Are concepts a natural kind?

**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 present a new challenge to concept eliminativism focusing on its most developed version by Edouard Machery (2005, 2009, 2010). I argue that his eliminativist argument depends on two assumptions that we have reason to reject. First, that concepts are *defined* as stable and context-independent bodies of information. Second, that distinct bodies of information in long term memory are partly individuated by the *type of content* of the information they store. Finally, I argue that certain pieces of information are connected enough to meet the conditions for being a natural kind. Therefore, we have currently no reason to think that concepts in cognitive science are not a natural kind.

**Keywords:** Concepts, Concept Eliminativism, Concept Pluralism, Concept Hybridism, Contextualism, Natural Kinds

**1 Introduction**

According to Margolis and Laurence (2018, 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 concepts are very dear to philosophers and cognitive scientists alike: The majority of them consider concepts to be indispensable for studying thought and language. Unsurprisingly, the proposal to eliminate the term from cognitive science (Machery, 2005, 2009) has faced much opposition (see e.g., commentary to Machery, 2010; Piccinini and Scott, 2006; Weiskopf, 2009; Rice, 2016; Vicente & Martínez Manrique, 2016; Taylor & Vosgerau, 2019). It may even be considered an affront to the profession, given that it is still widely assumed that philosophers’ main methodology is the analysis or engineering of concepts (cf. Cappelen, 2018; Jackson, 1998).

However, since there are arguably different notions of concept – at least one philosophical and one psychological one (Machery, 2009; Rey, 2010; Löhr, 2020) – there are different kinds of concept eliminativism, some of which are more revisionary than others. The version defended by Edouard Machery (2009) is a philosophically conservative one. Machery essentially 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* denotenatural kinds. This conservative kind of concept eliminativism will be discussed here. A more radical version that aims at eliminating the notion of concept as constituents of propositional attitudes will only be addressed briefly.

The question of whether we should be eliminativists, hybridists or pluralists of concepts has been critically discussed in various places (e.g., Piccinini and Scott, 2006; Weiskopf, 2009, Samuels & Ferreira, 2010; Vicente & Martínez Manrique, 2016; Rice, 2016). In this paper, I want to engage with the eliminativist argument against the psychological notion of concept in greater detail than previous discussions. To evaluate this *psychological concept eliminativism*, I focus on its most developed version by Edouard Machery (2005, 2009, 2010). However, my argument will have positive consequences for pluralism and hybridism as well. I argue that Machery’s eliminativism critically depends on two assumptions that we have reason to reject. First, that concepts are *defined* as being stable and context-invariant. Second, that *types* of bodies of information are individuated by the type of their content and processes they are used in. I conclude that there is currently no reason to think that ‘concept’ in cognitive science does not denote a natural kind.

**2 The argument for concept eliminativism**

To understand Machery's (2005; 2009; 2010) proposal to eliminate the notion of concept from cognitive science, it is crucial to first understand what exactly he takes the relevant notion of concept to be. According to Machery (2009, 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* 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. In fact, Machery argues that there is no single natural kind, but three different natural kinds that are picked out by the description.

Importantly, Machery (2005; 2009; 2010) makes explicit that he is *not* talking about the notion of concept that is commonly discussed in philosophy. Many philosophers use ‘concept’ to denote the constituents of propositional attitudes like beliefs and desires or the compositional truth conditional meaning of words and sentences. This is the notion used by various influential philosophers working on concepts like Rey (2010); Peacocke (1992); Fodor (1998); Prinz (2002), or Weiskopf (2009). Instead, Machery intends to propose a notion of concept that captures the *use* of the term 'concept' in contemporary cognitive science. This use differs, according to Machery, significantly from the use of the term in philosophy of language and mind in the sense that the focus does not lie on reference determination and compositionality and rather on categorization and decision making.[[1]](#footnote-1)

If it is right that the most common use of ‘concept’ in psychology is distinct from the most common use of the term in philosophy, then it is not clear why those philosophers who are more interested in compositional constituents of productive propositional attitudes and less interested in the way research on categorization is conducted should be worried about Machery’s concept eliminativism. Why should they worry about whether the “body of information (…) that is stored in long-term memory and that is used by default in the processes underlying (…) higher cognitive competences” (Machery, 2005, 449) picks out one, three or more natural kinds? Neither option is philosophically radical. We can still assume that the notion of concept prominent in philosophy has an important role to play in our theorizing about the mind and that it *does* pick out a single natural kind.

