Representational Kinds

Joulia Smortchkova (University of Oxford) & Michael Murez (University of Nantes)

Abstract

Many debates in philosophy focus on whether folk or scientific psychological notions pick out cognitive natural kinds. Examples include memory, emotions and concepts. A potentially interesting type of kind is: kinds of mental representations (as opposed, for example, to kinds of psychological faculties). In this chapter we outline a proposal for a theory of representational kinds in cognitive science. We argue that the explanatory role of representational kinds in scientific theories, in conjunction with a mainstream approach to explanation in cognitive science, suggest that representational kinds are multi-level. This is to say that representational kinds' properties cluster at different levels of explanation and allow for intra- and inter-level projections.

1. Introduction

Cognitive scientists seek to explain psychological capacities and often appeal to mental representations in doing so. However, not all representations are equally useful to cognitive science, and only some cluster together to form kinds and "carve cognition at its joints". What makes some representational categories *natural representational kinds* in cognitive science? This is the question we explore in this chapter.

Some examples of candidate representational kinds within cognitive science include: essentialist representations, prototype representations, exemplar representations, object-files, mental maps, analog magnitude representations, action representations, structural geon-based object representations, visual working memory representations, quasi-pictorial representations, face representations in the FFA (face fusiform area), mental models, body schema representations, singular terms in the language of thought, bug representations in the frog's brain, dominance hierarchy representations in baboons, naïve theories, core cognition representations, syntactic tree representations, Universal Grammar, content (or address-) addressable representations in memory, edge representations, etc. The status of each of these categories as a genuine representational kind is controversial. Some will strike the reader as more plausible or as better examples of representational kinds than others. We don't wish to take a firm stand on any case listed above in particular, but merely appeal to them in order to help fix the reference of what we mean by "representational kind" within the context of cognitive science.

The issue that interests us is not what a representation is, nor whether the notion of "mental representation" in cognitive science itself picks out a natural kind (Ramsey, 2007; Shea, 2018). Rather, assuming there are mental representations, we are interested in what it is for a particular class of them to form a natural kind (we will often refer simply to "representational kinds", implying that they are natural or real).

We are also not interested in entering into debates about whether folk psychological concepts (such as belief) aim at or succeed in picking out natural kinds (Jackson & Pettit, 1990; Lycan, 1988). Nor are we concerned with kinds of representations that are not mental (for instance, kinds of external representational artifacts, such as photographs). Our concern is with the notion of representation as it figures in cognitive science.

This question must also be distinguished from the question of what in general constitutes a kind in cognitive science. There are presumably many varieties of kinds within cognitive science, some of which are not representational. For example, "innateness" has been argued to be a cognitive natural kind (Khalidi, 2016). While there is a class of innate representations, there might also be innate cognitive structures, processes, or mechanisms that are not representational in nature.[1]

There are two parts to the notion of a "representational kind". First, its members must be mental representations. Second, they must form a kind. We argue that reflection on these two components of the notion, together with certain widely shared assumptions within cognitive science, lead to our central thesis: *representational kinds are multi-level.*

The chapter will unfold as follows. In section 2 we introduce a contrast between a non-natural representational category and a natural representational kind, and argue that the difference between them has to do with a contrast in their respective explanatory depths. In section 3, we outline a notion of non-classical natural kindhood which can be applied to representations. In section 4 we flesh out the multi-level proposal and the relevant notion of "depth" by appealing to the nature of mental representations and their role in multi-level explanations. In section 5 we refine the multi-level thesis. In section 6 we reply to objections.

2. A contrast case

A first way to get at the notion of "representational kind" is by noting an intuitive contrast between two exemplary categories of representations: the class of wombat-representations, on the one hand, and the class of action-representations, on the other (Gallese, Fadiga, Fogassi, & Rizzolatti, 2002).

The two classes have much in common. Both are characterized by the domain of entities in the world which they represent: wombat-representations refer to the domain of wombats, while action-representations refer to the broader domain of actions performed by living beings. Thus, both classes are clearly characterized semantically, or by their intentional content.

Based on this, one might hold that there are necessary and sufficient conditions for membership in each class. An entity is a wombat-representation just in case it correctly applies to all and only wombats. Similarly, an entity is an action-representation just in case it correctly applies to all and only (visible) actions. Yet the former is not a good candidate for representational kindhood, while the latter is. The study of wombat-

[1] Note that it wouldn't follow merely from the fact that "innate" is a cognitive kind, that "innate representation" is a representational kind, even if it turns out that "representation" is also a cognitive kind.

representations is of little interest to cognitive science. Intuitively, such a class is not likely to correspond to a joint in (cognitive) nature, and so is not a worthwhile target of cognitive scientific inquiry. By contrast, there is a dynamic and successful research program centered around the investigation of action-representations. What distinguishes the two cases?

A natural reaction might be that the difference lies entirely in the domains that each class of representations targets, as opposed to something about the representations themselves. Actions are a very general category of things, which are of evolutionary and psychological significance to humans and other animals. By contrast, wombats are an idiosyncratic domain, which is unlikely to be of much interest (unless you're an ethologist specialized in wombats).

This is obviously true, but it misses the point. There are other very general and significant domains, such as food[2] (Lafraire, Rioux, Giboreau, & Picard, 2016), or supernatural entities[3] (Atran, 2002), of which humans undoubtedly form representations. Yet upon investigation, these classes of representations do not appear, by cognitive scientists' lights, to correspond to natural representational kinds (Atran, 2002; Shutts, Condry, Santos, & Spelke, 2009). Members of these classes lack a sufficient degree of unity with respect to their representational properties, even if they are domain-specific.

Moreover, not all representational kinds are characterized by their specific target domains. Many plausible candidates for representational kindhood range across multiple referential domains (e.g. essentialist representations). Other plausible candidate representational kinds might not be characterized by the domain of things they represent, but rather in terms of other properties, such as their *format* (e.g. analog representations). And there are debates about whether representations of the very same domain constitute a representational kind, e.g. whether the different bodies of information that pick out the same category form a representational kind, corresponding to "*the* concept" of that category (Machery, 2009).

Consider an analogy. Dust and gold differ both in their significance for humans, and with respect to their natural kindhood. However, what makes it the case that dust is

---

[2] While food is an important category for human survival, we need to distinguish a category of representations, from the category they represent. In order for food representations to form a kind, it is not enough for food to be an interesting and important category; we would need to find properties that representations of food share *qua* representations. Yet food representations lack unity, from a cognitive perspective. For instance, young infants do not seem to possess a dedicated learning mechanism for food items (while they possess dedicated learning mechanisms for tracking agents and small quantities, for instance), and adults distinguish between edible and inedible items via a motley collection of domain-general mechanisms, which include texture, color, and smell (Shutts et al., 2009).

