makes these phenomena puzzling is that although they are interesting targets for scientific research, they are sustained by heterogeneous groups of physiological, psychological, and institutional factors, unlike phenomena in the natural sciences. I suggest treating these problematic phenomena as natural kinds supported by *extended mechanisms*. Nothing in the scientifically interesting conception of natural kind or in the notion of mechanism precludes treating kinds with extended mechanisms as natural kinds. The question of whether to include extended kinds in our scientific ontology must be approached by assessing their explanatory power.

1 Introduction: Problematic natural kinds

In the social sciences, psychology and psychiatry, there are interesting phenomena that seem to be supported by mechanisms that don't respect the spatial boundaries of the system. These phenomena form a very heterogeneous group: culture-bound mental illnesses, extended cognitive systems, and homo economicus – the ideal rational actor, just to mention a few. To illustrate the problems surrounding these "extended" phenomena, let's take a brief look at the eating disorder *bulimia nervosa*: Although recent research has suggested that the propensity to become bulimic is associated with – or perhaps partially caused by – abnormalities in central serotonin mechanisms (Steiger et.al. 2001), it is clear that bulimia is a case of psychiatric illness, whose explanation has an important social component. Bulimia is much more common in contexts, where Western ideas of female beauty are prevalent; the idealization of thin people plays a role in explaining why people get ill. It seems that a satisfactory explanation of bulimia requires making reference to a heterogeneous group of factors. Among these various factors are the social practices maintaining norms regarding beauty, the individual psychological mechanisms for becoming aware of the norms, and psychological and physiological factors that explain people's different susceptibilities of becoming ill.

In this paper, I examine an important problem that complex phenomena like bulimia present to our scientific concept formation and revision practices: How is the real phenomenon, the natural kind constituting the target of scientific research, to be identified in these cases? Can culture-bound mental illnesses be natural kinds although they are clearly at least partially socially constructed phenomena? Or, in the case of bulimia, is the true natural kind constituted only by the psycho-physiological pathology inside the afflicted individuals?

There are several important reasons for treating complex phenomena like culture-bound mental illnesses as genuine natural kinds, and thus deserving scientific attention. Despite their context-dependence, they are perfectly sensible objects for scientific investigation, and the concepts that refer to them are far from conventional or arbitrary. These concepts can support lawlike generalizations and the knowledge stored in them makes medical and policy interventions possible. They seem, in an important sense, to be carving nature by its joints, and thus ought to be thought of as natural kinds. Nonetheless, calling these phenomena natural kinds sounds paradoxical.

I suggest that Richard Boyd's HPC-theory of natural kinds can be used to make explicit what makes complex phenomena like bulimia not seem like natural kinds: according to the HPC-theory, they can be seen as phenomena that rely on *extended mechanisms* consisting of a motley of heterogeneous elements. As suggested by Ian Hacking's (1995a, 1998) work on transient mental illnesses, Satz&Ferejohn (1994) on economic behavior, and Clark (1997, 2008) on the embodied and extended mind, extended mechanisms lurk behind a surprisingly large group of important phenomena. As many of these phenomena are important targets of scientific inquiry, it seems that natural kinds should be allowed to be supported by extended mechanisms. However, this goes strongly against the traditional conceptions of natural kinds. Theories of natural kinds often require that essential properties of natural kinds are intrinsic properties, following from the microstructure of the members of the kind. Adopting a mechanistic approach to natural kinds doesn't necessarily change the picture at all: Also in the literature on mechanisms and mechanistic explanation, it is common to assume that mechanisms have to be somehow internal to the target of explanation.[1]

---

[1] From a psychological perspective, these requirements of internality can be seen as manifestations of a well-known reductivist bias in our thinking, psychological essentialism (cf. e.g. Gelman 2003; Strevens 2000).

In sections two and three of the paper I examine the concepts of natural kind and mechanism, and show that from a philosophy of science perspective, these concepts do not give a priori reasons for why kinds with extended mechanisms couldn't be natural kinds, respectable members of scientific ontology. Section four takes a more systematic approach to explaining with extended mechanisms by assessing the epistemic advantages and disadvantages of scientific explanations that employ extended mechanisms.