In fact, the notion of concept that is preferred by philosophers like Margolis and Laurence (2018), Peacocke (1992), Prinz (2002), Fodor (1998), Weiskopf (2009), or Rey (2010) is compatible with Machery’s notion of concept as bodies of “knowledge” retrieved by default. The set of representations we rely on to judge whether we should apply a word and the truth conditional meaning of this word may be different sets. For example, we may agree that the meaning of the word ‘bachelor’ is the complex concept (in the philosopher’s sense) UNMARRIED MALE and also concede that we judge whether this concept applies by relying on typical features of bachelors or exemplars, i.e., on our concept of bachelor in the psychologist’s sense (Fodor, 1998; Camp, 2015).

Similarly, note that by ‘knowledge’ or ‘information’, Machery refers to representations that he takes to *constitute* prototypes, exemplars or theories (as opposed to propositional attitudes). Even those who, unlike Machery (2005; 2009) in earlier publications,[[2]](#footnote-2) resist to characterize prototypes, exemplars or theories in terms of the notion of ‘belief’ will admit that these bodies of information are *bodies of intentional states*, i.e., bodies of mental representations with a certain truth-conditional content that can be ascribed to us on a personal level. It is not implausible that this content may be represented by means of the kind of concepts that the above-mentioned philosophers tend to focus on.

The main argument Machery (2005, 2009) puts forth is that the notion of concept that he takes to be central in cognitive science picks out different kinds of representations that are too heterogeneous to be considered a single natural kind that could be of interest for scientific psychology. What makes the types of bodies of information that, according to Machery, explain categorization, analogy- and decision making so different that they should not be classified together to form a single natural or at least scientific kind? Based on Richard Boyd's (1991) requirements for being a natural or scientific kind, Machery (2005, pg. 449) argues that bodies of information retrieved by default should only be considered a natural kind if they possess the following three characteristics:

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

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

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

What are the relevant (to the psychologist) properties that bodies of information could have in common? Generally, cognitive psychologists study bodies of information based on a number of dimensions:

(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 organization (whether and how the pieces of information that are stored are connected)

(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 perceptional or language-like amodal format)

(g) their neural implementation

Machery (2009, chapter 4) argues that there is overwhelming evidence that there are at least three types of bodies of information retrieved by default that underlie higher cognitive competencies (categorization, decision making, analogy making etc.). These types of bodies of information differ with respect to dimensions (a) and (b). Importantly, he insists that these dimensions are the critical ones for explaining higher cognitive competencies, in particular categorization (ibid.). Such an explanation, according to Machery, boils down to the question of what underlies the, what I will call “regularity effects” that were discovered in the psychology of categorization (see e.g., Murphy, 2004). In other words, because (a) and (b) explain regularity effects, according to Machery, they are the critical dimensions that we should rely on to individuate different types of bodies of information.

One especially important kind of regularity effect is that representations of typical features of a category can have processing advantages (“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. Another type of effect, what I call “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. Third, what I call “causal effect” is based on the observation that people tend to judge category membership not just by means of its superficial features but based on more theory-like causal or essentialist properties, e.g., that birds are not mammals or that the atomic number of gold is 79 (see Machery, 2009, Prinz, 2002 or Murphy, 2004 for summaries).

To explain categorization and higher cognition in general, Machery argues that psychologists discovered at least three fundamentally different types of bodies of information retrieved by default. These kinds, according to Machery, differ with respect to what he takes to be the relevant dimensions to explain higher cognition, i.e., (a), the type of content of the pieces of information stored in the concept, and (b), the type of cognitive processes they are used in (Machery 2005; 2009; 2010). He identifies the fundamentally different types of bodies of information as that which psychologists have called “prototypes”, “exemplars” and “theories”. Each of these psychological structures meets, according to Machery, at least the first and third of Boyd's requirements for being a scientifically interesting kind; each representational kind, taken individually, shares features (a) and (b).[[3]](#footnote-3)

