PhiMiSci
Philosophy and the Mind Sciences

Are concepts a natural kind? On concept eliminativism

Guido Löhr[a] (loehrg@icloud.com)

Abstract
Concept eliminativists argue that we should eliminate the term 'concept' from our vocabulary in psychology because there is no single natural kind that is picked out by it. I argue that the most developed version of concept eliminativism by Edouard Machery depends on the assumption that concepts are *defined* as stable and context-independent bodies of information. It is this assumption that leads Machery to eliminativism and it is an assumption we have reason to reject. Another assumption that leads to the eliminativist conclusion and that we have reason to reject is that the type of content represented in long-term memory is the relevant property based on which we should individuate certain natural kinds in cognitive psychology. Finally, I argue that certain pieces of information are functionally integrated enough to meet the conditions for being a natural kind.

Keywords
Concepts · Concept eliminativism · Concept pluralism · Concept hybridism · Contextualism · Natural kinds

1 Introduction

According to Margolis and Laurence (Margolis & Laurence, 2018, p. 1):

> Concepts are the building blocks of thoughts. Consequently, they are crucial to such psychological processes as categorization, inference, memory, learning, and decision-making. This much is relatively uncontroversial.

If we take Margolis and Laurence's quote as a starting point, it becomes immediately clear why the notion of concept is very dear to philosophers and cognitive scientists alike: The majority of them consider it indispensable to studying thought and language. Unsurprisingly, the proposal to eliminate the term 'concept' from cognitive science (Machery, 2005) has generated much opposition (Machery, 2010;

[a]Eindhoven University of Technology.

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632
 https://philosophymindscience.org ISSN: 2699-0369

Piccinini & Scott, 2006; Rice, 2016; Taylor & Vickers, 2017; Taylor & Vosgerau, 2019; Vicente & Martínez Manrique, 2016; Weiskopf, 2009). It may even be considered a serious challenge to some of philosophers' favorite methods of analyzing or engineering concepts (Blackburn, 1999; Cappelen, 2018; Jackson, 1998).

However, Machery also argues that philosophers and psychologists employ different notions of concept (Machery, 2009). Concepts in philosophy can be characterized by Margolis and Laurence's quote above. The notion of concept common in psychology is less focused on building blocks of propositional attitudes and truth-conditional content. Instead, it aims at explaining categorization and higher cognition (Löhr, 2020; Rey, 2010). Of course, both notions are related and may both be necessary for a full account of cognition. Still, if they do turn out to be different, as Machery argues, then there are also two different kinds of concept eliminativism, one being more revisionary than the other.

The kind of concept eliminativism defended by Machery (2009) is a philosophically modest one. He argues that the term 'concept' in cognitive science picks out different kinds of stable bodies of information that are too heterogeneous to be considered a single natural kind. Assuming that scientists intend to discover natural kinds, Machery concludes that we should abandon the term 'concept' and instead use more fine-grained notions that *do* denote natural kinds. This modest kind of concept eliminativism will be discussed here. A more radical version that aims to eliminate the notion of concept as constituents of propositional attitudes will not be addressed. As far as I know, Machery is skeptical of this notion as well. However, he does not explicitly argue for its elimination. [1]

The question of whether we should be concept eliminativists, hybridists, or pluralists of concepts in psychology has been critically discussed in various places (Piccinini & Scott, 2006; Rice, 2016; Samuels & Ferreira, 2010; Vicente & Martínez Manrique, 2016; Weiskopf, 2009). In this paper, I focus on the eliminativist argument in more detail. I am interested in the question of whether eliminativists have given us sufficient reason to accept that concepts in cognitive science are not a natural kind. To evaluate this *psychological concept eliminativism*, I focus on its most developed version by Machery (2009). However, my argument will have consequences for pluralism and hybridism as well. [2] I reconstruct Machery's argument in the following way:

> (1) 'Concept' in psychology refers to the "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

[1] But see Machery, 2017 for a discussion of the more philosophical notion.

[2] Note that the question of whether the term 'concept' in cognitive science picks out a natural kind or not (whether it should be eliminated) is different from the question of whether the term picks out a single hybrid kind or several kinds. We can argue that 'concept' picks out a hybrid and still recommend its elimination based on the view that this hybrid is too heterogenous to fulfill the conditions for being a natural kind.

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

> when these processes result in judgments about x." Machery (2009, p. 12) (see section 2.1)
>
> (2) The description labeled 'concept' picks out three fundamentally distinct types of bodies of knowledge: prototypes, exemplars, and theories. (see section 2.2)
>
> –These types of bodies of knowledge are individuated by their *type of content* and *type of computation*. Prototypes store information about typical features, while exemplars store information about exemplars. Theories store information about causal features.
>
> (3) Natural kinds are individuated by means of sets of scientifically relevant properties. In the psychology of concepts, the relevant properties are whatever explains higher cognitive competencies, in particular, what I call 'regularity effects' (typicality effects, exemplar effects, and causal effects). (see section 2.3)
>
> (4) What explains regularity effects are the type of content of a body of information and the computational process that calculates similarity to this content. *Type of content* and *type of computation* are therefore scientifically relevant properties. (see section 2.3)
>
> (5) If (2–4), then prototypes, exemplars, and theories are each a natural kind. (see section 2.4)
>
> (6) If (1,5), then the term 'concept' picks out three different natural kinds. (see section 2.4)
>
> (7) If (6), then this is confusing psychologists. Therefore, we should eliminate the term 'concept' and replace it with more specific terminology like prototype, exemplar, and theory. (see section 2.4)

I argue that we have reason to reject the first premise (section 3.1). If we reject (1), then 'concept' in cognitive science can in theory pick out more than bodies of information retrieved by default. I then argue that *type of content* of representations is not the critical property that we should be looking for when looking for a natural kind in cognitive science. Instead, we should focus on the way information in long-term memory is functionally integrated, which determines how and when it is retrieved (section 3.2). Prototypes, exemplars, and theories do not however differ in terms of the way their stored pieces of information are functionally integrated, at least not in the relevant sense that takes deliberation into account (section 3.3). I conclude that we have no reason to be eliminativists at this point (section 3.4).

