The European Philosophy of Science Association Proceedings

Volume 2

Series Editor

Friedrich Stadler, Wien, Austria

For further volumes:
http://www.springer.com/series/8888
Natural Kinds and Concept Eliminativism

Samuli Pöyhönen

Abstract Recently in the philosophy of psychology it has been suggested that several putative phenomena such as emotions, memory, or concepts are not genuine natural kinds and should therefore be eliminated from the vocabulary of scientific psychology. In this paper I examine the perhaps most well known case of scientific eliminativism, Edouard Machery's concept eliminativism. I argue that the split-jump-eliminate scheme of conceptual change underlying Machery's eliminativist proposal assumes a simplistic view of the functioning of scientific concepts. Conceiving of scientific concepts as natural kind terms is an important reason for the impasse between Machery and anti-eliminativists, as both sides allude to properties of natural kinds in their contradicting arguments. As a solution I propose that, in order to develop a more satisfactory theory of conceptual change in science, one needs to distinguish between three different types of scientific concepts, hitherto conflated under the loaded notion of natural kind.

1 Introduction

Eliminativism has a venerable history in the philosophy of mind and the philosophy of psychology, and the arguments for abandoning independent mental substances and properties from our ontological catalogue have played an important role in the development of both philosophical as well as scientific thinking about the mind. In the latter part of the twentieth century, eliminativist arguments were often directed at mental states posited by common sense psychology, and in the 1980s they received broad attention in the debates over eliminative materialism. Recently a new mechanistic variant of eliminativism has emerged. Based on a model of
conceptual change that I call the split-lump-eliminate scheme (SLE scheme), it has
been suggested that familiar notions such as emotion, memory, or concept
do not correspond to genuine natural kinds, and should therefore be eliminated
from scientific vocabulary (Griffiths 1997; Machery 2005, 2009; Piccinini and
Scott 2006). Edouard Machery’s concept eliminativism is perhaps the most hotly
debated example of these recent eliminativist projects. According to Machery’s
heterogeneity hypothesis, the human capacity for conceptual thought is supported
by at least three different kinds of representations and processes, and thus concept
is not a natural kind. However, a large group of philosophers and psychologists
alike have resisted Machery’s eliminativist conclusion (see the peer commentary
on Machery 2010). It has been a common reaction to Machery’s position to argue
that despite the differences between different kinds of concepts, the notion has an
important theoretical role in psychology, and thus cannot be abandoned.

Machery presents a strong case for the claim that concept is not a useful notion
for describing open explananda in psychological research on human conceptual
abilities, i.e., psychological phenomena whose properties are still at least partly
unknown and that are under ongoing inquiry. However, this would be a sufficient
reason for concept eliminativism only if referring to explananda was the only
epistemic function for scientific concepts.1 This is where I part ways with Machery
and side with the defenders of concepts. I contend that functionally identified kinds
sustained by abstractly characterized causal mechanisms often play an epistemically
important explanatory role in the sciences. Hence, if concept turns out to be such
a functional kind, its heterogeneity alone is not a sufficient reason for elimination.
Machery’s eliminativist inference is therefore premature, and his heterogeneity
hypothesis must be qualified for it to be sound.

My diagnosis of the conflict between Machery and the anti-eliminativists about
categories is that the disagreement can be traced to a simplistic picture of the
functioning of scientific concepts, the natural kinds model. To reveal the inadequacy
of this approach shared by both sides of the debate, I introduce the SLE scheme and its
problematic application to concept in Sects. 2 and 3. I then examine the notion
of natural kind underlying the model. I show that, unlike what its proponents suggest,
in most cases the SLE scheme does not provide unambiguous recommendations for
categorical change in science, and fails as a normative foundation for concept
eliminativism. In Sect. 6, I offer my positive contribution. I introduce a distinction
between three types of scientific concepts that hitherto have been conflated under
the notion of natural kind. My threefold division of kinds classifies scientific

1 In this paper, the term ‘concept’ appears in two different meanings. ‘Concept’ in the psychological
sense refers to a putative cognitive structure of an individual. ‘Scientific concepts’, on the other
hand, are things featured in scientific theories and employed collaboratively by scientists in their
research practices. Although there certainly are continuities between these two uses of the term, it
is important to clearly distinguish between them.