[3] Beliefs about supernatural beings are similar to food representations. According to Atran (2002), there is no special cognitive system for thinking about religion or representing supernatural beings. Religious beliefs are instead underpinned by "a variety of cognitive and affective systems, some with separate evolutionary histories, and some with no evolutionary history to speak of. Of those with an evolutionary history, some parts plausibly have an adaptive story, while others are more likely by-products." (Atran, 2002, p. 265). Religious beliefs are not a natural kind, and representations of supernatural entities are not kinds either, but are distributed across different systems, have different formats, and different evolutionary histories (p. 113).

not a natural kind has little to do with its lack of significance: what prevents dust from being a natural kind is simply that what instances of dust have in common that makes them dust is something entirely superficial. Dust is a nominal kind. Whatever appears dusty, or whatever we are disposed to categorize as dust (at least in normal circumstances) just *is* dust. There can be fool's gold – stuff that seems like it is gold but turns out not to be upon further investigation. Yet there cannot be "fool's dust", nor "shmust" on Twin Earth (Dennett, 1994). This is because dust has no underlying nature, or essence. In turn, there is no science of dust as such.

The intuition we hope to elicit, and build upon in what follows, is that wombat representations are to action-representations what dust is to gold. The class of wombat-representations is superficial, whereas there is something deeper that action-representations have in common, and which makes them fruitful targets for cognitive scientific inquiry. As we see it, the important difference is that the only thing that wombat-representations have in common is their intentional or semantic content. Wombat-representations may constitute an intentional (or semantic) category. But intentional categories are not necessarily representational kinds, which are those that matter from the perspective of cognitive science.

Action-representations' status as a representational kind depends on its being possible to discover other sorts of properties that this particular class of representations, which we initially characterized merely in terms of their target domain, non-accidentally share. First, there are specific psychological effects associated with the category. For instance, in the apparent body motion paradigm, when people observe a hand action moving across the body, they perceive the longer but anatomically plausible path, rather than the shorter but anatomically implausible path (Shiffrar, 2008). Also, action representations are stored separately in visual working memory from colors, locations, and shapes, as evidenced by action-specific memory capacity limits (Wood, 2007). Furthermore, the category exhibits unity at the neural level. Action-representations seem to be reusing the same neural mechanisms that are used for action production (Rizzolatti & Sinigaglia, 2007). We can study their neural correlates, as they elicit activations in the superior temporal sulcus (STS) and in the premotor cortex area F5 (Jellema & Perrett, 2007).

Note that several of the interesting, unobvious properties of action-representations are possessed not merely in virtue of what the representations are about. None of the aforementioned properties of action-representations can be discovered by reflecting on their content (or by considering properties directly related to their content, such as entailment-patterns). In our view, this hints towards a central aspect of representational kinds: representational kinds in cognitive science are inherently multi-level. This connects to one of the central features of explanations in cognitive science: merely intentional categories belong to horizontal explanations, while representational kinds belong to vertical explanations, and figure in the functional decomposition of a capacity (Cummins, 1983). This is what lends certain classes of representations the "depth" which enables them to constitute genuine kinds, despite the fact that – by contrast with gold – representations plausibly lack classical essences. We will discuss multi-levelness in section 4, but first, in order to set the stage for this view of representational kinds, we need to say some more about the non-classical view of natural kinds that it presupposes.

3. Non-classical kindhood

Our aim is not to enter into debates on natural kinds in general outside of cognitive science, nor it is to defend a particular view. It is rather to sketch an independently plausible, broader approach to natural kinds that is compatible with our specific proposal for representational kinds.

Natural kinds figure in scientific explanations – traditionally, natural kinds are taken to be the categories which feature in natural laws – and in scientific predictions and investigations. It is often said that natural kinds are "inductively deep" (Carey, 2009; Quine, 1969). This means that a natural kind supports projection: there are numerous true, non-obvious generalizations that can be made about its members, which share scientifically interesting properties because there is an underlying factor that they have in common, rather than by accident. According to the classical account of natural kinds (Kripke, 1980; Putnam, 1975), this underlying factor is an essential property, which also provides necessary and sufficient conditions for kind membership, in addition to causing members to exhibit a shared syndrome of properties – a cluster of generally more superficial features that kind-members tend to exhibit on average.

Syndrome properties are epistemically useful in initial identifications of potential members. They may even be practically indispensable: one typically postulates a kind prior to grasping its essence, which is psychologically represented only by a "placeholder" (Gelman, 2004). Yet from a metaphysical perspective, syndromes are only contingently related to kind-membership. A member of a kind may fail to satisfy its syndrome to any significant degree, while a non-member may satisfy it perfectly. Ordinary users of kind-concepts tacitly know this, as shown by research on "psychological essentialism" (Gelman, 2004).

Kind-membership depends on possession of the potentially hidden essence, rather than superficial resemblance. So shared kindhood is inferable from shared essence even in the absence of superficial resemblance. In turn, knowledge that superficially dissimilar entities are of a kind enables one to justifiably expect them to bear some deeper commonalities than otherwise evident – which enables kind-concepts to play a major heuristic role in guiding scientific inquiry.

One well-known problem with the classical account of kinds is that it fails to apply to special sciences, such as psychology (Dupre, 1981; Fodor, 1974; Millikan, 1999). For instance, it seems implausible that there is a classical essence that all and only action-representations share. If all kinds were classical, there would presumably be no representational kinds in cognitive science whatsoever.[4]

Note that this is true even if membership in the category can be defined strictly, by reference to its content. Some representations can correctly apply to all and only

[4] An anonymous reviewer suggested that there might be representational kinds that possess a classical essence. One such candidate kind is "perceptual demonstrative": there might be an essence (not in the sense of a microstructure, as for chemical natural kinds, but in the sense of a sufficient and necessary condition) for how perceptual demonstratives have their reference fixed. We are not convinced there is such an essence. But even if we were to discover such an essence for the class of perceptual demonstratives, it's unclear to what extent this class would be a representational kind in cognitive science (the focus of the chapter), rather than a kind outside the domain of cognitive science (for instance, in metasemantics or normative epistemology).

actions (i.e. have the content: being an action) without being members of the relevant representational kind. For example, imagine someone were to design an artificial intelligence with the ability to detect actions, yet whose action-representation capacities were based on entirely different computational mechanisms from those found in the human mind/brain: for instance, the artificial intelligence might detect actions thanks to a lookup table which accessed a huge database with stored templates for every single action it ever encountered (Block, 1981).[5] In such a case, cognitive scientists would not classify the artificial intelligence's action-recognition capacity as belonging to the same representational kind as our own.

We will return to the issue of what – beyond intentional content – unites representational kinds. But whatever it is, there is little hope that it will be a classical essence. Presumably, there are no interesting necessary and sufficient conditions for being an "action-representation" in the sense in which this expression picks out a representational kind for cognitive scientists.

Luckily, there are alternatives to the classical account of kinds. Some accounts, such as Boyd's influential homeostatic property cluster (HPC) view (Boyd, 1991), preserve the idea that kinds play a privileged epistemic and scientific role without postulating essences. Many of our scientific practices tacitly rely on, as well as support, the contrast between merely nominal kinds (those for which syndrome-possession constitutes membership), and natural ones – even within the special sciences where classical essences are plausibly absent.[6]

According to HPC, natural kinds are the result of an interplay between human classificatory interests and mind-independent joints in nature. HPC thus seeks to combine a pragmatist conception of kindhood, as partially dependent on human practices and concerns, and a realist approach, which captures the traditional intuition that natural kinds mark non-accidental and non-conventional divisions in the world.