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

2 The eliminativist argument

2.1 Machery's concept of concept in psychology

To understand Machery's proposal to eliminate the notion of concept from psychology, it is crucial to first understand what exactly he takes the relevant notion of concept to be. According to Machery (2009, p. 12):

> A concept of x is a body of knowledge about x that is stored in long-term memory and that is used *by default* [immediately and in a context-independent manner] in the processes underlying most, if not all, higher cognitive competences when these processes result in judgments about x. (My emphasis).

Note that Machery, in this quote, is proposing a *description* of how psychologists in fact use the term 'concept'. He does not yet propose a *theory* of what concepts are or of what is picked out by his description. Machery argues that there is no single natural kind that is picked out by the description. Instead, he argues that there are at least three distinct kinds picked out by it – prototypes, exemplars, and theories (see below).

2.2 'Concept' picks out three distinct types of bodies of information

Machery (2009) argues that the notion of concept that he takes to be central in psychology ("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") picks out different kinds of representations. In other words, he argues that there are three or more distinct types of bodies of information retrieved by default. Machery (in chapter 4 of *Doing Without Concepts*) identifies these as what psychologists have called 'prototypes', 'exemplars' and 'theories'. Following Machery, they are individuated primarily in terms of their *type of content*.

According to Machery (2009, p. 84):

> [P]rototypes consist of knowledge about properties that objects either possess or do not possess or about properties that objects possess to some degree. The property having wings is an instance of the first type of property. The property being sweet is an instance of the second type of property: a substance can be more or less sweet. [3]

[3] He adds: "According to some theories, prototypes merely store some knowledge about which properties are typical (or cue-valid). According to other theories, prototypes also store the degree of typicality (or of cue-validity) of the typical (or cue-valid) properties." (p. 84), but this technicality need not concern us here.

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

Furthermore, according to Machery, these bodies of information are used in cognitive processes that compute the similarity between the properties "summarized" in the prototype and the properties detected in the object in question. [4] He takes the similarity measurement in the case of prototypes to be *linear*. [5]

According to Machery (2009, p. 90):

> [A]n exemplar is a body of knowledge about the properties believed to be possessed by a particular member of a class. When we categorize, when we reason, and so on, we have by default in mind a set of exemplars (or, in some models, an exemplar drawn from a set of exemplars stored in long-term memory).

With respect to their use in higher cognitive processes, Machery (2009, p. 96) argues that

> Exemplar-based models assume that cognitive processes involve the computation of the similarity between exemplars and other representations. (…) The second central property of exemplar-based models of cognitive processes is that the similarity measure is usually supposed to be non-linear. [6]

Finally, Machery takes theories to be sets of beliefs retrieved by default *about nomological, causal, functional, or generic properties associated with a category*. In Machery's (2009, p. 101) words:

> Thus, a theoretical concept is supposed to store some nomological, causal, functional, and/or generic knowledge about the members of its extension. For instance, a theoretical concept of dog stores some nomological, causal, functional, and/or generic knowledge about dogs.

According to Machery, theories are employed in cognitive processes that are different from similarity measures and more akin to causal reasoning. [7]

[4] He writes: "Their most important property is that cognitive processes are assumed to involve the computation of the similarity between prototypes and other representations. (…) The second property of these models is that the similarity computation is usually assumed to be linear." (Machery, 2009, p. 90)

[5] This means that the more representations two sets of representations have in common, the more similar are they judged as.

[6] Machery (2009, p. 98) summarizes this feature of exemplar models well: "an object that is extremely similar to a specific known category member, but only moderately similar to others is more likely to be categorized as a category member than an object that is moderately similar to most known category members."

[7] Machery (2009, p. 106) summarizes an example from Murphy & Medin (1985): "If at a party, a guest jumps in the swimming pool with her clothes on, we may conclude that she is drunk. This categorization judgment does not result from matching the concept of drunken people with a representation of this guest. On the contrary, we infer that the most plausible explanation of the behavior of this guest is that she is drunk."