2 Roughly, investigatory kinds are vehicles for representing targets of ongoing empirical research, instrument kinds function
as explanantia, and non-mechanistic framework kinds are tools for coordination between different research perspectives on complex targets of scientific inquiry.

2 The Split-Lump-Eliminate Scheme

The SLE scheme underlying Machery’s concept eliminativism builds on the idea
that scientific concepts should refer to natural kinds. I call this approach the
natural kinds model of scientific concepts. The theory of natural kinds employed
by Machery and other scientific eliminativists is an interpretation of Richard
Boyd’s homeostatic property cluster theory (HPC), according to which natural kind
concepts should be aligned with causal mechanisms in reality. According to Boyd’s
(1991, 1999) theory, a natural kind is characterized by

(a) a cluster of typical properties that is supported by
(b) a homeostatic mechanism that brings about their co-occurrence.

The SLE scheme of conceptual change, which builds on this foundation, is based
on three operations: If a concept refers to several different mechanisms, one should
split it so that each mechanism gets its own corresponding concept. On the other
hand, a concept should capture the maximal class of phenomena sustained by the
same mechanism. Therefore, if we can find the same mechanism behind a group of
phenomena that were previously considered as separate, we should lump them
under the same concept. And thirdly, were it to turn out that there is no well-
defined mechanism corresponding to a concept, we should eliminate this notion
from scientific usage. In sum, the core idea underlying these three operations is
that there should be a one-to-one correspondence between scientific concepts and
mechanisms in reality (Griffiths 1997, 2004; Machery 2009; Craver 2009).

In comparison to probably the most well-known account of conceptual change
in the philosophy of psychology, eliminative materialism, the SLE scheme is an
advancement in several respects. First, it offers a more fine-grained picture of
conceptual revision by not only focusing on elimination but by also including cases
of unification and non-eliminative conceptual refinement as species of conceptual
change. Second, the model does not rely on semantic intuitions about reference as a basis for conceptual change, but instead draws on the widely accepted realist judgment that our scientific classifications ought to be aligned with the causal structures of reality. It thus conceives of the conceptual dynamics of scientific psychology as being continuous with those of other scientific fields, whereas the domain of eliminativist materialism is limited to folk psychological predicates only. Moreover, as the work of scientific eliminativists suggests, perhaps the most convincing evidence in favor of the SLE scheme comes from its ability to account for several recent episodes of theoretical development in the human sciences (cf. Griffiths 1997; Craver 2004; Wilson et al. 2007; Machery 2009).

3 Concept Eliminativism and its Discontents

The instance of scientific eliminativism that has recently raised the most debate is Edouard Machery’s concept eliminativism (2005, 2009, 2010). In Doing Without Concepts, Machery defines concepts in psychology in the following way:

A concept of x is a body of knowledge about x that is stored in long-term memory and that is used by default in the processes underlying most, if not all, higher cognitive competences when these processes result in judgments about x (Machery 2009, p. 12).

Based on empirical research in the cognitive sciences, Machery (2009, p. 4) then formulates his heterogeneity hypothesis:

1. The best available evidence suggests that for each category an individual typically has several concepts.
2. Co-referential concepts have very few properties in common. They belong to very heterogeneous kinds of concept.
3. Evidence strongly suggests that prototypes, exemplars, and theories are among these heterogeneous kinds of concept.
4. Prototypes, exemplars, and theories are typically used in distinct cognitive processes.
5. The notion of concept ought to be eliminated from the theoretical vocabulary of psychology.

Machery reviews plenty of empirical evidence for each of the tenets 1-4, and despite dealing with contestable issues, they have raised relatively little controversy. Not so for tenet 5. Machery’s normative conclusion and the main result of his book. It appears that Machery regards the last tenet as an implication of the conjunction of tenets 1-4 together with the principles described above as the SLE scheme. According to Machery, evidence suggests that scientifically interesting generalizations about concepts are actually sustained by mechanisms corresponding to the different subkinds, and because these mechanisms are sufficiently distinct, there is no well-defined mechanism underlying CONCEPT as such. Based on the SLE scheme, Machery then concludes that the notion of concept should be eliminated from scientific psychology, and replaced with lower-level notions referring to prototypes, exemplars, and theory-based concepts.

