



# Cognitive ontology and the search for neural mechanisms: three foundational problems

Jolien C. Francken<sup>1</sup> · Marc Slors<sup>1</sup> · Carl F. Craver<sup>2,3</sup>

Received: 8 October 2021 / Accepted: 15 April 2022 / Published online: 7 September 2022  
© The Author(s) 2022

## Abstract

The central task of cognitive neuroscience to map cognitive capacities to neural mechanisms faces three interlocking conceptual problems that together frame the problem of cognitive ontology. First, they must establish which tasks elicit which cognitive capacities, and specifically when different tasks elicit the same capacity. To address this *operationalization problem*, scientists often assess whether the tasks engage the same neural mechanisms. But to determine whether mechanisms are of the same or different kinds, we need to solve the *abstraction problem* by determining which mechanistic differences are and are not relevant, and also the *boundary problem* by distinguishing the mechanism from its background conditions. Solving these problems, in turn, requires understanding how cognitive capacities are elicited in tasks. These three problems, which have been noted and discussed elsewhere in the literature, together form a ‘cycle of kinds’ that frames the central problem-space of cognitive ontology. We describe this cycle to clarify the intellectual challenges facing the cognitive ontologist and to reveal the kind of iterative process by which ontological revision in cognitive neuroscience is likely to unfold.

**Keywords** Kinds · Neuroscience · Cognition · Operationalization · Philosophy · Task

---

✉ Jolien C. Francken  
jolien.francken@ru.nl

✉ Marc Slors  
marc.slors@ru.nl

Carl F. Craver  
ccraver@wustl.edu

<sup>1</sup> Faculty of Philosophy, Theology and Religious Studies, Radboud University, P.O. Box 9103, 6500 HD Nijmegen, Netherlands

<sup>2</sup> Department of Philosophy, Washington University in St. Louis, 63150 St. Louis, MO, USA

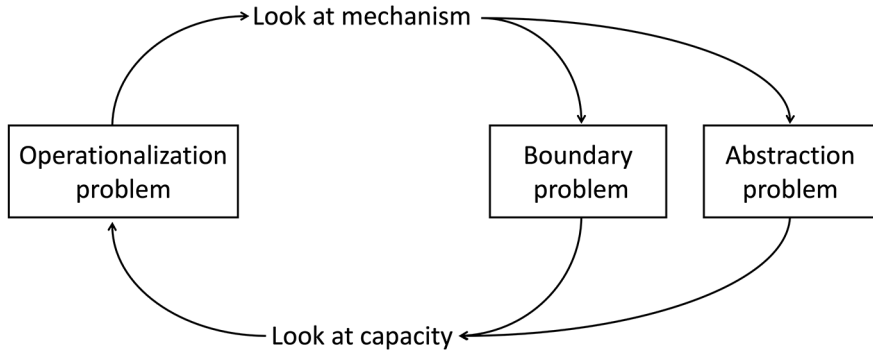
<sup>3</sup> Philosophy-Neuroscience-Psychology Program Wilson Hall, Washington University in St. Louis, 63105 St. Louis, MO, USA

## 1 Introduction

A fundamental goal of cognitive neuroscience is to explain the cognitive capacities that collectively enable humans to live their complex lives. To achieve this goal, we must answer two basic questions: (1) What are these capacities? And (2) How do facts about brains explain them? In the past decade, there has been a tremendous upsurge of work on both these questions, resulting in an exciting new, interdisciplinary field of Cognitive Ontology, dedicated to regimenting the scientific taxonomy of cognitive capacities (see e.g. Poldrack et al.'s (2011) compelling case for the Cognitive Atlas; Lenartowicz et al.'s (2010) case study on the ontology of cognitive control). Here, we describe three core problems any such regimentation must solve. These problems have been identified individually by ourselves and others before. We show that these three problems require simultaneous solutions, as the solution to any one of them constrains and helps to frame what the acceptable solutions to the others must be. We think the resulting circularity speaks strongly against a revisionist bottom-up, brain-driven reform and in favor of careful attention to the behaviors and tasks that frame the mechanisms bottom-up approaches privilege in their ontological reforms. The bottom-up approach, we argue, reflects a scientifically untenable and incomplete understanding of the inferential constraints driving conceptual development in cognitive neuroscience.

In answering (2), a prevailing idea in both science and philosophy is that the brain has these definitive capacities in virtue of containing mechanisms that underlie (or mediate, or implement) those capacities (Craver, 2007; Craver & Darden, 2001; Bechtel, 2008; Piccinini, 2020). If capacities are defined in functional terms, for example as relating input to output, the mechanism for a capacity involves the causally organized interactions of entities and activities in virtue of which the input is transformed into output. Such mechanistic explanations frequently span multiple levels of organization: The activities and interactions composing a neural mechanism are themselves explained by lower-level mechanisms.