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences*, *4*, 16. https://doi.org/10.33735/phimisci.2023.9632

At this point, I want to emphasize that Machery does not just argue that we have intentional states with the aforementioned three kinds of content or that there may be different kinds of processes that they are used in. Few people would deny, for example, that we represent certain properties as typical or essential for a category, or that we represent exemplars of a category. The crucial point is that Machery argues that there is evidence that they are *default* bodies of knowledge. Other bodies of information do not belong to the respective concept but to what Machery calls 'background knowledge'. [8]

2.3 A natural kind in science is individuated in terms of scientifically relevant properties

Based on Richard Boyd's (1991) criteria for being a natural or scientific kind, Machery (2005, p. 232) argues that bodies of information retrieved by default should only be considered a natural kind if they meet the following set of conditions:

> A class C of entities is a natural kind if and only if there is a large set of scientifically relevant properties such that C is the maximal class whose members tend to share these properties because of some causal mechanism. [9]

What *could* be relevant (to the psychologist) properties that bodies of information must have in common to be considered a natural kind? Generally, cognitive psychologists study bodies of information based on a number of dimensions:[10]

a) their content (what their constituents represent, e.g., whether they represent a feature as typical or necessary for category membership)

b) the cognitive processes in which they are used (e.g., whether they are used for linear or non-linear similarity measurements)

c) their functional integration (roughly, how the pieces of information are stored or connected in long-term memory)

[8] Machery agrees that there is no reason to stop at positing just three fundamentally different kinds of bodies of information individuated in terms of what kind of information they store. He does not elaborate on what other fundamental sets of beliefs there could be, but I take it that the following two may be candidates: For example, there has been some evidence for beliefs about situations associated with a category (Wiemer-Hastings & Xu, 2005), as well as beliefs about feelings associated with a category (Kousta et al., 2011). Both types of representations may form their own sets of default beliefs that function to aid categorization. Another possibility could be that image-like modal representations could form its own natural kind (but this is certainly not Machery's view).

[9] Machery (2009) says nothing about the second requirement for being a natural kind (that these bodies of knowledge possess shared properties because of some causal mechanism).

[10] For this list, I follow loosely Machery (2009, p. 77).

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

d) their function (what competencies they explain, e.g., categorization, linguistic understanding, analogy making, etc.)

e) how they are acquired (e.g., whether they are learned or not)

f) their format (e.g., whether this content is represented in an image-like perceptual or language-like amodal format)

g) their neural implementation

The crucial step in Machery's argument is that the dimensions (a) and (b), i.e., content and processes are the critical ones when it comes to explaining higher cognitive competencies, in particular categorization (see chapter 4 in *Doing Without Concepts*). The main reason he takes this to be the case is that psychologists rely on them to explain categorization. The key idea is that we store information about the world based on previous experiences. When we see a new object, we compare the properties of this object with this set of information, i.e., the information about properties in the world we cluster together.

A major part of explaining categorization, according to Machery, is to explain what generates three types of what I call 'regularity effects'. Those are the effects found when studying the knowledge psychologists believe we rely on when comparing our memory to our perceptual input in classification processes. It has been argued that these regularity effects show that we do not normally seem to carve up the world by means of necessary and sufficient conditions.

One kind of regularity effect is that representations of typical features of a category can have a processing advantage ('typicality effect'). For example, animals that have many typical features of the category 'bird' are recognized much more quickly as birds than atypical exemplars of the same kind of animal (Murphy, 2004). It is generally thought that prototype theory and its posited prototypes are especially good at explaining typicality effects.

A second type of effect, what I call the 'exemplar effect', is roughly based on the observation that we often rely on remembered exemplars of a category (say, Looney Tunes' Tweety for BIRD) – rather than typical features – to judge category membership. Exemplars were posited by exemplar theorists to explain especially such exemplar effects.

Third, what I call the 'causal effect' is based on the observation that people tend to judge category membership not just by means of its superficial features but also by more theory-like causal or essentialist properties, e.g., that birds are not mammals or that the atomic number of gold is 79 (Machery, 2009; Murphy, 2004; Prinz, 2002). Theories are posited by theory theorists to explain these causal effects.

2.4 The eliminativist conclusion

The eliminativist conclusion is now easy to derive. Machery argues that the term 'concept' in psychology picks out three types of body of information retrieved by

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences*, *4*, 16. https://doi.org/10.33735/phimisci.2023.9632

default (i.e., immediately and in a context-insensitive way): prototypes, exemplars, and theories. These bodies are essentially distinguished in terms of their different types of content and cognitive processing. Machery argues that they do *not* differ with respect to (c) their functional organization[11], (d) their function[12] , and (f) their format[13]. Machery does not say much about potential differences concerning (e) how they are acquired or (g) their neural implementation.[14]

Why do (a) *type of content* and (b) *type of cognitive processing* matter the most? According to Machery, these are the critical features of bodies of information that psychologists use to explain regularity effects. Again, prototype theory was developed to explain typicality effects. Exemplars are posited to explain exemplar effects and theories are posited to explain causal effects. Since *type of content* and *type of cognitive processes* explain regularity effects, they are the ones that are scientifically relevant for individuating natural kinds in cognitive psychology. So, 'concept' picks out not one but three different natural kinds – prototypes, exemplars, and theories.

Finally, if 'concept' picks out three distinct natural kinds, this is confusing psychologists. It has motivated them to pointlessly debate what 'concept' properly refers to – whether concepts are prototypes, exemplars, or theories. We should instead acknowledge that they all are natural kinds and abandon the term 'concept' and replace it with 'prototype', 'exemplar' and 'theory'.

Note that while the purpose of the remainder of this paper is to reject premises 2 and 4 of this argument, I agree with Machery that the question of whether concepts are prototypes, theories, or exemplars is pointless. However, we can reach this attractive conclusion without embracing eliminativism. As we will see in the next section, while prototypes, exemplars, and theories are individuated in terms of the type of content they store, we should not do the same with the notion of concept in psychology.