Perhaps the most common challenge to Machery’s eliminativist conclusion has been to emphasize the indispensable theoretical role that the notion of concept plays in psychological research. Machery’s critics have argued that only CONCEPT captures a set of questions and generalizations that have to do with the human capacity for conceptual thought in general. Abandoning the notion would therefore de-emphasize this set and hinder scientific progress because there would be no notion to integrate results from research on subkinds of concepts (Couchman et al. 2010; Edwards 2010; Hampton 2010). A problem with many of these replies is that while they draw on psychologists’ intuitions about the epistemic role of the notion of concept, they have not often been based on systematic theories of the functioning of scientific kind terms.

However, Richard Samuels and Michael Ferreira (Samuels and Ferreira 2010) have replied to Machery on his own ground. They argue, in contrast to Machery’s claim, that there are good reasons to accept CONCEPT as an HPC natural kind. First, there is a reliably occurring property cluster associated with the kind:

1. Concepts consist in bodies of information, and are
2. Stored in long-term memory,
3. Promiscuous (the same information is employed by several higher cognitive abilities),
4. Internally connected, and
5. Internally coherent.

In his defense of concept pluralism, Daniel Weiskopf (2009) has introduced some further shared properties of concepts:

6. Concepts are sensitive to logical form,
7. They combine productively and systematically, and
8. Are acquired by employing similar cognitive processes.

The argument thus goes that CONCEPT has a proprietary cluster (α) of projectible properties. Secondly, Samuels and Ferreira (2010, p. 222) suggest that this property cluster is sustained by a functionally identifiable causal process: the cognitive mechanism (β) corresponding to the cluster above is closely related to processes behind long-term memory, and their relations to other higher cognitive processes. Together these considerations suggest that CONCEPT would qualify as an HPC natural kind. Henceforth, I call this way of defending concepts the anti-eliminativist position.

Importantly, Samuels and Ferreira largely agree with Machery on the empirical facts about concepts, but deny his normative conclusion. The sticking point appears to be at the correct level of description of our conceptual abilities. To shed light on the disagreement, in the next section I examine the role and motivation of natural kind taxonomies in science.
4 Why Worry About Natural Kinds?

The notion ‘natural kind’ is discussed in several areas of philosophy, and in different contexts it serves slightly different conceptual aims. In the philosophy of language it has played a central role in arguments against descriptivism, and in metaphysics, the concept features in discussions concerning laws of nature, natural necessity, and essentialism (cf. Bird and Tobin 2008). The discussions on natural kinds in the philosophy of science constitute the third, partly independent strand of the tradition of natural kinds (Hacking 1991). Within this epistemology-oriented approach, questions of natural kinds concern primarily scientific concept formation (Reydon 2009). It is a common intuition that only when our concepts correspond to natural kinds, do they succeed in referring to genuine phenomena in reality, and can be reliably employed in the empirical practices of prediction, explanation, and manipulation of these phenomena. Moreover, it is now agreed by many that since the epistemic aims of the human sciences are similar to those of natural sciences, the categories in the human sciences should also conform to the natural kinds model (Sterelny 1990, Chap. 3; Boyd 1991). Thus, from the perspective of the philosophy of science, the reason to worry about the natural kindhood of concepts is primarily in order to maximize epistemic power and reliability.

However, even within the philosophy of science, there are multiple competing conceptions of what being a natural kind amounts to. Often the notion is used without explicating what is exactly meant by something being a natural kind, but a review of the current literature suggests the following list of criteria commonly attributed to natural kinds in the special sciences (cf. Hacking 1991; Boyd 1999; Murphy 2006, Chap. 9; Bird and Tobin 2008; Samuels 2009):

\{NK\}

1. Induction justification: Natural kinds should license inductive inferences.
2. Causal grounding: Natural kind concepts should track the causal structure of reality. The unity of a kind is causal, not conceptual.
3. Non-analyticity: Members of a natural kind share a large number of (logically unrelated) non-trivial properties in addition to the ones that are used to identify the kind.
4. Semantic openness: The semantics of natural kind terms is such that it makes sense to attempt to refine their meaning through empirical inquiry.
5. Lawfulness: Natural kinds are referred to in laws of nature.
6. Essentialism: Natural kinds have essences constituted by their intrinsic properties.
7. Uniqueness: There is a unique best taxonomy of reality in terms of natural kinds that represents nature as it is.