2 Various kinds of natural kinds

The concept of natural kind has many uses in philosophy. In metaphysics, the concept features in discussions concerning laws of nature, natural necessity, and essentialism. In these contexts, natural kind terms are often supposed to refer to the fundamental building blocks of reality, in a sense. They tell us what kinds of objects, substances, events and processes there "really" are in the world, as opposed to things referred to by conventional, socially constructed classifications. Many metaphysical theories of natural kinds are wedded to the idea that kinds are individuated by their intrinsic dispositional properties (Bird & Tobin 2008; Ellis 2001). Hence, the essence that the members of a natural kind share is common microstructure. According to such microstructural theories, natural kinds are scarce: they can only be found in physics, maybe chemistry, but certainly not in the human sciences, which investigate contingent phenomena.

However, the original, and perhaps the most significant use for the concept 'natural kind' is in picking out the phenomena that are sensible targets for scientific investigation (Mill 1891). In this usage, natural kind terms are contrasted to terms, whose meaning is exhausted by linguistic conventions. Mill's example of this distinction still applies: the members of the class 'white objects' share no other interesting properties other than whiteness, but generations of research have not exhausted the common properties of natural kinds such as animals and plants (Mill 1891, I, vii, §4). Unlike conventional terms, natural kind concepts cannot be given closed form definitions, but it makes sense to seek to clarify their meaning through empirical inquiry (Griffiths 2004).

From this point of view, there is no reason why historically contingent phenomena could not be natural kinds. Natural kinds are categories about which we can make inductive scientific discoveries. Natural kind terms can be used to make reliable inductive predictions about unobserved instances. Therefore, successful scientific theories are

about natural kinds. As several authors (e.g. Boyd 1999; Griffiths 2004; Sterelny 1990) have pointed out, this view of science should apply to the human sciences as well as to the natural sciences. This is because the natural and human sciences share the same epistemic goals: reliable inductive inference and explanation of phenomena.[2]

A widely accepted, and from the philosophy of science perspective perhaps the most promising, account of natural kinds is Richard Boyd's analysis of natural kinds as homeostatic property clusters (Boyd 1991, 1999). Boyd's theory is quite well known and not in need of a detailed exposition. A brief summary of the main points will do. According to the theory, a natural kind consists of two elements: (1) A cluster of typical properties and (2) a homeostatic causal mechanism responsible for the co-occurrence of the prototypical properties. In short, the instances of a kind are similar, because there is a causal mechanism that governs the co-occurrence of the typical properties of the kind.

Boyd's classic example of a homeostatic cluster kind is a sexually reproducing biological species. The members of the species share a number of morphological, physiological and behavioral properties. These similarities between the members of the kind are explained by the exchange of genetic material: the members of the species produce offspring only with individuals that belong to the same species, not with animals from other species. (Boyd 1991, 142; 1999, 67.)

Assuming for now that natural kinds could have extended homeostatic mechanisms, bulimia nervosa could be approached in a similar manner. The etiology of the illness, the cluster of symptoms, and response patterns to treatments belong to the cluster of properties characteristic of the kind. The homeostatic mechanism is in this case a complex causal mechanism, consisting of neurophysiological, psychological and social elements that together govern the manifestation of the illness.

What makes the HPC-theory a valuable tool for the philosophy of science is that it offers a plausible picture of conceptual change in science. The idea could be captured in the slogan "the kinds are where the mechanisms are." In science, we should put like with like. Things that share a common causal structure (vs. merely observable similarities) should be grouped together in classifications. Scientific concepts are tools for our

---

[2] Bridgandt (2003) and Griffiths (2004) have adopted the new concept 'investigative kind' to replace the concept of natural kind in the context of philosophy of science. This terminological move could help avoid unnecessary proprietary conflicts between different users of the concept natural kind.

inductive and explanatory practices. We should sharpen these tools by revising the contents of our concepts in light of results from empirical research on causal mechanisms. As a result we get the tightest possible fit between our concepts and the structures of reality, and thus make reliable induction and explanation possible.[3]

To summarize: In light of its role in philosophy science, nothing in the concept of natural kind rules out the possibility of treating kinds with extended mechanisms as natural kinds. However, a worry remains. The concept of mechanism might carry a commitment to internality, and as the HPC-theory of natural kinds depends on the concept of mechanism, this would imply that natural kinds need to be based on non-relational properties after all. The literature on the HPC-theory offers little help here: the idea of homeostatic mechanism is usually introduced only by giving examples of mechanisms behind kinds, and thus the meaning of the concept is left vague. Furthermore, typically the exemplary HPC-kinds have internal, not external mechanisms. In the next section, I focus on the concept of homeostatic mechanism to find out whether there is something in the notion that rules out the possibility of extended mechanisms.