3 The eliminativist argument debunked

3.1 Machery's concept of concept (against premise 1)

Significant weight in Machery's argument rests on the assumption that the term 'concept' picks out only those bodies of information that are retrieved by default.

[11] According to Machery, the different pieces of information that make up a single default body of information are coordinated in the sense that the retrieval of one is accompanied by the retrieval of the other.

[12] Machery (2009) takes all fundamental kinds of information to fulfill a number of different cognitive competencies, which is an essential part of his "heterogeneity hypothesis".

[13] Machery (2016) is very explicit that he endorses a monist amodal view of concepts.

[14] However, there is no reason to think that Machery takes the different pieces of knowledge retrieved by default to differ fundamentally based on these last two dimensions. He also does not seem to think that they explain regularity effects and are therefore not the relevant property we should care about.

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

What exactly are Machery's reasons for restricting his *notion* of concept in cognitive science to only those bodies of information that are retrieved in an immediate and context-independent manner? This question is interesting because if the term 'concept' is extensionally allowed to pick out more context-dependent structures, we might find a single natural kind after all.

The reason why Machery includes the restriction is that (i) he takes context-insensitivity to be required to explain the observed stability or regularity of our behaviors (Machery, 2009, p. 24) and (ii), more simply, because he takes this notion to be the one used by psychologists (Machery, 2005, pp. 196–197). Both arguments can be rejected as reasons for restricting the concept of concept in psychology in the way Machery suggests.

That concepts have to be stable in the sense of being context-independent is not new (Fodor, 1998; Rey, 2010). Many philosophers of concepts assume that concepts – as compositional constituents of propositional attitudes – have to be context-independent to account for the fact that we can successfully reason and communicate. If the concept of table in the premise "all tables are made of wood" differed from the concept of table in the premise "x is a table", we could not be able to rationally infer that x is made of wood.[15]

However, if, as Machery says, the term 'concept' in cognitive science does not pick out the kind of entity that *constitutes* our premises or beliefs and instead denotes bodies of beliefs (not their constituents) that underly categorization, then 'concepts' not need to be as strictly stable as required by the logician or philosopher of language. Note that this can also be agreed upon by those philosophers who argue that philosophers and psychologists refer to the same kind of thing when using the term 'concept' (Prinz, 2002). For example, the bodies of information that underlie categorization may be highly unstable even if they still have the same "intentional content" (Prinz, 2002), or "r-content" (Del Pinal, 2015). This is how Prinz and Del Pinal can defend the view that the same concept may be instantiated or realized by different bodies of beliefs or belief-like bodies of information.

Thus, there is no *conceptual* reason why we could not explain regularity effects by means of context-dependent structures. Hence, stability should not feature in the definition of the notion of concept in cognitive science. Some have argued that contextualists can even better account for regularity effects (Casasanto & Lupyan, 2015; Kiefer, 2018; Lalumera, 2010; Löhr, 2017). Context-dependent structures can account for regularity effects (typicality, frequency, and exemplar effects) assuming that some pieces of knowledge in some contexts may usually be more available than others. For example, the belief that dogs bark may be reliably available in most if not all contexts. All that so-called contextualists of categorization are committed to deny is that most sets of information underlying higher cognition

[15]Note that these strong requirements of context-independence of concepts – as constituents of truth-conditional thoughts – may not always be met by ordinary minds and speakers. We may be able to converse in ways that facilitate coordination in joint action without necessarily always having identical concept types.

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

are stable and retrieved by default *come what may as a set*, i.e., whether or not this body of information is largely inappropriate for the respective sentence or discourse context.

Second, it is not clear whether Machery's description of the use of 'concept' in cognitive science is accurate. Many leading contemporary psychologists working on categorization like (Barsalou, 1999; Kiefer, 2018; Pulvermüller, 2013; Rosch, 2011) now prefer a more contextualist notion of concept, according to which the body of knowledge underlying higher cognitive competencies may change from situation to situation. Even Rosch's not-so-recent writings (Rosch, 1999, p. 69) suggest that the field has moved on to a more flexible notion, arguing that "context effects show that category prototypes and graded structure are not pre-stored as such in the mind, but rather are created anew each time 'on the fly' from more basic features or other mental structures."

Note that the reader need not accept contextualism to reject this premise. What I wanted to emphasize is that there is neither a conceptual nor an empirical reason to *presuppose* invariantism in this discussion. It should not be presupposed (if you agree with the reasons given above) and instead, we should leave it open to empirical investigation whether categorization for example is best explained by stable bodies of information retrieved by default or whether it is best explained by context-dependent bodies. The notion of concept in psychology should instead be defined as "bodies of beliefs or information underlying higher cognition".

3.2 The relevant property is not content (against premise 4)

The second crucial premise in Machery's argument in favor of eliminativism is that the relevant property that we should look for when searching for natural kinds in cognitive psychology is primarily the type of content of bodies of information. This is because *type of content*, along with *type of cognitive processes*, according to Machery, explain what we are interested in: categorization and other higher-order cognitive abilities like reasoning or decision-making.