The list is not meant as an exhaustive characterization of the properties of natural kinds – alternative suggestions abound. Moreover, most theories of natural kinds only subscribe to some of the criteria. In fact, I suggest that different combinations of the criteria can be used to isolate somewhat independent dimensions in the meaning of ‘natural kind,’ which I employ in Sect. 6 in distinguishing between different epistemic roles that kind concepts play in scientific research practices.

Assessing HPC theory in the light of \{NK\} reveals its liberal nature. The definition of HPC kinds offered in Sect. 2 suggests that they obviously satisfy criteria 1 and 2: an HPC kind consists of a reliably occurring cluster of projectible properties allowing for reliable extrapolation, and its epistemic reliability follows from the kind being anchored in existing causal structures. However, HPC theory is not committed to the whole group of the remaining criteria: The concept of mechanism has a central role in the theory primarily in order to avoid employing the problematic notion of law of nature. Moreover, several proponents of the theory have emphasized that mechanisms underlying kinds need not consist only of intrinsic properties of kind members (Boyd 1991; Griffiths 1997; Murphy 2006). HPC theory is thus not committed to properties 5 or 6.

Moreover, as a recent argument by Carl Craver (2009) shows, HPC classifications do not satisfy the uniqueness criterion (7). In brief, drawing on recent research on the notion of causal mechanism, Craver observes that while mechanisms are constituted by real causal structures, decisions regarding the correct level of mechanistic description and demarcation of the boundaries of mechanisms require considerations of explanatory relevance. This is because descriptions of mechanisms are explanatory devices – mechanisms are identified in order to explain properties of an explanandum phenomenon. Although “kinds are where the mechanisms are,” mechanism individuation in turn requires prior fixing of explananda.

Hence, mechanisms sustaining HPC kinds can be identified at various levels of abstraction, depending on the epistemic aims of the research perspective. 3 Here HPC theory seems to capture a genuine aspect of scientific practice: examples of classificatory pluralism abound in the life sciences. As Robert Richardson (2008) and Mark Coulth (2009) have observed, “multiply realized” concepts such as EYE or ENZYME capture sufficiently homogeneous units for certain epistemic purposes, whereas from other perspectives they appear as heterogeneous kinds. Such higher-level categories are usually functionally individuated, and their corresponding mechanisms are abstractly specified causal structures. In sum, once the often-unnalyzed notion of mechanism is spelled out properly, it turns out that causally sustained functional kinds qualify as HPC kinds. Therefore, it appears that criteria 3 and 4 are not necessary conditions for HPC kinds: If the abstract mechanism exhausting the unity of the kind (e.g., ENZYME, TURING MACHINE) is already known, it does not appear useful to further examine the internal makeup of the members of the kind in order to learn more about the cluster of kind-properties (~4). Moreover, in cases in which the relationship between the mechanism and the

---

3Boyd (2010) himself recognizes this conventionalist aspect of scientific classification, when he states that there are no kinds that are natural simpliciter, but instead kinds are natural with respect to the inferential architectures of particular disciplinary matrices.
corresponding kind properties is transparent, it is often questionable whether it can meaningfully be said that the kind is characterized by a large group of properties not accounted by the definition (i.e., the mechanism description) of the kind.4

5 Inadequacy of the SLE Scheme

This analysis of the commitments of HPC theory points to a serious shortcoming in the SLE scheme: In situations where our classifications are outright incorrect and do not correspond to any well-defined causal mechanisms (e.g., sublunary objects, phlogiston, phrenology, drapetomania), the model rightly suggests elimination. However, in most cases discussed by scientific eliminativists, this is not the case. Instead, decisions of splitting and lumping are often done between causally sustained classifications at different levels of abstraction. This is also the case with concept eliminativism: both the lower-level categories (prototype, exemplar, and theory-based concepts) as well as the higher-level notion of concept facing elimination can be understood as mechanistic HPC kinds. In such cases, none of the operations of the SLE scheme apply.

It appears that in order to save the eliminativist conclusion, supporters of the SLE scheme must adhere to a stricter notion of natural kind. Perhaps the intuitively most obvious option would be to require that members of a natural kind must share the same internal structure. However, this option is not open in the domain of psychology. Cognitive kinds in general are supported by abstractly characterized mechanisms: they are implemented in plastic neural structures, and therefore implementation-level differences between their instances are unavoidable.