3 What is a homeostatic mechanism?

In many fields of contemporary science it is common to think that a genuine explanation of a phenomenon requires providing a description of a mechanism. The concept has acquired an important role also in the philosophy of science (cf. Machamer et.al. 2000; Glennan 2005). Several different conceptions of mechanism have been suggested, but the important contributions agree, that a mechanism consists of (i) parts that are (ii) organized together in order to manifest some (iii) regular behavior. (Machamer et.al. 2000; Glennan 2002, 2005; Woodward 2002). From an explanatory perspective, a mechanism can be characterized as a (sufficiently modular) part of the causal structure of reality that can be used to explain *how* the explanandum phenomenon is brought about.

---

[3] For examples of the use of the HPC-theory as a heuristic for concept formation, see Griffiths (1997), Craver (2004), Machery (2004, 2009, ch. 8), Murphy (2006), Samuels (forthcoming).

In the case of homeostatic mechanisms, the behavior explained by the mechanism is the *reliable co-occurrence of the observable properties* characteristic of the natural kind. To function normally, a homeostatic mechanism must reliably produce its associated property cluster despite changes in the environment parameters, thus sustaining a stable phenomenon. In the case of culture-bound mental illnesses, the mechanism is needed to explain how different individuals with unique life histories and physiological make-ups all end up showing the cluster of properties typical of the illness.

It is often thought that mechanisms have to be somehow internal to the system, whose behavior is being explained. The explanatory role of mechanisms supports this intuition: When mechanistic explanations produce understanding by revealing how the cause brought about the effect, they do this by describing the causal structure *between* the original cause and effect. However, one must be clear what kind of internality the betweenness here implies. The reference to the parts of a mechanism implies no commitments about the spatial ordering of the components of the mechanism, but refers only to the *topology of causal (or constitutive) relations*. A description of a mechanism shows the operations and transformations between the initial and final state. This causal-constitutive internality should not be understood in a too concrete way.

Besides this conceptual confusion, the internality requirement for mechanisms is implied by several explicit strategies for drawing the boundaries of systems and mechanism. In his discussion about these strategies, Carl Craver (2007, 141–144) brings up some points pertinent to the internality issue. As pointed out by Craver, *compartmentality*, equating mechanism boundaries with the boundaries of physical compartments (e.g. nucleus, cell, body), is an obvious strategy of delineation, but it has rarely been defended with systematic arguments. Another popular alternative is to appeal to the *intensity of interactions* between elements, and draw the boundary between the mechanism and its environment at the causal bottleneck, the local minimum of the strength of the causal interactions.

As Craver shows, neither one of these strategies is in itself sufficient for drawing the boundaries of a mechanism, because all delineations have to be based on considerations of causal relevance. A mechanism should describe the parts (and no others) of the causal structure that make a difference to the explanandum phenomenon.[4] As can be seen from the characterization of mechanism at the beginning of this section, a mechanism is

[4] See section 4 for further discussions on explanatory relevance and difference making.

always a mechanism *for* something. Mechanisms can only be identified when the explanandum has been fixed. In order to know what parts of the causal structure to include in a mechanistic explanation, one needs to be clear of what one is trying to explain.

Some might find the conclusion rather disappointing: focusing on mechanisms is not the magic bullet that would solve the difficult questions concerning natural kinds and classification. Mechanisms will not be our guides to THE correct classification of reality. Rather, how we should draw the boundaries of mechanisms seems to depend on what we are trying to explain. However, despite this initial kickback, the mechanistic approach to natural kinds offers a very plausible pluralist view of classification: The causal structure of reality is very complex, and in light of different epistemic aims, different parts of this causal web can be taken into our descriptions of mechanisms. The classifications based on such mechanisms can capture only some aspects of the complex reality. Still, as long as they can be used reliably in inductive inference and explanation, they can be called natural kinds.