As in the classical account, HPC kinds are defined by two factors: co-occurring property clusters (syndromes) that members tend to instantiate more than non-members; and a second factor which accounts for the non-accidentality of resemblance between members. By contrast with classical essences, however, property clustering is grounded in a collection of "homeostatic mechanisms". These mechanisms are dynamic (they may evolve over time) and context-dependent (they may cause the syndrome only in certain "normal" environments, rather than in all circumstances).

---

[5] An anonymous reviewer suggests that the brain could use something like lookup-based convolutional networks to categorize actions, and that this is not a possibility that we can rule out a priori. We agree that if this turns out to be the case, then we would classify the AI's action-recognition system as deploying the same representational kind as our own. In this example we suppose that the look-up table AI and the brain do not use the same computations, to make the point that when computations are radically different, classificatory practices in cognitive science tend to split the kind.

[6] While we're not discussing in detail how classical and non-classical kinds differ, we mention here some of the main dissimilarities: non-classical kinds admit of imperfections in the members of the kind (for example Sphynx cats are still cats, even if they don't have fur), they have fuzzy boundaries between categories and admit of overlaps between members of different kinds (such as mules, which are the outcome of a cross between a jack and a mare).

For Boyd, the relevant notion of "mechanism" is a broad, inclusive one, not to be confused with the notion of "mechanism" recently popular elsewhere in philosophy of cognitive science (Machamer, Darden, and Craver, 2000). Boydian (non-classical) "essences" need not be reducible to more basic interacting entities and their activities. Nor do they need to be intrinsic to their members. For instance, Boydian kinds can possess historical essences – meaning that property clustering is sustained over time by mechanisms of replication or information transfer between members, some of which may be external to kind-members.

Though the classical account has its defenders (Devitt, 2008), and is sometimes still taken to apply to "microstructural" kinds in the fundamental natural sciences[7], non-classical accounts like Boyd's are dominant in philosophy of the special sciences. Indeed, Boyd's view, and other similar ones (Millikan, 1999) are designed to allow for functional kinds (which are defined in terms of extrinsic causal interactions) to potentially count as a subspecies of natural ones, and do not require that kinds necessarily figure in law-based explanations. As a result, non-classical views are attractive to defenders of broadly functionalist conceptions of the mind (though see Ereshefsky and Reydon (2015) for a contrary opinion). In particular, such views are compatible with the widespread assumption that psychological kinds are multiply realized (Boyd, 1999).[8]

Henceforth, we will assume that there are representational kinds, and so that some version or other of the non-classical account of kindhood will apply to them (for our purposes, we will not commit to one specific non-classical account of kinds). However, it remains unclear how exactly to apply a non-classical account to representational kinds. Although we have contrasted intentional categories, which are merely nominal or superficial from the perspective of cognitive science, with genuine representational ones, we have yet to explain in any detail what gives representational kinds their characteristic "depth". In the next section, we suggest that the answer to this question is to be found by reflecting on the nature of mental representations, and their role in vertical, multi-level explanation – which is the paradigmatic form of explanation that takes place in cognitive science.

4. Representational kinds as multi-level

Merely intentional categories – like the class of wombat-representations – illustrate the fact that shared intentional content is not sufficient for representational kindhood in cognitive science (even if there might be intentional kinds outside of cognitive science, see section 6). Of course, there will be a cluster of other *intentional* properties that tend to be co-instantiated with the property of having a given intentional content. For instance, people who possess wombat-representations probably tend to use them to make certain inferences, such as moving from the premise "X is a wombat" to the conclusion that "X is a mammal" or "X is brown". But this is a consequence of the meaning of "wombat" – or at least of widely shared world-knowledge about wombats.

[7] Though see Needham (2002).

[8] The classical account of natural kinds is incompatible with the multiple realizability thesis, because psychological kinds are defined not in terms of their internal structure (what they are made of), but in terms of their functional properties (what they do). These properties can be implemented in different micro-structures and materials.

Other sorts of properties of the representational category are hard to derive from its content. Indeed, it is plausible that there are simply no non-intentional properties that non-accidentally cluster with the aforementioned intentional ones.

For action-representations, on the other hand, what we can learn about the category is not exhausted by its intentional content. Rather, intentional content plays a role akin to ostension in the case of physical kinds, like gold: it fixes the reference of the potential representational kind and guides the (largely empirical) search for other properties which are not obvious consequences of its having the content that it has.

What, then, makes a cluster of representations in cognitive science a kind rather than a mere category? Taking the example of action-representations in section 2 to be representative, we suggest that talk of "representational kinds" is justified when the co-clustering of properties is multi-level, that is, when clusters of properties associated with the kind are discovered at different levels and are used in vertical, decompositional, explanations.

Thus, our master-argument in favor of the multi-levelness of representational kinds stems from the role played by natural kinds in scientific endeavors, and from a mainstream approach to explanation in cognitive science:

> 1. Natural kinds play a central role in the epistemic project of scientific disciplines: they are used to project properties when making predictions, and for explaining systematic observations.
>
> 2. The epistemic project of cognitive science is to decompose a complex cognitive capacity into simpler components and their relations. This decomposition is articulated at multiple levels of explanation, and is iterative: a complex cognitive capacity is composed of subsystems that can in turn be broken up into further sub-subsystems until no more decomposition is possible.
>
> 3. From 1 and 2, the best account of natural representational kinds is one that allows them to play a role in these multi-level decompositional explanations, with different aspects of representations lining up at different levels of explanation.

The picture of the mind suggested by such an approach should be familiar to the reader: according to the computational-representational theory of mind, the mind is a computational and representational system (Fodor, 1975), which accounts for thinking by viewing mental processes as computational processes operating over representations, with "formal" or "syntactic" properties lining up with semantic properties. At the semantic level, the (logical) thinker moves rationally from premises to conclusion in a truth-preserving way. This semantic relation between mental states corresponds, at the syntactic level, to a causal transition, ensured by appropriate formal relations between thought-vehicles. Token mental representations are thus concrete, though potentially spatio-temporally distributed, particulars with a dual nature: on the one hand they have intentional, semantic properties; on the other hand, they possess non-intentional, non-semantic, vehicular properties.

Mental representations' vehicular properties include not only syntactic ("formal" properties), but also physical properties which correspond to the physical structures

that realize (or "implement") the representations (for humans, neurons and neural configurations). We do not wish to endorse any specific approach to CRTM, such as the classic approach put forward by Fodor and Pylyshyn: our account is compatible with connectionist as well as classical views of computations. What we do endorse is the idea that both semantic and vehicular properties of mental representations are in the domain of cognitive science, an idea at the core of CRTM (in any form).

In CRTM mental representations are individuated by an appeal to both contents and vehicles: classes of representations that are alike with respect to their contents, might not be alike with respect to the properties of their vehicles, and thus play different roles in processing (Shea, 2007). This is why merely intentional categories do not necessarily correspond to representational categories of use in cognitive science.

Understanding how cognitive scientists explore non-observable vehicular properties of mental representations throws light on the process by which representational kinds are uncovered. Representations' vehicular properties cannot be discovered a priori, nor (usually)[9] directly observed. But in many cases they can be inferred from characteristic psychological effects. Following Cummins (1983, 2000, 2010), we distinguish between psychology's primary explananda, that is cognitive capacities (such as language acquisition or depth perception), which need not be discovered empirically, and secondary explananda, that is psychological effects, which have to be discovered empirically (for example the effects of the algorithms and representations employed for language acquisition or for depth perception). The discovered effects are not the endpoints of exploration. Instead, effects need explanations in turn, which appeal to constituents of the cognitive systems (including functional and physical components) and their organization.