In his response to concept pluralists, Machery (2009, pp. 243–245) has adopted a different strategy. He claims that there has to be further empirically discoverable generalizations to be made about natural kinds, and thus in treating functional kinds as natural, concept pluralists misconceive the nature of natural kinds. Machery thus simply assumes that natural kinds should satisfy criteria 3 and 4. This judgment appears to stem from the fact that heterogeneous categories like CONCEPT are not plausible explananda, as empirical research conducted by using such notions would result in disjunctive theories and explanations. However, this move begs the question against pluralists because, for them, the ability to ground reliable theoretical inferences and explanations -- to function as an explanandum -- is sufficient for CONCEPT's kindhood (cf. Weiskopf 2009). As suggested in the previous section, functionally individuated HPC kinds can serve this epistemic purpose: they are by definition characterized by reliably occurring property clusters sustained by abstract causal mechanisms. Therefore, as long as the generalizations made by employing a

concept concern this property cluster, the implementation-level differences between instances of the kind can safely be ignored. Functionally individuated HPC kinds satisfying only criteria 1 and 2 can thus be treated as inferential tools that “black-box” the non-pertinent implementation-level differences between instances of the kind.

Hence, Machery’s eliminativist conclusion appears to be blocked by a competing account of what natural kindhood amounts to. The competing conceptions emphasize two different but equally important epistemic roles that scientific concepts can play. Whereas eliminativism is driven by the idea that natural kind concepts stand for plausible explananda, an anti-eliminativist can emphasize CONCEPT’s role as an indispensable explanans in psychological theories. In the following penultimate section of my paper, I suggest that these two distinct epistemic roles should be clearly distinguished, and that trying to prove the primacy of either one of them is a misguided effort, only motivated by the monolithic natural kinds model of scientific concepts. As my positive contribution, I suggest a threefold division between different types of scientific concepts.

6 Splitting the Notion of Natural Kind

My reconstruction of the notions of natural kind used by eliminativists and their opponents suggested that an eliminativist needs to adhere to a notion that includes all of the criteria 1–4 as necessary conditions of natural kinds, whereas anti-eliminativists employ a more liberal notion that only clearly satisfies 1 and 2. In this section I suggest that these clusters of criteria can be used to identify two different types of concepts employed in science, each with their corresponding epistemic niche. Furthermore, I suggest that also non-mechanistic concepts often play an important role in research practices. By grouping types of scientific concepts according to their epistemic roles, we get the following threefold classification:

(A) Investigative kinds. Adopting a term from Brigandt (2003) and Griffiths (2004), I call ‘investigative kinds’ the group of scientific concepts that capture many of the intuitions behind traditional conceptions of natural kinds. Treating a concept as an investigative-kind concept means that in addition to justifying inductive inference, members of the kind are assumed to share yet unknown similarities, and thus we can learn more about them by empirically investigating the properties of their instances. For this reason, investigative kind concepts are vehicles for representing targets of ongoing empirical research, and often stand for explananda in scientific theories. Examples of investigative kinds would include elementary particles and neutron stars, but also psychological explananda such as schizophrenia or confirmation bias.

(B) Instrument kinds. Unlike investigative kinds, instrument kind concepts typically function not as explananda but as explanans: they serve as vehicles for explanation and storage of scientific knowledge. As argued above, despite being functionally identified kinds, they can serve in these epistemic roles.

4Peirce (1903) observed this tension already in Mill’s account of kinds: Mill (1891) requires that a small group of properties must not account for the rest of a real kind’s properties but, on the other hand, the aim of scientific research to find law-like relationships between the properties of kinds appears to undermine their independence.
because members of the kind share a robust cluster of projectable properties supported by an abstractly specified causal mechanism. However, instrument kinds are not characterized by the same semantic open-endedness as investigative kinds and are thus poor devices for reductive research: there is no reason to assume that their members share non-trivial properties apart from ones governed by the known homeostatic mechanism of the kind. Instead, the epistemic power of instrument kinds like eye, enzyme, market, or Turing machine stems from their ability to capture general patterns and abstract mechanisms common to several different targets and domains.