Machery (2009, 2010) construes prototypes as sets of intentional states retrieved by default that *represent certain properties as typical for a category*. These sets 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. Machery takes the similarity measurement in the case of prototypes to be *linear*.[[4]](#footnote-4) He construes exemplars as sets of beliefs retrieved by default *about concrete exemplars of this category*. These sets of intentional states are used in *non-linear* similarity measure processes.[[5]](#footnote-5) Finally, he takes theories to be sets of beliefs retrieved by default *about nomological, causal, functional or generic properties associated with a category*. These beliefs are employed in cognitive processes that are different from similarity measures and more akin to causal reasoning.[[6]](#footnote-6)

It may be helpful 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 intentional states of a certain type (whose stored representations have a certain type of content)*.*[[7]](#footnote-7) Moreover, he argues that the bodies of knowledge retrieved by default in higher cognition do *not* differ with respect to (c) their organization[[8]](#footnote-8), (d) their function[[9]](#footnote-9) and (f) their format[[10]](#footnote-10). Machery also does not say much about potential differences with respect to (e) how they are acquired or (g) their neural implementation. 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.

The debate between Machery and his opponents (e.g., pluralists like Weiskopf, 2009 or hybrid views as Vicente & Martínez Manrique, 2016) is thus essentially about whether the bodies of information that are retrieved by default and that can explain typicality, exemplar and causal effects, form a single natural kind or whether we have to do with three or more natural kinds. In the remainder, I first discuss whether Machery’s notion of concept accurately describes the use of the term in cognitive science. Second, I object that the dimensions that his eliminativist arguments rest on (a and b), cannot explain typicality, exemplar and causal effects. Instead, what explains these effects is (c), the organization of our intentional states in long-term memory. In the final section, I argue that (c) justifies the notion that concepts are a natural kind as argued by hybrid theorists like – most recently – Vicente & Martínez Manrique (2016).

**3 Machery’s concept of concept revisited**

Significant weight in Machery’s argument that rests on the assumption that the term ‘concept’ picks out only those stable bodies of information that are retrieved by default is significant. 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’ could be used to potentially 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, p.196–197). Both arguments can be rebutted.

That concepts have to be stable is not new (e.g., Fodor, 1998; Rey, 2010). Many philosophers of concepts assume that concepts – as compositional constituents of propositional attitudes – have to be stable in order 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. Stability is *the* most important requirement for being a concept in the philosophical sense and if concepts are individuated by means of their content then a concept that changes its content will no longer be the same concept. Note that these strong requirements of stability 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.

However, if, as Machery says, the term ‘concept’ in cognitive science picks out not the kind of entities that constitute our premises and beliefs and instead denotes the bodies of beliefs we rely on to apply our words and concepts when categorizing, then “concepts” may not need to be as strictly stable as demanded by the logician or philosopher of language. Note that this can also be agreed upon by those philosophers who argue that philosophers and psychology refer to the same kind of thing when using the term ‘concept’ (e.g., 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.

Second, there is no conceptual reason why we could not explain regularity effects by means of context-dependent structures. Hence, it should not be presupposed. In fact, some have argued that contextualists can even better account for it (Lalumera, 2010; Casasanto & Lupyan, 2015; Löhr, 2017; Kiefer, 2018). Context-dependent structures can easily 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 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.

Third, it is not clear whether Machery’s description of the use of ‘concept’ in cognitive science is really accurate. Many leading contemporary psychologists working on categorization like Barsalou (1993; 1999), Pulvermüller (2013), Kiefer (2018) or even Eleanore 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 (e.g., 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.”

Finally, we should not stipulate that ‘concept’ only correctly applies to stable bodies of beliefs that are retrieved by default. 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. Machery is of course aware of this fact and argues in great length for the empirical claim that they are in fact stable bodies of information retrieved by default. However, there are many opponents to this empirical view (again, e.g., Barsalou, 1999; Kiefer, 2018), which raises the question of why Machery recommends restricting the notion of what he calls “concept” to bodies of beliefs retrieved by default in a context-independent manner as opposed to simply “bodies of beliefs or information underlying higher cognition”. This more relaxed notion would be more scientifically neutral as it would not rule out a controversial empirical hypothesis in an *a priori* definitional manner. However, it would also make concept eliminativism more difficult to defend, as I will argue.