This suggests adopting an explanatory approach to the extended mechanisms problem. As was shown above, there are no a priori reasons requiring that homeostatic mechanisms of scientific kinds should be internal to the classified entities. However, neither can the HPC-theory offer direct support to the extended mechanisms view. Rather, the way we draw the boundaries of kinds and mechanisms has to be based on epistemic considerations. I conclude this section by introducing three different strategies for dealing with the problematic extended phenomena. These strategies correspond to three different ways of conceptualizing the kinds and their corresponding mechanisms. I will call these strategies the conservative, extended systems, and extended mechanisms approach. The epistemic advantages and disadvantages of the strategies will be assessed in the next section by appealing to a systematical approach to explanatory power.

(CO) *Conservative strategy.* One can be minimalist about the extended phenomena. The conservative strategy identifies the explananda so that each entity contains its homeostatic mechanism. In the case of bulimia, for example, this would imply separating the physiological core disorder from its "cultural phenotype" and having separate explanatory programs for each.

(ES) *The extended systems approach.* Another alternative is to hold on to the idea that mechanisms have to be internal to the explananda, but redraw the boundaries of systems

so that they include all the variables belonging to the homeostatic mechanism in question. The extended mind thesis is an example of this approach: Because the cognitive mechanisms constituting our mental capacities are often based on intricate feedback loops between our brains, bodies, and artifacts, the mind-system is extended to include also the relevant parts of the external world.

(EM) *The extended mechanisms approach* is similar to the extended systems approach, but more modest. It does not imply major revisions to our scientific ontology, but emphasizes that relevant causal structures behind the many interesting phenomena (e.g. bulimia) extend beyond the system manifesting the phenomenon, and therefore these kinds of phenomena are supported by homeostatic mechanisms transgressing the spatial boundaries of the system.

4 Explanatory power dissected

In this last part of the paper, I will assess the three kind-formation strategies introduced above. The strategies are evaluated based on how powerful theories they will be likely to produce. As argued by Ylikoski & Kuorikoski (2010), the concept of explanatory power[5] is very important in assessing explanations, but its meaning in the literature is vague. Y&K show that our intuitions regarding explanatory power stem from five different dimensions. The dimensions are partly conflicting and trade-offs between dimensions exist. Y&K's framework of explanatory virtues is based on *a contrastive-counterfactual* model of scientific explanation (Woodward 2003). The authors suggest, however, that this theory can also accommodate the judgments of explanatory depth stemming from other competing theories of explanation, and thus be of more general interest.[6] As a similar account of explanation is assumed in this paper, a few words on the contrastive-counterfactual model are in order.

According to the contrastive-counterfactual model, explanation is an epistemic activity. Explanations aim at creating understanding, which in turn is cashed out as the ability to answer questions concerning counterfactual situations.[7] That is, understanding is the

---

[5] The concepts 'explanatory power' and 'explanatory depth' are largely synonymous.
[6] There are several competing theories of scientific explanation on the market: DN-account, unificationist theory, causal process account, and pragmatic account.
[7] Ylikoski (2009) emphasizes the distinction between *understanding* as an ability, and the *sense of understandin*g, the feeling we get when we think we understand something.

ability to answer *what-if-things-had-been-different questions* (*what if*-questions) (Woodward 2003). A somewhat more technical way to approach the idea is to say that explanations track invariant dependencies between values of variables: X explains why Y if Y depends on X in the sense that if X had not happened, then Y would not have happened either.

Explanations also have a contrastive structure. They are typically answers to questions of the form: why fact rather than foil, in which the foil is an exclusive alternative to the fact.[8] An explanation might for example show that blue litmus paper turned red (rather than remained blue) under acidic conditions (rather than under neutral or alkaline conditions).