Again drawing on Cummins, we can illustrate the notion of 'psychological effect' as follows. Multiplications can be performed via two types of mechanisms: the first type of mechanism multiplies each digit of one factor by each digit of the other, then adds the results; the second mechanism multiplies by repeated addition, and it represents 3 x 3 as 3 + 3 + 3. The 'linearity effect' (i.e. computational cost as a linear function of multiplier size) is only shown by the second type of mechanism: for example, 3 x 6 requires twice as many operations as 3 x 3, and as a result takes twice the time (all other things equal). The linearity effect is characteristic of the particular algorithm through which the semantically-specified task of multiplication is accomplished, but it is tangential to *what* the mechanism does. What it shows is something about *how* the mechanism works, and what kind of algorithms and representational vehicles it uses.

Much valuable work in psychology simply demonstrates a certain effect. Ultimately, however, psychology appeals to effects not in their own right, but in order to study processes and properties of the representations involved in various cognitive

[9] An anonymous reviewer casts doubts on this affirmation. On the one hand, they notice that there are recent claims that suggest that representations can be directly observed (Thomson & Piccinini, 2018) or causally manipulated (Widhalm & Rose, 2019). On the other hand, they doubt that psychological effects are directly observed. Regarding the first point, the proposal in this chapter is compatible with representations sometimes being observed: if and when they are, this would facilitate the discovery of representational kinds. However, direct observation of representations does not seem to be frequent, given the present state of our knowledge. Concerning the second point, we're not arguing that effects are directly observed: effects are uncovered through the usual tools of cognitive science (reaction time, priming, etc.).

capacities. One example is the study of memory capacity limits defined in terms of the cost of processing a given number of stimuli. Limits on working memory depend on the number of representations to be retained (typically 3 to 5 chunks of stimuli) (Cowan, 2010). To return to our previous example, one of the reasons to consider "action-representations" a representational kind is precisely that distinctive capacity limits can be observed for this category of representations. How many action-representations can be stored at one time is entirely incidental to the overall semantic function of standing for actions. But it puts constraints on the type of vehicles, *qua* vehicles, that can be involved (whether they are stored by the number of objects or by the number of features).

Another example are analog magnitude representations, representations of spatial, temporal, numerical, and related quantities (Beck, 2015). The main evidence used to uncover the analog format of magnitude representations is Weber's law, according to which the ability to discriminate two magnitudes is determined by their ratio (Jordan & Brannon, 2006). The closer the magnitudes are in their ratio, the more difficult it is to distinguish them until the point (Weber's constant) where they are not discriminable. Again, the presence of this effect shows something about the nature of the representations involved and requires moving beyond the semantically-characterized capacity (in this case the capacity to represent approximate quantities).

The strategy of decomposing complex capacities into their components, which are then investigated, points toward a central feature of explanations in cognitive science: they are vertical rather than horizontal explanations (Drayson, 2012). The difference between the two kinds of explanations is that horizontal explanations explain an event's occurrence by referring to a sequence of events that precede it, while vertical explanations explain a phenomenon synchronically by referring to its components and their relations. Folk psychological explanations are horizontal, because they appeal to sequences of mental events that precede the behavior to be explained. Cognitive psychological explanations are vertical, because a complex capacity is decomposed in simpler components, via functional analysis (Cummins, 1975). This does not mean that there are no valuable horizontal explanations of behavior using belief-desire psychology (or more sophisticated elaborations on the belief-desire model, such as one finds in certain areas of social psychology, for example). However, not all explanatorily useful psychological categories are representational kinds. Talk of "representational kinds" occurs and is justified only when explanation takes place vertically, and projection of properties of representations (at least potentially) crosses explanatory levels.[10]

What is the relevant notion of levels in multi-level explanations? There are many accounts of levels (e.g., Dennett, 1978; Marr, 1982; Pylyshyn, 1984), with Marr's being the most influential. According to Marr, we can study a cognitive system at three levels (and a "complete explanation" of a cognitive phenomenon spans all three):

[10] An example is provided by the literature on object perception and cognition (Carey & Xu, 2001; Scholl, 2001), which led to the postulation of the representational kind 'object files'. Similar effects were observed both in adults and in infants, such as similar set-size capacity limits, or similar prioritization of spatio-temporal over featural information, leading to adults' and infants' object representations being "lumped" together into a common kind. Carey & Xu argue that "[y]oung infants' object representations are the same natural kind as the object files of mid-level vision [in adults]" (Carey and Xu, 2001, p. 210).

1. At the **computational level**, the task of the system is identified and redescribed in terms of the information-processing problem that the system has to solve, and the constraints that the system has to satisfy to perform this task are identified (this level is concerned with *what* the system does and *why* it does it).
2. At the **algorithmic-representational level** the transformation rules (algorithms) and the syntactic properties of the representations that accomplish the task identified at level 1 are specified (including the format of the representations that are the inputs, the intermediaries, and the outputs of the processes).
3. At the **implementational level** the physical structures (neural mechanisms) that realize the processes described at level 2 and the neural localization are specified.

Marr's account of levels has been challenged: for example, it has been argued that these are not special levels of explanation, but merely heuristic clusters of questions (McClamrock, 1991), that there are intermediate levels between the computational and the algorithmic levels (Griffiths, Lieder, & Goodman, 2015), or that the implementational level is not as autonomous as the original account made it out to be (Kaplan, 2017). There is also a debate about what exactly Marr's levels are: are they levels of abstraction, levels of organization, levels of analysis, or levels of realization (Craver, 2014)?

An exhaustive overview of Marr's levels and their different interpretations is a topic for another paper (or book). Here we endorse the view that these are levels of explanation, and that they are identified by the type of questions cognitive scientists might ask in their investigation. We side with the view that Marr's computational level is specified with reference to intentional representational contents (which makes it similar to Pylyshyn's semantic level).[11] On the other hand, we do not assume that there are only three levels: our account is open to the inclusion of further levels in addition to the ones above.

What matters to our purpose is the close relation between levels and kinds. In many domains, it is fruitful to study cognitive systems as multi-level when trying to uncover whether they are natural kinds. For example, this is the strategy used by Michaelian (2011) for exploring whether memory is a natural kind. For memory systems this implies asking three questions:

> a) is there an information-processing task common to the relevant memory systems? (computational level)
>
> b) is there a procedure for performing that task common to the systems? (algorithmic-representational level)
>
> c) is there an implementation of the procedure common to the systems? (implementational level)

[11] In the debate on whether Marr's computational level refers to representations and their contents, or to a mathematically specified function (Shagrir, 2010).

A lack of unity at each of the levels, and an impossibility of projecting across levels, implies that the memory system is not a cognitive natural kind (if Michaelian is right). Though this example concerns a cognitive system, it highlights an important feature of inquiry in cognitive science more generally: natural kinds in cognitive science are plausibly necessarily multi-level. Similarly, for representational kinds, clusters of non-accidentally co-occurring properties need to be discovered at multiple levels in order for there to be kindhood.