**4 Machery’s eliminativist argument revisited**

So far, I summarized Machery’s description of what he takes the term ‘concept’ to mean in cognitive science and why he believes that this notion does not pick out one but at least three different natural kinds. I also argued that his eliminativist argument is based on the assumption that what explains “regularity effects” are (a) the type of content of the information stored in the bodies of information psychologists call “concepts” and (b) the processes they are used in. Most of the argumentative weight lies on this assumption (an assumption that Machery, 2005, 2009, 2010 does not provide evidence for). Thus, I will raise the following question: why should it be these two variables to individuate different types of bodies of information? Why should it be these variables that best explain categorization, as opposed to others, say, the organization of our intentional states in our long-term memory, i.e., (c)?

Before I get into details, 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 (e.g., Barsalou, 1999; Prinz, 2002; Pulvermüller, 2013) may, in principle, “forestall concept eliminativism” (Prinz, 2010). Prinz suggests that all bodies of knowledge underlying higher cognition could have something in common after all: all are perceptual symbols. In other words, if they were all of the same format (dimension f), this might be a unifying parameter that could justify the claim that concepts in psychology form a single natural kind.

However, even though Machery (2009, 2010) devotes much 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 (2010, 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. Again, Machery focuses on the type of content of information stored in a concept and the type of processes this information is used in to make his eliminativist claim. He focuses on type of content and type of cognitive processes, as opposed to its format, because he considers the former two, but not the latter, to be the key features that explain typicality, exemplar and causal effects.

I think that Machery's overall strategy is on the right track: the best way to look for a natural kind when it comes to higher cognition is to search for similarities with respect to the entities that explain the most interesting facts about a central part of higher cognition, namely categorization, i.e., typicality-, exemplar- and causal effects. However, I disagree with Machery about what actually explains these effects. I argue that what explains effects of higher cognition best is neither type of content (a) nor type of cognitive processes (b). What explains regularity effects in categorization, even according to Machery’s own account, is how long-term memory is organized, i.e., dimension (c). This, however, requires, as I will argue below, that we abandon the notion of concept described by Machery above, according to which long-term memory is structured in terms of stable bodies of information that are retrieved by default. Again, instead, we should leave it open to empirical investigation whether this is in fact the case.

That the dimension that gives us a genuine explanation of regularity effects is not type of content (a) or type of cognitive process (b) can be shown by a surprisingly simple observation: What explains, e.g., that typical members of a category are recognized faster is not whether subjects retrieve the representations with the content *birds typically fly* or whether they compare this piece of information to perceptual representation in a linear or non-linear way. What explains typicality effects is not the *kind of content* (individuated in terms of modal concepts like *typical*, *necessary*, *rarely* or *never*) subjects stand in a relation to or the type of process this representation is used in. What really explains typicality effect is that the representation with the content *birds typically fly* is much more readily available than the information that some birds may not be able to fly. We could observe the same typicality effects if we had the representation with the content *birds necessarily fly* or *birds* *rarely* fly immediately available (cp. Armstrong et al., 1983) – even if we do not believe both to be true upon reflection. As long as the information that birds fly is available in the relevant situation and irrelevant information is suppressed, we are faster at recognizing animals as bird if they fly.

Even with respect to Machery's own account, what seems to do most of the explanatory work is not his theory that there are different types of bodies of information individuated by their different types of *content* (typical features, exemplars, theory-like beliefs). What explains typicality, exemplar or causal effects, according to his own theory, is how Machery thinks knowledge is organized in long term memory. Machery argues that our knowledge of a single category, say *bird,* is organized in terms of stable chunks or “bodies” that become immediately available as a complete context-independent set in the right context, e.g., when hearing the word ‘bird’. Again, Machery argues that there is overwhelming evidence for these bodies of information, and I do not deny that there is such overwhelming evidence. However, if *type of* *content* does not actually explain regularity effects, this empirical argument is a red herring. Even on Machery’s own account, the three or more types of bodies of information that we rely on to apply the category of bird or the word ‘bird’ have something in common that explains categorization: they are prioritized when thinking about the same category. The information about typical features and exemplars of tomatoes are prioritized when thinking about tomatoes even if their prioritization profile may differ in different contexts.