The problem is that *type of content* and *type of computations* do not play the role of explaining our higher cognitive abilities that Machery thinks they play. Let's focus again on regularity effects. Believing that dogs *typically* bark (prototypes), or that Rex is a dog (exemplars), does not explain why we recognize typical dogs faster than atypical dogs or why we recognize certain dogs via exemplars. What explains why we immediately (or after deliberation) recognize Rex as a dog is that the information that barking things are usually dogs or that Fido looks like Rex (or that Rex has a certain genetic profile) is readily available in the relevant context.[16] Even if we stored (somewhere in our background knowledge) the false informa-

[16]"Being readily available in the relevant context" does not imply the truth of contextualism. Invariantists also assume of course that the relevant information is available in the relevant context. If you are an invariantist like Machery, you think that only part of the knowledge we *normally* think is part of a single concept is readily available in the relevant context. Thus, in most contexts,

tion that things that bark are cats or that Rex is a cat, we would still display the same behavior if these pieces of information were suppressed in the right context. What matters to explain categorization is how our information is functionally integrated (Vicente & Martínez Manrique, 2016) in long-term memory such that certain pieces of information become available at the right moments while others are suppressed. Without the belief that dogs bark, we would probably have a hard time recognizing Rex as a dog, but we should not conclude from this that the content of the information that dogs bark is what makes this piece of information special or interesting. What is important is how this content is accessed and organized in long-term memory such that it is available when we need it.

Consider for example Machery's (2005, p. 89) example of a prototype model of apple. This model contains several pieces of information about the diagnostic properties of apples (properties that easily set them apart from other properties, say *red* and *round*) including some information about how diagnostic this information is taken to be (largely based on typicality ratings). We can then say that one of the pieces of information stored in this prototype has the complex *content* that apples are *typically* red and round. This kind of content distinguishes prototypes from, say, exemplars, which store beliefs with a different kind of content, say, the content *that the piece of fruit I ate in the morning is an apple* (this might be an exemplar-like piece of information) or *that apples have seeds because this is how apple trees reproduce* (this would be a theory-like piece of information).

Intuitively, this model of our information of apples makes sense. One might think that when looking at an object, I tell you it is probably an apple because apples typically have this shape and color. This is a reliable reason for me to classify this object as an apple. It is the content of my belief that justifies my classification of the belief *that apples typically look that way*. However, possessing a body of information about whether apples *typically* or *necessarily* have a certain shape helps me classify things as apples only if this information is available to me in the right context. Again, what explains categorizing something as an apple is that *apples are red* is more readily available than *apples can be black and rotten* in this context. And it is this difference in availability in typical apple contexts that is represented by cue validity ratings or typicality ratings and not the content of a piece of information.

Again, I might even have the wrong modal information about apples and still make the right rough and ready classification in typical contexts (Armstrong et al., 1983). I might believe that an apple *necessarily* has a certain shape or that it is never yellow. If both are suppressed at the right moment, I might still display the appropriate behavior. Again, what is important and what is reflected in the cue validity ratings is not the content of our bodies of information but their level of accessibility in different contexts. In other words, what explains regularity effects is how long-term memory is organized or how pieces of information are functionally

the Machery-style invariantist and the contextualists will make the same predictions regarding to which information is available in which context.

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

related to one another or integrated in long-term memory (in the list of dimensions above this was dimension c).

This organization can be studied by employing familiar psychological experiments, such as typicality ratings. Whether we represent a feature as typical or necessary is important when it comes to justifying our categorizations (e.g., when asked why we think that it is a bird). It is also important to explain *why* the mind is organized such that it is information about typical features of birds, rather than atypical features for example, that are more readily available. However, when it comes to *explaining* regularity effects, *accessibility* to this content is what counts. An answer to the question of why information p becomes available rather than information q likely involves the structure of the world – because birds and not dogs have wings and you asked me about birds and not dogs.

That regularity effects are not explained by the type of content of the information stored, but by which bodies of information come to mind most quickly (to put it simply) may come as a surprise because psychologists often frame their theories, like prototype theory, in terms of *type of content*. For example, what distinguishes prototypes from exemplars is that the former represents information about typical properties and the latter information about exemplars. This is not problematic. Prototype theories and exemplar theories *do* differ in terms of their ideas about what content is stored and what processes this stored content is used in.[17] Machery is right about that.

But the question here is not how to individuate different kinds of bodies of information in long-term memory and not even how to individuate concepts in psychology. The question is how to individuate *natural kinds* in cognitive psychology. The minimal common denominator is that they should be individuated in terms of what explains typicality, exemplar, and causal effects. Again, by ruling out a contextualist notion of concepts from the beginning (by definition), Machery neglects the possibility that the prototypes, exemplars, and theories of birds may all have at least one thing in common besides being about the same kind: they are made most available when we come into causal contact with this kind even if this availability is context-sensitive and not default-like.[18]

[17] And it is of course interesting to ask why information about features of a category we take to be typical come to mind more quickly in some situations and why information about more theoretical features of the same category may be more available in others. The answer may simply be that features we take to be typical are, in fact, typical of the respective class of objects and that typical features are highly diagnostic and therefore useful for everyday categorizations. The most interesting fact however remains thereby unexplained: why and how does the system adapt to contexts by prioritizing certain types of contents in different situations?