As previously mentioned, representations have a content and a vehicle with syntactic and physical properties. These properties of representations entail that they can be studied at different levels. The alignment between properties of representations and levels might not be perfect, but it provides a heuristic for the discovery of representational kinds. At the computational level we're interested in finding out what the representations represent, and how their contents are connected with other mental states and with behaviors. For example, at the computational level the action detection system's aim is to detect actions in the environment. At the algorithmic-representational level, scientists discover the features of the processes that the representations enter into, as well as some properties of the vehicles, such as their format. Here, for example, scientists bring to light the processes implicated in the differentiation between actions and non-animated movements, as well as the format of action representations. At the implementational level, scientists look for the features of the neural structures underlying action-representations, as well as for their localization in the brain.

Our claim is that for a representational category to constitute a kind, it must allow for cross-level projection of properties. This strongly suggests that the representations correspond to a joint in (cognitive) nature: they have more in common than mere content, and further resemblance between them is unlikely to be accidental. To return to the example we used earlier, the semantic syndrome of action-representations provides a first step in the positing of a potential representational kind: there is evidence for domain-specific detection of stimuli that involve visible movements. But action-representations also show effects that are due to non-semantic properties of the representations involved, which we reviewed in section 1.

At this point one could object (as one anonymous reviewer does) that there is no real contrast between wombat-representations and action-representations, because one can always find a cluster of properties at different levels, even for wombat-representations. As a result, our account over-generates. This objection can be formulated in two ways. In the first version of the objection, when I deploy a wombat-representation, there are some activations in my brain, as well as some psychological processes (such as categorization) that occur at the same time. For example, I can imagine a picture of a wombat and I can think "wombats are cute". In both of these cases we can find syntactic and implementational properties. Why not say that they cluster together to form the (generic) kind "wombat-representation"?

We contend that in this case the features would be too heterogeneous for natural kindhood: very few shared features would be found at the algorithmic-representational and implementational levels. Indeed, such a proposal would be subject to the same arguments Machery uses against prototypes, exemplars, and theories belonging to the same natural representational kind "concept" (Machery, 2009).

In the second version of the objection, the focus is only on wombat-representations that are prototypes (a similar argument could be formulated for exemplars or theories). In this case, there are certainly properties that cluster together at the algorithmic-representational and implementational levels. Yet, in this case the properties associated with wombat-prototypes are also shared with other "basic level categories" (basic for thinkers familiar with the Australian fauna), such as kiwi-prototypes, kangaroo-prototypes, and so on. The representational kind here is not wombat-representation, but "prototype category-representation", of which wombat-prototypes are a sub-category, similarly to how Dalmatian, Pomeranian and Chow Chow are all different breeds of the same kind "dog".

By contrast, action-representations show a significant clustering of properties at different levels: intentional properties at the semantic level, special format and effects at the syntactic level, and characteristic patterns of brain activations at the implementational level. Moreover, this cluster of features is also distinct from other representations in the vicinity, such as representations of non-action movements, including random spatio-temporal trajectories (see Smortchkova, 2018).

Thus, contrary to wombat-representations, action-representations in cognitive science are good candidates for kindhood. They exhibit commonalities which allow them to be identified at various levels of inquiry. They form a kind not merely on the basis of their intentional content, but on the basis of the discovery of psychological effects and physical-implementational properties, which provide evidence of multi-level resemblance between representations of the relevant sort – in other words, they exhibit the relevant sort of "depth" which justifies talk of "kinds", despite the absence of a classical essence.

5. How many levels of explanation are needed?

A strong version of the multi-level thesis would require non-accidental clustering of properties across *all* levels, i.e. that the kinds go "all the way down" to the implementational level. One reason to endorse such a view would be if one conceived of the algorithmic-representational level as a mere "sketch" of what is going on at the implementational level, and insisted that functional constituents of a system have to line up with spatiotemporal ones (Piccinini & Craver, 2011). In this section we would like to explore how some representational kinds could be unified at some (more than one) levels without being necessarily unified at all levels.

The first way in which not all levels might be needed is tied to the standard version of the multiple realization thesis (Putnam, 1967): a kind that is unified at the computational and algorithmic-representational levels might not be unified at the implementational level. The most obvious reason to reject the all-level version of the multi-level thesis is thus to allow for representational kinds to be genuinely psychological, as opposed to necessarily neuropsychological, i.e. multiply realized at the implementational level.

A second way in which a representational kind could be multi-level without encompassing all levels would be for the kind to exhibit what we could call "inverse multiple-realizability": a category of vehicles is inversely multiply realized if it forms a kind at the algorithmic-representational and implementational levels, yet has multiple

different types of contents. Such a kind would be unified at the algorithmic-representational level, as well as in terms of the implementational mechanisms that realize the relevant algorithmic function; yet at the computational level, there would be few or no semantic properties in common between its members. There is clearly room in logical – and indeed empirical space for such a kind.

One potential example is "address addressable representations" and "content addressable representations" used in information processing approaches inspired by computer science (Bechtel & Abrahamsen, 2002; Gallistel & King, 2011). "Address-addressable representations" are used to retrieve mnemonic contents stored at different locations (addresses) via indexes that (non-semantically) identify the addresses. "Content-addressable representations" search for storage locations in memory by probing them with partial contents, that are used to retrieve the rest of the stored information. When applied to explanations of human cognition, these representations would acquire their kindhood status if properties that enabled the brain to realize these particular sorts of computational operations were discovered at the implementational level. What would be stored at the various memory addresses would, by hypothesis, be representations, endowed with content. However, there might be nothing in common to the different contents of the representations within the (broad) class of "address-addressable memory representations". By analogy with typical examples of multiple realizations, we would find the address-addressable representations to have a fairly open-ended disjunction of semantic properties – though what enables them to all be addressable in the same way might be empirically discovered to be some similar brain mechanism. For such a kind, semantic content, though present, is merely "quantified over" rather than made reference to – it does not play a role in unifying the kind. If one adopts an approach to the brain strongly driven by concepts from computational theory, *à la* Gallistel and King, one should expect – indeed hope – to find such kinds.

It is also worth noting that representational kinds are sometimes identified at the computational and implementational levels without reference to the intermediary algorithmic-representational level. This seems to be an approach that is commonly used in some areas of neuroscience to find representational kinds. When studying how the brain represents types of objects, the methodology is often based on differential activation of brain regions: for example, regions in the face fusiform area (FFA) are said to represent faces because they are active when face stimuli are present, and not when other kinds of stimuli are present, such as tools (Grill-Spector, 2003). Prima facie this practice suggests that kinds could be identified on the basis of content and implementation alone, with little or no regard for "syntax".

One could object that, in such cases, the gap at the algorithmic-representational level is only *epistemic* – it reflects temporary ignorance, or division of labor between neuroscientists and psychologists, rather than the true nature of the kind. According to such a view, in reality, face-representations will be known to constitute a representational kind only when cognitive scientists also discover shared algorithms underlying face perception, and other syntactic properties of face-representations.

However, there is a stronger thesis that would suggest that we can bypass the algorithmic-representational level altogether and identify representational kinds via their semantic content and their implementational features only: such a thesis might be taken to follow from a version of the semantic view of computation in cognitive

science recently defended by Rescorla (2017). According to Rescorla's view computation individuation proceeds via semantic properties along with non-representational implementational mechanisms (p. 302) without any contribution from formal syntactic properties. Thus, Rescorla writes that "[w]e can model the mind computationally without postulating formal mental syntax" (p. 281).