The contrastive-counterfactual view gives a plausible picture of the explanatory value of describing mechanisms. Describing a mechanism behind a phenomenon creates understanding by revealing how the explanandum phenomenon is dependent on the parts of the mechanism and their organization. It allows one to answer *what if*-questions regarding what would happen to the explanandum, if there were changes to the parts or their organization. In the bulimia case, knowing the complex homeostatic mechanism behind the phenomenon, one could predict how interventions to the physiological, psychological or social variables in the mechanism would make a difference to the manifestation of the illness (Ylikoski 2010).[9]

Extended mechanisms in light of the explanatory virtues

With this picture of the nature and aim of explanations, the concept of explanatory power can be stated more precisely: The degree of understanding conveyed by an explanation can be defined as the number and importance of counterfactual inferences that the explanatory information makes possible (Ylikoski 2009). The amount of inferences that can be drawn from an explanation depends on five separate virtues: non-sensitivity, cognitive salience, precision, factual accuracy, and degree of integration. Out of these five virtues, only factual accuracy (the amount of idealizations in the explanation), seems largely orthogonal to the choice between the conservative and extended strategies. I will now discuss the remaining four virtues.

---

[8] This idea, though somewhat controversial, has earlier been defended by Van Fraassen (1980), Hitchcock (1996), Schaffer (2005).

[9] For the use of the notion of mechanism in the social sciences, see Hedström & Ylikoski (2010).

Non-sensitivity. The sensitivity of an explanation is determined by how easily the explanatory dependency between the explanans and the explanandum variables breaks down as a response to changes in the background conditions (Woodward 2003, ch.6). Non-sensitivity seems like an especially important property of natural kind mechanisms: homeostatic mechanisms, by their nature, should be resistant to changes in the environment values, thus supporting reliable inductive inferences made with natural kind concepts. It seems that ES and EM are superior to the CO strategy in this dimension. I will demonstrate this with the now already familiar bulimia example.

In the bulimia case, three different variations of the conservative strategy are available. (1) The most stringent one would be to take only the psycho-physiological core of the illness as one's explanandum, and to strictly focus on the internal mechanism, in this case the hormonal abnormality. This way of describing the phenomenon would make the explanation more sensitive than the ones produced by extended strategies (ES & EM): changes in the cultural norms relevant to the eating disorder make a difference to the phenomenon, but the truncated physiological mechanism cannot explain how this happens. Instead, the dependencies between the psycho-physiological variables and the occurrence of the illness continue to hold only when relevant environment variables (e.g. cultural norms regarding beauty) are held constant.

(2) Another, more promising conservative strategy would be to again focus on the psycho-physiological phenomenon, but to indirectly refer to the external factors as well. Instead of including the cultural norms as proper parts of the mechanism, one could have the *representations* of cultural norms as inputs to the homeostatic mechanism. It seems that this would be a way to avoid extended mechanisms, but still incorporate the external factors in the explanation.

However, referring to the shared cultural norms themselves rather than their representations on the perceptual surface of an individual seems to increase the understanding produced by the explanation. By including the social mechanism by which the representations are distributed and propagated in the population, one gets a causal account of how culturally shared meanings interact with the individual psychology, and

thus gains the ability to answer a set of new *what if*-questions.[10] Moreover, if there are strong feedback connections between the internal and external parts of the mechanism, drawing the boundary of the mechanism on the perceptual surface of the individual and focusing on the internal part prevents one from understanding the nature of these complex causal connections.

(3) The third possible alternative for the conservative would be to divide the extended mechanism into parts, each corresponding to its own phenomenon. This piecemeal approach seems to be encouraged by the current disciplinary structure of science: each discipline focuses on the aspects of phenomena that are most easily available to its research methods and instruments. This strategy creates detailed explanatory texts, but in order to answer the same amount of *what if*-questions as the extended strategies (ES & EM), additional work is needed for integrating results from different disciplines. The relationships between variables described in different sub-mechanisms must be investigated so that it can be seen how they work together to manifest the extended phenomenon.

The *precision* of an explanation refers to how sharply the explanandum and its contrast classes are specified. For example, suppose a population neurons is given electrical stimulation. An explanation that shows why this causes the neurons to fire at rate f_1 rather than rate f_2 is more precise than an explanation that shows why they fire rather than do not fire at all. In the eating disorder case, including cultural variables in the mechanism improves the precision of the explanation: Based on knowledge of internal physiological mechanisms, one could perhaps tell whether a particular person has bulimia, but including the cultural variables makes it possible to explain differences between different cultural manifestations of the same physiological pathology.