[18] Note that my proposed argumentative strategy to look for dimensions other than (a) and (b) to judge whether we should be eliminativists or not is not new. It has been suggested that a radical form of so-called "concept empiricism" or "neo empiricism" (as Machery calls it), i.e., the view that all or most concepts are perceptional symbols (Barsalou, 1999; Prinz, 2002; Pulvermüller, 2013) may, in principle, "forestall concept eliminativism" (Prinz, 2002). Prinz suggests that all bodies of knowledge underlying higher cognition could have something in common after all: all are perceptual symbols (dimension f). However, even though Machery (2009) devotes much

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

3.3 Higher cognitive competencies and deliberation (against premise 3)

Higher cognition involves more than rough and ready quick categorization and concepts should explain this. What is just as important is this: Explain judgments about cats, dogs, or apples *when deliberating*. What is just as important is explaining why we believe that a typical dog (e.g., a bulldog), an atypical exemplar of a dog (e.g., a Komondor), or a burned dog with only dog DNA, are all of the same kind – dogs. What is just as important is to explain why the different pieces of information are intuitively and inferentially connected and coordinated. Again, what explains this is not the type of content or computational use of bodies of information. What explains this is how the information we have is integrated in memory – how accessible it is in different contexts and how it is inferentially related to other pieces of information.

Machery denies that different fundamental types of information in long-term memory (prototypes, exemplars, theories) are usually "coordinated" (chapter 3 of *Doing Without Concepts*) based on the observation that we can hold inconsistent beliefs about the same category.[19] For example, we often hold to be true that tomatoes are vegetables in one context – relying on prototypes and exemplars – and that they are fruit in another context – where theory-like information is more readily available.

Machery is certainly right to point out that the conceptual system is far from perfectly consistent. We often hold contradictory beliefs. However, Machery tends to downplay the fact that we often cease to hold such contradictory beliefs *after deliberation*. Again, psychologists have to explain not just *fast* categorization but categorization *after deliberation*. Moreover, Machery tends to downplay the important fact that there is actually a tremendous amount of coordination and consistency. I currently classify gold and water into fundamentally different categories. This could be wrong and both categories may be co-referential. However, I currently have no reason to hold this belief as gold and water present themselves in fundamentally different ways (they have very different *modes of presentation*). It would be irrational to merge the two categories without further evidence. Thus, it seems that my beliefs about gold and water are highly coordinated

space attempting to refute concept empiricism in order to support his view that 'concept' picks out fundamentally different kinds of representations, recall that he, too, argues for a monism of format on many occasions. In particular, Machery (2016) argues that the bodies of knowledge underlying higher cognition are represented in a purely amodal format. Moreover, Prinz' (2010) attempt to avoid Machery's eliminativism has, despite its perhaps somewhat misleading title, little to do with finding a common format (he focuses on organization of long-term memory as I do in this paper) and Machery's eliminativist argument is not committed to any theory of format.

[19] "These parts are coordinated: the parts of a given concept do not produce inconsistent outcomes, for instance, inconsistent categorization judgments." Coordination is not the same as linkage: "…when one of the parts is used, say, to categorize, we can *ipso facto* use the other parts of the concept for other purposes." (Machery, 2009, p. 64)

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

This familiar Fregean insight must be reflected in a theory of how my knowledge about the world is organized in my mind. It justifies the assumption that the bodies of information we rely on to recognize objects as *gold* and *water* are not functionally connected in the mind (at least not in the relevant sense that explains how we categorize both) and that, unlike our different contradictory assumptions about tomatoes, they do therefore *not* belong to the same concept. Similarly, even if we retrieve different information about tomatoes by default, this does not mean that we have two distinct concepts of *tomatoes.* The reason is simply that we are aware that tomatoes are not really vegetables. Once we give up on the assumption that concepts have to be default structures in the mind, we can acknowledge that the beliefs that tomatoes are fruit and that they are vegetables are linked in a way that, when deliberating, we realize that one of those pieces of information is false. This does not mean that both belong to separate concepts of tomato (*concept* in the psychological sense).

We can take the idea of deliberation as an additional factor in individuating concepts further. At this point, it is perhaps important to stress that I agree with Machery's key observation that different parts of a concept are *sometimes* not coordinated and, therefore, produce inconsistent outcomes. But once we lose the definitional requirement of concepts as bodies of information retrieved by default, we can think of more creative ways of individuating concepts in psychology. We might think of deliberation as a long process that could even include collective deliberation – a kind of process of conceptual engineering or revision (Cappelen, 2018; Machery, 2017) to increase coordination between different parts of a concept.

For example, we can think of tomatoes as being (botanically speaking) fruit and also as located in the vegetable section. Once we think about it a bit, we do not need to hold the belief that tomatoes are vegetables *and* fruit. We can just think of tomatoes as fruit, which, for historical or practical reasons, are placed in the vegetable section of supermarkets. We improve our concept of tomato by decreasing inconsistency and we do this in a way that reflects the fact that the different pieces of information are connected in a way that the pieces of information associated with gold and water are not.

3.4 Why concepts might still be a natural kind (against the eliminativist conclusion)

It is one thing to consider concepts hybrids of prototypes, exemplars, and theories. It is another thing to argue that they are natural kinds. To return to the question of whether concepts are a natural kind, recall Machery's (Machery, 2009, p. 241) adaption of Boyd's three conditions for natural kindness.

1. There is a large set of properties that these bodies of knowledge tend to possess.

2. These bodies of knowledge possess these properties because of some causal mechanism.

3. This set of properties is specific to this class of bodies of knowledge.

I take it that there is currently no reason to think that these conditions are not met in the case of the concept of concept in cognitive science. A concept in cognitive science is a body of information that underlies higher cognition and that explains how we apply a category x to an object, for example. This body of information has to have *at least* one central property to be considered the concept of x (there might be others): It has to be functionally integrated in the right way.