Views about computation individuation do not necessarily apply to accounts of computational explanation and the claim that we should do without mental syntax altogether is obviously highly controversial. Therefore, Rescorla's view may or may not be true of computational explanation in cognitive science in general. But the view certainly does not seem downright contradictory. As a result, it strikes us as at least coherent to suppose that some explanations could in principle take place only at the computational and implementational levels. If that is the case, it would behoove us to make room for multi-level kinds that span levels 1 and 3 of the tri-level approach: if there is no need for level 2, then level 2 properties are not required for kindhood. While we do not wish to endorse Rescorla's view, the important point for our purposes is that it is compatible with ours: it leaves the room for some multi-level representational kinds that do not span all levels.

Finally, we do not claim that one of the levels is more basic, in the sense of having to play the role of the sustaining mechanism for the syndrome of properties at other levels. For example, we are not claiming that the implementational level or the algorithmic-representational level is where one finds the mechanism sustaining the clustering of semantic properties at the computational level. Indeed, the sustaining mechanisms might be extrinsic to the representational kind. The most obvious argument for such a view derives from *(meta)semantic externalism*, the view according to which what determines the content of a mental representation are causal-historical relations to one's environment or community (Deutsch & Lau, 2014). For instance, according to teleosemantic approaches to intentional content, the mechanism which determines the content of mental representations is at least partly historical (Millikan, 1984). Our multi-level view of representational kindhood is entirely compatible with externalism about content.

What matters for representational kindhood is the existence of robust clustering of properties and cross-level projections, which enable the representational kind to support many generalizations and predictions. Such multi-levelness is sufficient, in our view, to support the claim that the clustering is non-accidental. The multi-level thesis for kindhood need not include the claim that one of the lower levels is the sustaining mechanism for the cluster of properties that are exhibited by the kind. Even if research is guided by the tacit assumption that there is some sustaining mechanism for the kind, such a mechanism might be extrinsic.

There might also be a plurality of mechanisms. This plurality might be synchronic: for instance, if a two-factor view of content determination were correct (Block, 1986) the semantic level would be partially sustained by internal causes/functional role, and partially by external environmental causes. Note that not all aspects of functional role have to be equally content-determining: incidental effects may be incidental also in the sense of being metasemantically inert. A plurality of cluster-sustaining mechanisms may also be involved diachronically: it is compatible with Boyd's view that different

mechanisms play the role at different times, or on different time-scales (e.g. ontogenetically and phylogenetically). [12]

6. Objections and replies

In this section we consider some objections to our account, and offer replies that we hope will help clarify the proposal.

Objection 1: Some representational kinds in cognitive science are single level. For instance, some are *purely* semantic (e.g. singular thought, belief), or *purely* based on format or syntax (e.g. analog representations, or address-addressable memory representations).

Reply: Our reply to this objection is that, in reality, single-level kinds fall into two cases: either they are not kinds in cognitive science, or they are (secretly) multi-level.

To illustrate the first case, take "singular thought" as an example: to the extent that it is a merely semantic (or merely epistemological) kind, identified by a clustering of semantic or epistemic properties, it is not relevant to cognitive science. It becomes a candidate representational kind from a cognitive scientific perspective when one engages in the project of vertical explanation by looking at properties at other levels whose existence could support kindhood (Murez, Smortchkova, & Strickland, forthcoming).

To clarify our view on this point, it is worth stressing that we do not deny that some categories of entities may constitute kinds merely by virtue of sharing semantic properties. For instance, formal semanticists debate whether adverbs or quantifiers form natural kinds (see Balcerak Jackson, 2017, who defends an account of semantic kinds inspired by Evans, 1985). But these are, precisely, candidate *semantic* kinds, which are studied by linguists, not mental *representational* ones studied by cognitive scientists. The latter are those that are the primary targets of investigation in cognitive science, the science which studies mental representations.

What about solely format-based kinds? "Analog" picks out a candidate representational kind only to the extent that analog representations share interesting properties at other levels, in addition to their being analog. For example, analog representations play a role both in number cognition (Dehaene, 1997) and in mental modeling (Johnson-Laird, 2006) where they pick out entities in different domains. These different roles and domain-specificities might come with different properties at the computational and implementational levels, which might lead to the postulation of different representational kinds with analog format.

This brings us to the second reply to the objection. Many purportedly single level kinds might actually turn out to be multi-level, if one adopts a more accurate view of levels. Indeed, various philosophers and cognitive scientists have argued that the usual tripartition between levels is insufficiently fine-grained. For example, there might be levels between the computational and the algorithmic (Griffiths et al., 2015; Peacocke, 1986). Likewise, the semantic level might usefully divide into semantic (what is the

[12] We hope to further explore the question about the sustaining mechanism in future work.

content?) and metasemantic (what determines or grounds the content?) levels (Burgess & Sherman 2014). In some cases, the appearance of a single-level kind could be explained by an insufficiently fine-grained account of what constitutes a level. A kind could span multiple *sub*-levels, thereby acquiring the requisite sort of "depth".

Objection 2: Our proposal goes against the independence of the levels and multiple realizability.

Reply: The notion of "kindhood" we use relates kindhood to explanation in cognitive science. Our view is entirely compatible with a methodological or merely epistemic independence of levels: it might be useful, for various epistemic or methodological purposes, to abstract away from other levels in studying a representational kind, for example when the implementation is unknown, or when the algorithms haven't been yet discovered. Furthermore, as we made clear, we are not committed to the claim that kinds go "all the way down" (i.e. to a "tri-level" view of representational kinds). Nevertheless, to talk of "kinds" is to commit to there being more to the representations being studied than what is found at a single level. The view is compatible with representational kinds being multiply realizable: all it requires is that there be clusters of properties at different levels and inter-level *constraints* (a requirement commonly accepted in multi-level explanations).

Objection 3: Levels are not the only source of "inductive depth", history (evolutionary history) also plays a role. For example, Ereshefsky (2007) suggests that we should study psychological categories as homologies, and not only as functions or as adaptations.

Reply: We agree with this. As we noted, there are different possible views of the sustaining mechanism, that is the factor that unifies a kind and that causes the properties in the kind-syndrome to non-accidentally cluster together: intrinsic (e.g. a common neural mechanism) or extrinsic (e.g. a common developmental or evolutionary history). A representational kind might correspond to a grouping of representations that share properties in virtue of a common evolutionary ancestor, for example. This assumption is often implicit in cognitive science and guides research. Indeed, the discovery of essentialist representations in the cognition of adults and infants (Gelman, 2004) guides the search for the same representational kind in apes (Cacchione, Hrubesch, Call, & Rakoczy, 2016).

What is the role of historical factors for representational kinds? According to one view, some representational kinds are unified by extrinsic, historical factors. We could talk of such representational kinds having a "historical sustaining mechanism". Another possible option is that historical properties can be counted among those that non-accidentally cluster together to form a kind, without necessarily being the mechanism that causes the clustering. Yet another option would be that, if levels of explanation in cognitive science are properly understood, they would make room for a historical or evolutionary level. Our view of representational kinds is compatible with these different options for the role of history, which might each be true of different representational kinds.