There remain two important virtues, degree of integration and cognitive salience. Unlike sensitivity and precision, they seem to favor the conservative strategy. The *degree of integration* into existing knowledge makes a difference to the power of an explanation,

[10] Sperber's (1996) epidemiology of representations is a naturalist theory of cultural representations that could provide a framework within which to understand the interaction of psychological and sociological factors.

because the amount of *what if*-questions one is able to answer regarding a phenomenon depends on how well its explanation is connected with the rest of one's scientific knowledge. Integration facilitates the creation of inferential connections between phenomena.

By *cognitive salience* Y&K mean the ease of use of an explanation. A complex model with hundreds of variables might be an accurate representation of a system, but still create little understanding because of its impenetrability. A model can be useful for accurately predicting the behavior of a system, but still be non-explanatory, if it cannot be used for making relevant *what if*-inferences. Cognitive salience would seem to favor mechanisms with limited scope: for scientists with training in mainly one discipline, understanding extended mechanisms with physiological, cognitive, psychological, and social components would be a formidable task. The piecemeal strategy, in which scientists from different disciplines focus on different sub-mechanisms, seems a more cognitively ergonomic approach. In addition, the division of labor between scientific disciplines implies that truncated mechanisms (CO) are easier to integrate with the rest of our contemporary scientific knowledge: as explanations tend to mainly refer to the intrinsic properties of systems, and most of the time follow disciplinary boundaries, truncated mechanisms seem to form modules compatible with such an organization of scientific knowledge.

5 Extended systems vs. extended mechanism

As mentioned earlier, there are trade-offs between different dimensions of explanatory power: some dimensions favor truncated mechanisms, and others extended mechanisms and systems. Luckily, the HPC-theory is compatible with *classificatory pluralism*: There need not exist facts of the matter concerning the correct delineation of scientific kinds. Hence, there is no need to find the one correct way to draw the boundaries of natural kinds. Rather, different systems of classification can be used in different research programs that have differing epistemic aims. For certain purposes, it can be best to focus on very narrow low-level kinds, whereas for the purposes of integrative research, extended approaches might be indispensable.

In the above treatment of explanatory virtues, the extended systems and extended mechanisms approach were together contrasted to the conservative strategy. From an explanatory point of view, the choice between ES and EM does not seem to make a big

difference. In the end, the difference between these approaches is largely a matter of ontological radicality. The extended systems approach implies radical revisions to our scientific ontology, whereas the extended mechanisms approach sticks to a rather traditional conception of the explananda of our theories, but maintains that comprehensive explanations of these phenomena require appealing to extended mechanisms. However, considerations of cognitive salience suggest that to preserve explanatory power, radical deviations from the rest of our scientific ontology should be avoided. This appears to speak against the extended systems strategy.

In conclusion, I suggest that adopting the explanatory approach to the topic of kinds and mechanisms could give a fresh perspective on the disputes surrounding the extended mind. As suggested above, from the scientific perspective the considerations concerning explanatory power are more important than ontological disputes on whether the mind really is extended or not. Moreover, the theory of natural kinds presented in this paper suggests that our scientific ontology should follow the results of empirical research and accompanying epistemic considerations: natural kind concepts are tools for reliable induction and explanation. They create an accommodation between our scientific practices and the causal structure of reality. How this accommodation is to be achieved cannot be based on a priori intuitions concerning the scope of the mental.

References

Bird, A. and E. Tobin (2008). "Natural kinds", in E. N. Zalta (ed.), *Standford Encyclopedia of Philosophy*. http://plato.stanford.edu/.

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

—— (1999). "Kinds as the 'Workmanship of Men.'" In Nida-Rümelin, Julian (2000). *Rationalität, Realismus, Revision*. Berlin & New York: Walter de Gruyter.

Brigandt, Ingo (2003), "Species Pluralism Does Not Imply Species Eliminitivism", *Philosophy of Science* 70, (Proceedings), 1305–1316.

Clark, Andy (1997). *Being There*. MIT Press.