(C) Framework kinds. As observed already by Hilary Putnam (1965, p. 379), many central scientific concepts are not defined by their role in a single law or theory, but are law-cluster concepts residing at the intersection of several theories. Putnam’s example was energy, but several cases can be found in the human sciences as well: gene, rationality, information, and representation are examples of important concepts that however have slightly different meanings in different research programs (Griffiths and Stotz 2007; Bermúdez 2005, pp. 9–10). I suggest that despite not being anchored in any specific causal mechanisms, framework kinds often play an important epistemic role. As suggested by Susan Leigh Star in her work on boundary objects, in science we need concepts simultaneously inhabiting several social worlds. They must be malleable enough to adapt to the informational requirements of different disciplines, but still maintain the identity of the target across different sites (Leigh Star and Giessmer 1989). To put the matter in terms of {NK}, framework kinds do not satisfy criteria 1 and 2, but 3 and 4 capture important aspects of their functioning. Open-endedness and indexicality are semantic properties that allow the reference of a concept to be fixed independently of particular descriptions, and framework concepts can thus correspond to targets of research whose mechanisms and best levels of description are still unknown.

This tentative classification of scientific concepts according to their epistemic roles is still coarse, and the details of the proposal need to be worked out. However, the scheme is arguably more useful than the monolithic natural kinds model: it appears that all the three types are manifested in scientific research and correspond to distinct epistemic niches. Moreover, all stand apart from conventional or erroneous classifications. This more refined picture of scientific concepts is also useful for making sense of the debate on concept eliminativism, because it can accommodate both Machery’s and anti-eliminativist insights: Machery convincingly shows that concept does not qualify as an investigative kind, and thus trying to uncover the whole set of projectable properties of concepts would be misguided. On the other hand, in several theoretical contexts in psychology concepts are explananda rather than explanans (Lombrozo 2011). As argued above, for these purposes it suffices that a concept satisfies the requirements for instrument kinds.

Moreover, my scheme suggests a third possibility. Retaining concept as an instrument kind requires that it correspond to a well-defined causal mechanism.

The jury is still out on this question, partly due to the fleeting nature of the notion of mechanism, and partly to inconclusive empirical research. Even conceding the eliminativist the judgment that concept is not a mechanistically grounded kind, the notion could survive as a framework kind coordinating research between several fields investigating higher cognitive abilities (e.g., psychology, social sciences, and AI).

7 Conclusion: The Fate of Concepts

I have argued that not all scientific concepts serve the same epistemic purpose. Working out the consequences of this insight suggests that Machery’s eliminativist conclusion does not follow from his heterogeneity hypothesis. However, the genuine insight of Machery’s position can be saved by qualifying his argument: Heterogeneity of concept does not recommend its elimination but it does show that the notion does not pass as an investigative kind, and hence cannot serve the corresponding epistemic role in scientific research practices. Acknowledging this change in the inferential status of the notion can have the same epistemic benefits for psychology as Machery uses to motivate his eliminativist position (cf. Machery 2009, p. 248): being explicit about concept’s status as an instrumental (or framework) kind should discourage useless primacy debates between different theories of concepts and direct attention towards more relevant questions.

My more general aim in this paper has been to highlight an overlooked form of conceptual change in science. In addition to the operations described by the SLE scheme, conceptual change consists also in often-subtle changes in the inferential potential of concepts. The labels ‘investigative,’ ‘instrumental,’ and ‘framework kind’ correspond to such inferential statuses, and keeping track of how scientific concepts move from one concept-type to another is one way of representing such conceptual change. The trajectory of concept might provide a typical example of the life course of a scientific concept: starting off as an investigative kind, the notion first promotes research on a phenomenon that is considered unitary. However, after the heterogeneity of the processes behind the phenomenon is revealed, the notion might persist as a tool for the storage of higher-level generalizations, or as a more malleable notion coordinating research and communication between different perspectives on the target. During this process of conceptual change, the splitting and lumping operations suggested by the SLE scheme might lead to the emergence of more precise (or more general) mechanistic classifications, but these events need not be accompanied by the elimination of the original kind concepts.

Acknowledgements. An earlier version of this paper was presented at the 14th Congress of Logic, Methodology and Philosophy of Science (CLMPS) at Nancy. I thank audiences both at CLMPS and at EPSA11 for their comments. I also thank Petri Yliokoski and Marion Godman for valuable feedback on earlier drafts.
References