Objection 4. What distinguishes representational kinds from other cognitive kinds? For instance, what distinguishes representational kinds from kinds of mental processes?

Reply: This issue intersects with the more general issue "What is a mental representation?". We will not attempt a general definition. We assume that representations have semantic properties, but also non-semantic ones. The semantic properties are "substantial" to the extent that e.g. the distinction between representational correctness and error plays a not easily eliminated explanatory role. A kind is not representational if it makes no reference to semantic properties. For instance, a classification of brain states according to merely physiological criteria might count as a cognitive scientific kind, but not as a representational one.

Any functional kind will allow for some sort of multi-level approach: one can distinguish role-level investigation, and realizers/implementational level. However, the division into levels introduced by Marr in the context of a computational theory of vision are (arguably) specific to – or at least paradigmatic of – representational explanation. Hence, it might be suggested that the sort of multi-level approach we describe is indeed specific to representational kinds, as opposed to other sorts of cognitive kinds.

Objection 5: Can't representational kinds be unified/individuated merely by the sorts of processes that operate on those representations? For instance, suppose you have an independent notion of what makes a process perceptual, as opposed to cognitive. Might one not then introduce a kind "perceptual representations" that would not satisfy our criteria?

Reply: Representations play a role within processes, including computational-inferential processes, and perhaps non-computational and non-inferential processes (such as, arguably, associations)[13]. In general, there is a close connection between representations and processes. In such cases, a kind of representation is picked out by reference to a kind of process (or to a component of cognitive architecture). Yet, kinds of processes and kinds of representations can be orthogonal. In some cases, classification of representations into kinds crosscuts processing, for instance when one entertains the possibility that visual imagery manipulates the same representations as visual perception. In other cases, the same process might make use of different kinds of representations, for example if perceptual processes use both iconic and discursive representations (Quilty-Dunn, forthcoming).

According to our requirement of depth, for there to be a genuinely representational kind, there must be more in common to the representations in question than merely that they are processed in the same way. For instance, "visual representation" might pick out a representational kind if it turns out that such representations tend (non accidentally) to share format-properties (e.g. they tend to be iconic) and effects at deeper levels. On this view, while "visual representation" may pick out a useful class of

[13] An anonymous reviewer challenges this claim by remarking that associations can be computational if they occur in artificial neural networks, and that their non-inferential status depends on the notion of inference one endorses. Here we accept the standard contrast between associative transitions and computational transitions in the computational theory of mind. See Mandelbaum (2015) for a review of the contrast.

representations, it only picks out a genuinely representational kind if the representations non-accidentally share properties at multiple levels.

The fact that representations participate in processes and are partially identified by the role they play in these doesn't undermine the existence of representational kinds, just like the fact that, in the biological domain, mitochondria participate in the process of cellular respiration doesn't diminish their status as a potential natural kind.

A related objection (due to an anonymous reviewer) states that appealing to processes in the identification of representational kinds runs the risk of over-generating. All representations that can be stored in a limited working memory would be of the same kind ("working memory representations"). This suggests that the properties that make representational categories kinds, are not properties of representations, but properties of the system that processes representations. If we could explain all features of representations by the type of processes to which they belong, then we wouldn't have representational kinds, but only process kinds. Our reply above also applies to this objection: Storage in working memory is only one of the features of a genuine representational kind.

Objection 6. Semantic content of the sort that matters to cognitive science is internalistic, and determined by syntax. Externalistic intentional content is largely irrelevant. What you call a "semantic syndrome" plays no role in representational kinds' individuation.

Reply: This objection appeals to an internalist account of mental representations that has been proposed (among others) by Egan (2013), who distinguishes representations with "mathematical content" from representations with cognitive content (content that refers to the environment). Representations with mathematical content are characterized functionally by the role they play in the computational account of a cognitive capacity. This role is specified in mathematical terms and does not make reference to worldly entities. Yet mathematical representations are representational because they can misrepresent, relative to the 'success criteria' of the cognitive capacity that has to be explained. Egan uses as an example Marr's description of the stage of early visual processing which takes light intensity values at different parts of the image as inputs and gives as outputs calculations of the change rate of the intensities in the image. This representation could be applied to an environment where light behaves very differently from ours (Egan, 2013, p. 122).

Regardless of whether Egan's account of "mathematical content" is correct, Egan's representational kinds are, in fact, multi-level in our sense, because in order for the mathematical description of the capacity to explain cognition, it has to be mapped onto the physical realizer via a realization function, such that transitions in the physical system mirror transitions between symbols in the mathematical computation (Egan, 1992).

Furthermore, though we are sympathetic to the claim that Egan's account applies to some representational kinds (but see Sprevak (2010) for a critical discussion), it strikes us as implausible that it will extend to *all* representational kinds studied in cognitive science. Action-representations are first identified by their worldly extension, and the fact that they refer to actions plays a central role in the study of this representational

kind. For many representational kinds, an externalist account of content (Burge, 1979) seems to be more adequate.

7. Conclusion: the notion of maximal representational kind.

To conclude, we would like to suggest that cognitive scientists operate with a (tacit) *ideal* of what a representational kind looks like – a *maximal* representational kind. A maximal representational kind is a category of representations that are maximally "deep", in that they non-accidentally share properties across all levels of description. Representational kindhood can come in degrees, to the extent that an actual class of representations can match more or less closely to this ideal of a maximal kind. We think that the notion of a maximal kind can be viewed as a heuristic device that guides research in cognitive science: to hypothesize that a certain class of representations forms a kind is to be disposed to investigate other properties that co-cluster with those that have been used to initially pick out members of the kind in the process of vertically decomposing the capacity to be explained.

There are still many details to fill in to fully develop an account of representational kinds. In this chapter we have sketched the main outline of a proposal that we hope to expand and refine in the future.

Acknowledgements

We are grateful to audiences in Bochum, London, Nottingham, and Oxford for insightful comments on previous versions of the chapter. In particular, we thank the participants of the Philosophy of Mind Work in Progress group in Oxford (Umut Baysan, Dominic Alford-Duguid, Sam Clarke, Anil Gomes, Jake Quilty-Dunn, Nick Shea, and Margot Strohminger) for helpful discussion of a previous version of this paper, two anonymous reviewers for insightful comments, and Krys Dolega for a last round of comments. Joulia Smortchkova's work on this chapter has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No. 681422.

References

Atran, S. (2002). *In Gods We Trust: The Evolutionary Landscape of Religion*: Oxford University Press.

Bechtel, W., & Abrahamsen, A. (2002). *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks*: Blackwell Publishing.

Beck, J. (2015). Analog Magnitude Representations: A Philosophical Introduction. *British Journal for the Philosophy of Science, 66*(4), 829-855.

Block, N. (1981). Psychologism and behaviorism. *Philosophical Review, 90*(1), 5-43.

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy, 10*(1), 615-678.

Boyd, R. (1991). Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies, 61*(1), 127-148.

Boyd, R. (1999). Kinds, Complexity and Multiple realization. *Philosophical Studies, 95*(1), 67-98.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy, 4*(1), 73-122.