To put the same point in different words: Machery (2009) insists that experimental evidence shows that the default chunks he argues structure long-term memory are individuated by means of the type of content of their constituents and the cognitive processes these constituents are used in. However, if the type of content of our representations does not explain regularity effects then this interpretation of the available evidence is beside the point. Again, few psychologists would deny that we store representations of typical features or exemplars and that we heavily rely on these representations when it comes to reliably applying our concepts and words. However, it is not the type of content that should interest us, e.g., whether we represent certain features as typical, essential or whether we represent exemplars. The question that should be of interest is when and based on what mechanism this information comes to mind more quickly than others. Thus, we should not individuate bodies of information in psychology in terms of type of content or type of cognitive process and instead focus on why and under which circumstances certain bodies of information become readily available.

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 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 of typical properties and the latter information about exemplars. But this characterization is misleading. 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?

To illustrate the argument further, consider Machery's (2009, p. 89) example of a prototype model of apple. This model contains a number of 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).

What according to Machery's own account explains why we recognize red and round objects earlier as apples than square and blue objects is *not* that we represent the redness or roundness of an apple *as diagnostic* or *as typical* of an apple. What, according to Machery, explains typicality effects is that the information that apples are red, and round comes to mind immediately when confronted with birds. What explains regularity effects is how long-term memory is organized (dimension c), namely in context-insensitive bodies of information that can be studied and accessed by means of familiar psychological experiments, such as typicality ratings. Whether we represent a feature as typical or necessary, for example, 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 in a way such that it is information about typical features of birds, rather than atypical features for example, that is more readily available. However, neither are relevant when it comes to explaining regularity effects.

**5 Are concepts a natural kind?**

Machery assumes that (i) the term ‘concept’ in cognitive science is defined as a stable body of information and (ii) that (a) *type of content* and (b) *type of process* –as opposed to, for example, the same intension or extension – individuate different scientifically relevant bodies of information, i.e., bodies of information that are natural kinds. However, (a) and (b) play little relevant explanatory work even on Machery’s own account and the claim that concepts *must* be stable does not have a solid basis. Where does this leave us with respect to the question of whether the concept of concept in cognitive science picks out a natural kind? It shows that Machery has given us no convincing reason to think that concepts in psychology are not natural kinds. Since I am not aware of any other worked out version of concept eliminativism in cognitive science, we have no reason to think that concepts in psychology are not natural kinds or that we should eliminate the term from our terminology in psychology.

On the contrary, by defining ‘concept’ in cognitive science as a body of information retrieved by default, Machery becomes vulnerable to overlooking that we *do* have reason to believe that the term picks out a single kind that meets Boyd’s conditions. To understand what these reasons are consider that explaining regularity effects is not the only task of concepts in cognitive science. What is just as important is to explain why the different pieces of information are intuitively connected and coordinated. Machery denies that they are “coordinated” (Machery, 2009; 2010) based on the observation that we can hold inconsistent beliefs about the same category. For example, we often hold to be true that tomatoes are vegetable in one context – relying on prototypes and exemplars – and that they are fruit in another context – where theory-like information is more readily available. However, Machery tends to ignore or downplay another interesting and important observation pertaining to higher cognition: we usually do not hold such contradictory beliefs about tomatoes after deliberation. Psychologists have to explain not just fast categorization but categorization after deliberation.

Even Machery will agree that the different kinds of information associated with the same category are not as disconnected or uncoordinated as we might understand him to suggest. We can compare the tomato case with categories where our bodies of information are more disconnected. 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. 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 our bodies of information we rely on to recognize objects as *gold* and *water* are organized very differently in the mind and that, unlike our different contradictory assumptions about tomatoes, they do *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 for *tomato* since we are aware that tomatoes are not really vegetables.