—— (2008). *Supersizing the Mind*. New York: Oxford University Press.

Craver, Carl (2004). "Dissociable realization and kind splitting", *Philosophy of Science*, 71, 960–971.

—— (2009). "Mechanisms and natural kinds", *Philosophical Psychology*, 22, 5, 575-596.

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

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

Glennan. S. (2002). "Rethinking mechanistic explanation"*, Philosophy of Science*, 69, 3 (supplement).

—— (2005). "Modeling mechanisms", *Stud. Hist. Phil. Biol. & Biomed. Sci.*, 36, 443–464.

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

—— (2004). "Emotions as natural and normative kinds", *Philosophy of Science*, 71,

Hacking, Ian (1995a). *Rewriting the Soul: Multiple personality and the sciences of memory*. Princeton: Princeton University Press.

—— (1995b). "The looping effects of human kinds." In Sperber & Premack & Premack (1995). *Causal Cognition. A multidisciplinary debate*. Oxford: Clarendon Press.

—— (1998). *Mad Travelers. Reflections on the reality of transient mental illnesses*. Cambridge: Harvard University Press.

Hedström, Peter; Ylikoski, Petri (2010). "Causal mechanisms in the social sciences", *Annual Review of Sociology*, 36, 49–67.

Hitchcock, C. (1996). "The role of contrast in causal explanatory claims", *Synthese*, 107, 395–419.

Kornblith, Hilary (1993). Inductive Inference and Its Natural Ground. Cambrigde & London: MIT Press.Machery, Edouard (2005). "Concepts are not a natural kind", *Philosophy of Science*, 72, 444–467.

Machamer, Peter; Darden, Lindley; Craver, Carl (2000). "Thinking about mechanisms", *Philosophy of Science*, 67, 1–25.

Machery, Edouard (2005). "Concepts are not a natural kind", *Philosophy of Science*, 72, 444–467.

—— (2009). *Doing without concepts*. New York: Oxford University Press.

Mill, John Stuart (2002/1891). *A System of Logic: Ratiocinative and inductive*. Honolulu:

University Press of the Pacific.

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

Samuels, R. (Forthcoming) "Delusions as a Natural Kind" In Broome, M. & Bortolotti, L. (eds.) *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford University Press.

Satz, Debra; Ferejohn, John (1994). "Rational choice and social theory", *Journal of Philosophy*, 91/2, 71–87.

Schaffer, Jonathan (2005). "Contrastive causation", *The Philosophical Review*, 114, 297–328.

Skipper,Robert; Millstein, Roberta (2005). "Thinking about evolutionary mechanisms: natural selection", *Stud. Hist. Phil. Biol. & Biomed. Sci.,* 36, 327–347.

Sperber, Dan (1996). *Explaining Culture: A naturalistic approach*. Oxford: Blackwell.

Steiger, H.; Gauvin, L.; Israel, M.; Koerner, N.; Ying Kin, N.;Paris, J.; Young, S. (2001). "Association of Serotonin and Cortisol Indices With Childhood Abuse in Bulimia Nervosa", *Archives of General Psychiatry*, 58, 837-843.

Sterelny, Kim (1990). *The Representational Theory of Mind*. Oxford: Basil Blackwell.

Strevens, Michael (2000), "The essentialist aspect of naïve theories", *Cognition*, 74, 149–175.

Van Fraassen, B. (1980). *The Scientific Image*. Oxford: Oxford University Press.

Woodward, James (2002). "What is a mechanism? A counterfactual account", *Philosophy of Science*, 69, 366–377.

—— (2003). *Making Things Happen. A theory of causal explanation*. Oxford University Press.

Ylikoski, Petri (2009). "The illusion of depth of understanding in science", in *Scientific Understanding: Philosophical Perspectives* (edited by H. De Regt, S. Leonelli & K. Eigner), Pittsburgh University Press, 100-119.

—— (2010) "Social mechanisms and explanatory relevance", in *From Social Mechanisms to Analytical Sociology* (edited by P. Demeulenaere), Cambridge University Press, forthcoming.

Ylikoski, Petri; Kuorikoski, Jaakko (2010). "Dissecting explanatory power", *Philosophical Studies,* 148/2, 201–219.