Long-term memory is apparently functionally organized such that it prioritizes different bodies of information in different contexts, but it is also organized such that certain bodies of information that may seem disconnected under time pressure (by default), actually turn out to be inferentially related, at least when deliberating. Even if they are not yet strongly related, we will most likely relate them in a *consistent* way when engaging in conceptual deliberation – in a way in which we would never connect our ideas about gold and water. And this simple insight, I take it, justifies that our prototypes, exemplars, and theories of x might all belong to the same concept of x and that our concept of x is a scientifically interesting kind worth positing and studying.[20] At least, Machery has given us no reason to reject this more conservative assumption.

Deliberation merely generates a context that makes all the different pieces of information that we take to belong to the same category relevant. This results in the realization that these pieces of information are connected in our minds (even if very loosely), i.e., that they are part of the same concept. Thus, it is not unreasonable to hypothesize that the same mechanism that explains why some information about the same category is prioritized in one context and suppressed in another also explains why these pieces of information are also connected at least when deliberating. Again, at least Machery has not given us reason to think otherwise. Put differently, certain bodies of information we take to be associated with the same category share a property that is *at least one crucial* determiner of both categorization under time pressure *and* after deliberation. This feature is the kind of coordination that hybrid theorists like Vicente & Martínez Manrique (2016) have been insisting on and called "functional integration".[21]

[20] Thanks to one of the reviewers for asking me to make the following clear: I do not here claim that concepts are really a natural kind, but that Machery has not given us a good reason for abandoning the idea that there might be something like a scientifically interesting kind worth studying. This is an incredibly conservative and modest claim but is enough I think to make us question whether Machery's (incredibly thought-provoking) proposal is to be adopted by the scientific community.

[21] Michel (2022) has suggested a more detailed model of such a hybrid in a predictive processing framework. The core idea is that the different types of knowledge are not "parts" of a hybrid but arise from different ways of processing a unified structure in which prediction error minimization

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

Functional integration is often suppressed under time pressure, but we know that it is there because when deliberating (when we put ourselves in the right context), we realize that a conceptual connection exists. When engaging in conceptual deliberation, this connection might be discovered as the most plausible (given our background assumptions). The mechanism that explains how the different pieces get to be connected and how they get suppressed depending on context is likely the same mechanism that explains why the connection can be intuited if it becomes relevant, e.g., when deliberating. Furthermore, such connected bodies of information can be contrasted clearly with categories that we consider completely disconnected, as we saw with the difference between water and gold.[22]

Finally, the functional integration is specific not to a sub-kind of stable bodies of information, i.e., it is not specific to prototypes, exemplars, or theories respectively. Instead, it applies to the bodies of information that we ordinarily consider concepts in cognitive science. Our prototypes, exemplars, and theories of tomatoes are connected while our prototypes, exemplars, and theories of water and gold are disconnected upon reflection. This means, the concept of, say tomato, is the right level of analysis as the concept contains a body of information that is connected in psychologically interesting ways that explain both fast and slow categorization.

4 Conclusion

The main motivation for considering whether we should be concept eliminativists or concept hybridists is that it has become apparent in recent years that no single psychological theory of concepts can explain all the data on categorization (Vicente & Martínez Manrique, 2016, p. 72). Prototype theory explains typicality effects but does poorly when it comes to explaining categorization using more theory-like representations. Exemplar theory explains how we classify an atypical but familiar object but is not ideal when it comes to classifying objects based on essential features. What distinguishes these different psychological theories is that they posit representations with different kinds of content. Prototype theory posits representations of typical features, exemplar theory posits representations of exemplars, and theory theory posits representations of less superficial features.

is being carried out. Prototypes and exemplars correspond to vertical downward processing of that structure and theories to horizontal and upward processing. Details do not matter here, the point is that we have a functionally integrated hybrid, which can be considered a natural kind. It is identified and accessed via the root node of that network structure. The structure is processed in different ways, depending on the context, giving rise to the appearance of different "types" of concepts.

[22] What if people keep insisting that sets of knowledge are not coordinated (Machery & Seppälä, 2011). Then one could either engage in more testing whether they are really not coordinated or functionally integrated or simply conclude that then the subject has indeed two or more concepts of, say, tomato. This does not mean that prototypes, exemplars and theories are never functionally integrated.

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

Concept eliminativism in psychology is the view that there is no reason to think that the different types of bodies of information stored in long-term memory share enough of the relevant properties to label them all 'concept'. I argued that the most developed eliminativist theory of concepts in cognitive science is based on two assumptions we have reason to reject. We should reject the idea that concepts are by definition stable bodies of information and that they should be individuated in terms of the type of content they represent. Since there are good reasons to reject these two assumptions, we have little reason (all things considered) to endorse concept eliminativism, at least until a more convincing kind has presented itself. On the contrary, I argued that we have reliable and familiar (e.g., Fregean) reasons to posit hybrid concepts and to assume that they are natural kinds after all.