I do not know what exactly the mechanism is that suppresses connected information in different contexts. For example, I do not know what mechanism makes it the case that when hearing the term ‘tomato’, representations of vegetables become readily available in a supermarket context, while I think of them as pieces of fruit in a biology class. However, without too much speculation, relevance seems to be the crucial factor. The information that tomatoes are vegetables has been more relevant in the former than the latter context. Importantly, if we assume that deliberation simply generates a context in which the information that our beliefs about tomatoes are connected becomes relevant, then we have no reason to suspect that the mechanism that makes information available is fundamentally different when it comes to categorizing quickly compared to when categorizing without time pressure. For example, the information that pianos are heavy might be privileged in a moving context and inhibited in a concert context. Still, both pieces of information appear to be part of the same concept of piano since when deliberating, i.e., when creating a context in which both pieces of information are relevant, we *do* understand that both representations belong to the same category.

To return to the question of whether concepts are a natural kind, recall Machery’s adaption of Boyd’s three conditions for natural kindness.

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

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

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. Long-term memory is organized such that it prioritizes different bodies of information in different contexts. The mechanism that prioritizes information likely depends on relevance. Bodies of information of typical features are more relevant in some contexts than others. Deliberation 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 they are connected in our minds. 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 likely explains why they are also connected when deliberating. 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 something that is the crucial determiner of both categorization under time pressure *and* after deliberation. This feature is the kind of connectedness that hybrid theorists like Vicente and Martínez Manrique (2016) have been insisting on. The connectedness 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. 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 clearly be contrasted with categories that we consider completely disconnected, as we saw with the difference between water and gold.

Finally, the connectedness is specific not to a sub-kind of stable bodies of information – contrary to what Machery has argued. 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. To focus only on the set of bodies of information that are immediately available would not be the right level of analysis. The reason is that the crucial properties that explain categorization are properties of the concept per se and not merely parts of the concept or bodies of information that are most readily available. In other words, the tomato concept may contain different bodies of information that are not always available and even perhaps inconsistent. They are still part of the same concept.

**Conclusion**

The main motivation for considering whether we should be concept eliminativists, concept pluralists 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 (cp. Vicente & Martínez Manrique, 2016, p. 72). Prototype theory explains typicality effects but does poorly when it comes to explaining categorization by means of 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. As argued above, what distinguishes these different psychological theories is that they posit representations with different kinds of contents. Prototype theory posits representations of typical features, exemplar theory posits representations of exemplars and theory theory posits representations of less superficial features.

Concept eliminativim in psychology is the view that there is no reason to think of the different types of information to be connected or coordinated enough to justify the common label ‘concept’. I argued that the most developed eliminativist theory of concepts in cognitive science is based on two assumptions we should 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 no reason to endorse concept eliminativism, at least until a more convincing kind has presented itself. On the contrary, I argued that we have very good and highly familiar (e.g., Fregean) reasons to posit 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.

Barsalou, L.W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In: A. C. Collins, S. E. Gathercole, and M. A. Conway(eds.). *Theories of memory*, 29–101. London: Lawrence Erlbaum Associates.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and brain sciences*, *22*(4), 637-660.

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical studies*, *61*(1), 127-148.

Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering*. Oxford University Press.

Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In L. Margolis & S.

Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543–566).

Cambridge, MA: MIT Press.

Camp, E. (2015). Logical concepts and associative characterizations. In E. Margolis & S. Laurence (Eds.), *The Conceptual Mind: New Horizons* (pp. 591–621). Cambridge, MA: MIT Press.

Churchland, P. M. (1981). Eliminative materialism and propositional attitudes. *The journal of Philosophy*, 78(2), 67-90.

Del Pinal, G. (2015). Dual Content Semantics, privative adjectives, and dynamic compositionality. *Semantics and Pragmatics*, 8, 7-1.

Dennett, D. C. (1991). Real patterns. *The journal of Philosophy*, 88(1), 27-51.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.

Jackson, F. (1998). From metaphysics to ethics: *A defence of conceptual analysis*. Oxford University Press.

Kiefer, M. (2018). Cognitive control over unconscious cognition: flexibility and generalizability of task set influences on subsequent masked semantic priming. *Psychological research*, 1-15.

Lalumera, E. (2010). Concepts are a functional kind. *Behavioral and Brain Sciences*, *33*(2-3), 217-218.