Cacchione, T., Hrubesch, C., Call, J., & Rakoczy, H. (2016). Are apes essentialists? Scope and limits of psychological essentialism in great apes. *Animal cognition, 19*(5), 921-937.
Carey, S. (2009). *The Origin of Concepts*: Oxford University Press.
Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition, 80*(1-2), 179-213.
Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science, 19*(1), 51-57.
Craver, C. F. (2014). *Levels*: Open MIND. Frankfurt am Main: MIND Group.
Cummins, R. C. (1975). Functional analysis. *Journal of Philosophy, 72*(November), 741-764.
Cummins, R. C. (1983). *The Nature of Psychological Explanation* (Vol. 19): MIT Press.
Cummins, R. C. (2000). "How does it work" versus "what are the laws?": Two conceptions of psychological explanation. In F. Keil & R. A. Wilson (Eds.), *Explanation and Cognition, 117-145*: MIT Press.
Cummins, R. C. (2010). *The World in the Head*: Oxford University Press.
Dehaene, S. (1997). *The number sense*. New York Oxford University Press.
Dennett, D. C. (1978). *Brainstorms*: MIT Press.
Dennett, D. C. (1994). Get Real. *Philosophical Topics, 22*(1/2), 505-568.
Deutsch, M., & Lau, J. (2014). Externalism about Mental Content. *The Stanford Encyclopedia of Philosopshy*.
Devitt, M. (2008). Resurrecting biological essentialism. *Philosophy of Science, 75*(3), 344-382.
Drayson, Z. (2012). The uses and abuses of the personal/subpersonal distinction. *Philosophical Perspectives, 26*(1), 1-18.
Dupre, J. (1981). Natural Kinds and Biological Taxa. *The Philosophical Review, 90*(1), 66-90.
Egan, F. (2013). How to think about mental content. *Philosophical Studies*(1), 1-21.
Ereshefsky, M. (2007). Psychological categories as homologies: lessons from ethology. *Biology and Philosophy, 22*(5), 659-674.
Ereshefsky, M., & Reydon, T. A. C. (2015). Scientific kinds. *Philosophical Studies, 172*(4), 969-986.
Fodor, J. A. (1974). Special sciences. *Synthese, 28*(2), 97-115.
Fodor, J. A. (1975). *The Language of Thought* (Vol. 87): Harvard University Press.
Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (2002). 17 Action representation and the inferior parietal lobule.
Gallistel, C. R., & King, A. P. (2011). *Memory and the computational brain: Why cognitive science will transform neuroscience* (Vol. 6): John Wiley & Sons.
Gelman, S. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences, 8*(9), 404-409.
Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science, 7*(2), 217-229.
Grill-Spector, K. (2003). The neural basis of object perception. *Current opinion in neurobiology, 13*(2), 159-166.
Jackson, F., & Pettit, P. (1990). In defense of folk psychology. *Philosophical Studies, 59*(1), 31-54.
Jellema, T., & Perrett, D. I. (2007). Neural pathways of social cognition. *Dunbar and Barrett, Oxford Handbook*, 163-177.
Johnson-Laird, P. N. (2006). *How we reason*: Oxford University Press, USA.

Jordan, K. E., & Brannon, E. M. (2006). A common representational system governed by Weber's law: Nonverbal numerical similarity judgments in 6-year-olds and rhesus macaques. *Journal of experimental child psychology, 95*(3), 215-229.
Kaplan, D. M. (2017). *Integrating Mind and Brain Science: A Field Guide*: Oxford University Press.
Khalidi, M. A. (2016). Innateness as a natural cognitive kind. *Philosophical Psychology, 29*(3), 319-333.
Kripke, S. (1980). *Naming and Necessity* (Vol. 217): Harvard University Press.
Lafraire, J., Rioux, C., Giboreau, A., & Picard, D. (2016). Food rejections in children: Cognitive and social/environmental factors involved in food neophobia and picky/fussy eating behavior. *Appetite, 96*, 347-357.
Lycan, W. G. (1988). *Judgement and Justification* (Vol. 100): Cambridge University Press.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1-25.
Machery, E. (2009). *Doing Without Concepts*: Oxford University Press.
Mandelbaum, E. (2015). Associationist Theories of Thought *Stanford Encyclopedia of Philosophy*.
Marr, D. (1982). *Vision: A computational investigation into*: WH Freeman.
McClamrock, R. (1991). Marr's three levels: A re-evaluation. *Minds and Machines, 1*(2), 185-196.
Michaelian, K. (2011). Is memory a natural kind? *Memory Studies, 4*(2), 170-189.
Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*: MIT press.
Millikan, R. G. (1999). Historical kinds and the "special sciences". *Philosophical Studies, 95*(1-2), 45-65.
Murez, M., Smortchkova, J., & Strickland, B. (forthcoming). The Mental Files Theory of Singular Thought: A Psychological Perspective. In R. Goodman, J. Genone, & N. Kroll (Eds.), *Mental Files and Singular Thought*. Oxford: Oxford University Press.
Needham, P. (2002). The discovery that water is H2O. *International Studies in the Philosophy of Science, 16*(3), 205 – 226.
Peacocke, C. (1986). Explanation in computational psychology: Language, perception and level. *Mind and Language, 1*(2), 101-123.
Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese, 183*(3), 283-311.
Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science, 7*, 131-193.
Putnam, H. (1967). Psychological predicates. *Art, mind, and religion, 1*, 37-48.
Pylyshyn, Z. W. (1984). *Computation and cognition*: MIT press Cambridge, MA.
Quilty-Dunn, J. (forthcoming). Perceptual Pluralism. *Noûs*.
Quine, W. V. (1969). *Ontological Relativity and Other Essays*: Columbia University Press.
Ramsey, W. M. (2007). *Representation reconsidered*: Cambridge University Press.
Rescorla, M. (2017). From Ockham to Turing–and Back Again *Philosophical Explorations of the Legacy of Alan Turing* (pp. 279-304): Springer.
Rizzolatti, G., & Sinigaglia, C. (2007). Mirror neurons and motor intentionality. *Functional neurology, 22*(4), 205.
Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80*(1-2), 1-46.

Shagrir, O. (2010). Marr on computational-level theories. *Philosophy of Science, 77*(4), 477-500.
Shea, N. (2007). Content and Its vehicles in connectionist systems. *Mind and Language, 22*(3), 246–269.
Shea, N. (2018). *Representation in cognitive science*: Oxford University Press.
Shiffrar, M. (2008). Embodied motion perception: Psychophysical studies of the factors defining visual sensitivity to self-and other-generated actions. *Embodiment, ego-space, and action*, 113-143.
Shutts, K., Condry, K. F., Santos, L. R., & Spelke, E. S. (2009). Core knowledge and its limits: The domain of food. *Cognition, 112*(1), 120-140.
Smortchkova, J. (2018). Seeing goal-directedness: a case for social perception. *The British Journal for the Philosophy of Science*.
Sprevak, M. (2010). Computation, individuation, and the received view on representation. *Studies in History and Philosophy of Science Part A, 41*(3), 260-270.
Thomson, E., & Piccinini, G. (2018). Neural Representations Observed. *Minds and Machines, 28*(1), 191-235.
Widhalm, M. L., & Rose, N. S. (2019). How can transcranial magnetic stimulation be used to causally manipulate memory representations in the human brain? *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(1), e1469.
Wood, J. N. (2007). Visual working memory for observed actions. *Journal of Experimental Psychology: General, 136*(4), 639.