References

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*(3), 263–308. https://doi.org/10.1016/0010-0277(83)90012-4

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(4), 637–660. https://doi.org/10.1017/S0140525X99532147

Blackburn, S. (1999). *Think: A compelling introduction to philosophy*. OUP Oxford. https://doi.org/10.1093/mind/110.438.430

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, *61*(1), 127–148. https://doi.org/10.1007/BF00385837

Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering*. Oxford University Press. https://doi.org/10.1093/oso/9780198814719.001.0001

Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543–566). MIT Press. https://doi.org/10.7551/mitpress/9383.003.0031

Del Pinal, G. (2015). Dual content semantics, privative adjectives, and dynamic compositionality. *Semantics and Pragmatics*, *8*, 7–1. https://doi.org/10.3765/sp.8.7

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press. https://doi.org/10.1093/0198236360.001.0001

Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. Oxford University Press. https://doi.org/10.1093/0198250614.001.0001

Kiefer, M. (2018). Cognitive control over unconscious cognition: Flexibility and generalizability of task set influences on subsequent masked semantic priming. *Psychological Research*, *83* 1556–1570. https://doi.org/10.1007/s00426-018-1011-x

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446

Lalumera, E. (2010). Concepts are a functional kind. *Behavioral and Brain Sciences*, *33*(2-3), 217–218. https://doi.org/10.1017/S0140525X10000403

Löhr, G. (2017). Abstract concepts, compositionality, and the contextualism-invariantism debate. *Philosophical Psychology*, *30*(6), 689–710. https://doi.org/10.1080/09515089.2017.1296941

Löhr, G. (2020). Concepts and categorization: Do philosophers and psychologists theorize about different things? *Synthese*, *197*(5), 2171–2191. https://doi.org/10.1007/s11229-018-1798-4

Machery, E. (2005). Concepts are not a natural kind. *Philosophy of Science*, *72*(3), 444–467. https://doi.org/10.1086/498473

Machery, E. (2009). *Doing without concepts*. Oxford University Press.

Machery, E. (2010). Précis of doing without concepts. *Behavioral and Brain Sciences*, *33*(2-3), 295–206. https://doi.org/10.1017/S0140525X09991531

Machery, E. (2016). The amodal brain and the offloading hypothesis. *Psychonomic Bulletin & Review*, *23*, 1090–1095. https://doi.org/10.3758/s13423-015-0878-4

Machery, E. (2017). *Philosophy within its proper bounds*. Oxford University Press. https://doi.org/10.1093/oso/9780198807520.001.0001

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632

Machery, E., & Seppälä, S. (2011). Against hybrid theories of concepts. *Anthropology and Philosophy*, *10*, 99–126. http://d-scholarship.pitt.edu/17471/

Margolis, E., & Laurence, S. (2018). Concepts. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/sum2019/entries/concepts/

Michel, C. (2022). A hybrid account of concepts within the predictive processing paradigm. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-022-00648-8

Murphy, G. (2004). *The big book of concepts*. MIT press. https://doi.org/10.1007/s13164-022-00648-8

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(2), 295. https://doi.org/10.1037/0033-295X.92.3.289

Piccinini, G., & Scott, S. (2006). Splitting concepts. *Philosophy of Science*, *73*(4), 390–409. https://doi.org/10.1086/516806

Prinz, J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge University Press. https://doi.org/10.7551/mitpress/3169.001.0001

Prinz, J. (2010). Can concept empiricism forestall eliminativism? *Mind & Language*, *25*(5), 612–621. https://doi.org/10.1111/j.1468-0017.2010.01404.x

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*(9), 458–470. https://doi.org/10.1016/j.tics.2013.06.004

Rey, G. (2010). Concepts versus conceptions (again). *Behavioral and Brain Sciences*, *33*(2-3), 221–222. https://doi.org/10.1017/S0140525X10000440

Rice, C. (2016). Concepts as pluralistic hybrids. *Philosophy and Phenomenological Research*, *92*(3), 597–619. https://doi.org/10.1111/phpr.12128

Rosch, E. (1999). Reclaiming concepts. *Journal of Consciousness Studies*, *6*(11-12), 61–77.

Rosch, E. H. (2011). Slow lettuce: Categories, concepts, fuzzy sets, and logical deduction. In R. Belohlavek & G. Klir (Eds.), *Concepts and fuzzy logic* (pp. 89–120). MIT Press.

Samuels, R., & Ferreira, M. (2010). Why don't concepts constitute a natural kind? *Behavioral and Brain Sciences*, *33*(2-3), 222–223. https://doi.org/10.1017/S0140525X10000452

Taylor, H., & Vickers, P. (2017). Conceptual fragmentation and the rise of eliminativism. *European Journal for Philosophy of Science*, *7*(1), 17–40. https://doi.org/10.1007/s13194-016-0136-2

Taylor, S. D., & Vosgerau, G. (2019). The explanatory role of concepts. *Erkenntnis*, *86*, 1045–1070. https://doi.org/10.1007/s10670-019-00143-0

Vicente, A., & Martínez Manrique, F. (2016). The big concepts paper: A defence of hybridism. *The British Journal for the Philosophy of Science*, *67*(1), 59–88. https://doi.org/10.1093/bjps/axu022

Weiskopf, D. A. (2009). The plurality of concepts. *Synthese*, *169*(1), 145–173. https://doi.org/10.1007/s11229-008-9340-8

Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science*, *29*(5), 719–736. https://doi.org/10.1207/s15516709cog0000_33

Löhr, G. (2023). Are concepts a natural kind? On concept eliminativism. *Philosophy and the Mind Sciences, 4*, 16. https://doi.org/10.33735/phimisci.2023.9632