Löhr, G. (2017). Abstract concepts, compositionality, and the contextualism-invariantism debate. *Philosophical Psychology*, *30*(6), 689-710.

Löhr, G. (2020). Concepts and categorization: do philosophers and psychologists theorize about different things? *Synthese*. 197(5), 2171-2191.

Margolis, E. and Laurence, S. (2018), "Concepts", The Stanford Encyclopedia of Philosophy (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2019/entries/concepts/>.

Machery, E. (2005). Concepts are not a natural kind. Philosophy of Science, 72, 444-467.

Machery, E. (2009). Doing without Concepts. New York: Oxford University Press.

Machery, E. (2010). Précis of Doing without Concepts. Mind & Language, 25, 601-610.

Machery, E. (2016). The amodal brain and the offloading hypothesis. Psychonomic Bulletin & Review, 23, 1090-1095.

Machery, E. (2017). *Philosophy within its proper bounds*. Oxford University Press.

Murphy, G. (2004). *The big book of concepts*. MIT press.

|  |
| --- |
|  |

Nosofsky, R. M. (1991). Relation between the rational model and the context model of categorization. *Psychological Science*, 2(6), 416-421.

Peacocke, C. (1992). *A study of concepts*. The MIT Press.

Piccinini, G., & Scott, S. (2006). Splitting concepts. *Philosophy of Science*, *73*(4), 390-409.

Prinz, J. (2002). *Furnishing the Mind: Concepts and their perceptual basis*. Cambridge University Press.

Prinz, J. (2010). Can concept empiricism forestall eliminativism?. *Mind & Language*, *25*(5), 612-621.

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in cognitive sciences*, *17*(9), 458-470.

Rey, G. (2010). Concepts versus conceptions (again) 1. *Behavioral and Brain Sciences*, *33*(2-3), 221-222.

Rice, C. (2016). Concepts as pluralistic hybrids. *Philosophy and Phenomenological Research*, *92*(3), 597-619.

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). Cambridge: MIT Press.

Samuels, R., & Ferreira, M. (2010). Why don't concepts constitute a natural kind?. *Behavioral and Brain Sciences*, 33(2-3), 222-223.

Taylor, S. D., & Vosgerau, G. (2019). The Explanatory Role of Concepts. *Erkenntnis*, 1-26.

Taylor, H., & Vickers, P. (2017). Conceptual fragmentation and the rise of eliminativism. *European Journal for Philosophy of Science*, *7*(1), 17-40.

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2013). The neural representation of abstract words: the role of emotion. *Cerebral Cortex*, *24*(7), 1767-1777.

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.

Weiskopf, D. A. (2009). The plurality of concepts. *Synthese*, 169(1), 145.

1. This does not mean that many philosophers are not interested in whatever is picked out by the psychological notion, or that psychologists should not be interested in the philosophical notion. It also does not mean that both notions are not related in important ways. See Löhr (2020) for a closer look at this distinction. [↑](#footnote-ref-1)
2. Compare this use to Machery (2017) where he merely talks about “belief-like states”. [↑](#footnote-ref-2)
3. Machery (2005, 2009, 2010) says nothing about the second requirement for being a natural kind (that these bodies of knowledge possess shared properties because of some causal mechanism).. [↑](#footnote-ref-3)
4. This means that the more representations two sets of representations have in common, the more similar are they judged as. [↑](#footnote-ref-4)
5. Machery (2009, 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.” [↑](#footnote-ref-5)
6. Machery (2009, 106) summarizes an example from Murphy and Medin (1985: 295) thus: “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.” [↑](#footnote-ref-6)
7. 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, 2009; 2010; 2016 view). [↑](#footnote-ref-7)
8. He takes them to be connected as part of the stable body of information meaning that they are coordinated in the sense that the retrieval of one is accompanied by the retrieval of the other. [↑](#footnote-ref-8)
9. Machery (2005, 2009, 2010) takes all fundamental kinds of information to fulfill a number of different cognitive competencies, which is an essential part of his “heterogeneity hypothesis”. [↑](#footnote-ref-9)
10. Machery (2009; 2016) is very explicit that he endorses a monist amodal view of concepts. [↑](#footnote-ref-10)