Question (1) asks for a taxonomy of cognitive capacities, which is the main focus of research on cognitive ontology. Fierce controversies arise both locally and globally with some frequency over how to define cognitive capacities and how to distinguish them from each other (e.g. Anderson 2015; Colaço 2018; Colaço 2020; Danziger 1997; Feest 2017; Janssen et al., 2017; Khalidi, 2017; Klein, 2012; Poldrack, 2010; Poldrack and Yarkoni 2016; Price and Friston 2005; Sullivan 2017; Uttal 2001). Characterizations of cognitive capacities often take the form of functional, 'causal role' definitions. For example, memory can be defined, broadly, as "experience-dependent modification of internal structure, in a stimulus-specific manner that alters the way the system will respond to stimuli in the future as a function of its past" (Baluška and Levin 2016, p. 2). If so, questions about the reality and/or delineation of such cognitive capacities may be answerable by looking at the 'realization base' of such concepts—that is, by invoking mechanistic answers to (2). For example, given the broad definition of memory mentioned above, the question whether memory is transferable after transplants is answered by researchers from the laboratory of Glanzman by referring to the "transplant" of sensitization via the transfer of RNA in *aplysia* sea slugs, a model organism (Bédécarrats et al. 2018). The legitimate ontology of cogni-



**Fig. 1** The Cycle of Kinds: An overview of the three problems and their interrelation for the mechanistic anchoring approach to cognitive ontology

tive capacities is hence thought to correspond to the correct catalogue of identifiable neural mechanisms. Facts about neural mechanisms are thus supposed to anchor facts about cognitive ontology.

This picture of the “mechanistic anchoring approach” to cognitive ontology is idealized. We argue that just as the delineation of existing cognitive capacities depends on the mechanisms that are implemented by the brain, the delineation of what counts as the relevant mechanisms depends on what we take these capacities to be. In order to see this circularity, we concentrate on three problems faced by the mechanistic anchoring approach: the Operationalization Problem (Sect. 2), the Abstraction Problem (Sect. 3), and the Boundary Problem (Sect. 4). None of these problems is new (neither are these the only problems the approach faces—see footnotes 1 and 5). What has not previously been noted is how closely interconnected these problems are or why they require simultaneous solutions. Together, they form what we will call “the Cycle of Kinds,” depicted in Fig. 1. The Operationalization Problem concerns a principled uncertainty about how specific experimental tasks correspond to the cognitive phenomena they are used to study. This uncertainty is frequently addressed by assessing whether the tasks engage the same or different neural mechanisms. However, we argue, the mechanistic structure of the world is not simply perceived as such but requires theoretical reconstruction to be discovered. To identify mechanisms as such, we need to abstract away from the buzzing blooming confusing of causal connections (the Abstraction Problem) and distinguish constituents of mechanisms from e.g. background conditions (the Boundary Problem). These problems can be resolved, we argue only by having recourse to a prior understanding of what the relevant capacities are and of how they are elicited in cognitive tasks. Thus, we arrive back at the Operationalization Problem (Fig. 1). In Sect. 5, we argue this circularity is neither unique to cognitive (neuro)science nor especially deadly (Chang, 2004); instead, we should expect progress at this key interface to be incremental, iterative, and ultimately assessed globally for a system of interrelated concepts.

## 2 The operationalization problem

The Operationalization Problem is the problem of justifying the indicative relationship between a behavioral task and a cognitive capacity, that is, of justifying the claim that one or more tasks measure(s) the capacity it/they aim(s) to measure. In psychology, this question has been addressed before through the concept of “construct validity” (see Cronbach and Meehl 1995), but the success of this strategy has been debated. However, this discussion has been conducted primarily within psychology and is hence beyond the scope of this paper (Borsboom et al., 2009). In philosophy, on the other hand, the question has received widespread attention in recent years (e.g. Colaço 2020; Feest

2005, 2010, 2011, 2017, 2020, *forthcoming*; Colaço 2018; Sullivan, 2007, 2010, 2017, 2016, 2015; Irvine 2013; Stinson, 2016; McCaffrey & Machery, 2016). Uljana Feest, for example, studies the relationship between “operationism” and phenomena in the cognitive sciences and argues that operational (task) definitions are tools for conceptual development. Jacqueline Sullivan argues that experimental practices in at least some areas of (cognitive) neuroscience are insufficiently coordinated, hindering explanatory and taxonomic progress. Indeed, the Operationalization Problem is built into Poldrack’s effort at building an atlas of cognitive ontology (see Poldrack et al. 2011). Our delineation of the cycle of kinds helps to make explicit how these operational choices are constrained by, and constrain, other decisions about how we think about the mechanisms in the brain.

The Operationalization Problem is faced by any project linking (cognitive) capacities to (brain) mechanisms, either explicitly or implicitly. One cannot assess or establish such linkages without the use of tasks designed to elicit behavior reflective of the capacity in question. Such tasks provide the stimulus conditions and behavioral measures that allow one to interpret brain activities (or the absence thereof) in cognitive terms. In lesion studies, for example, subtly different measures of behavior are necessary to interpret the lesion as producing a cognitive deficit. Neurologists use clinical tasks to localize possible lesions via functional loss. Imaging studies use tasks and subtractions to activate (and localize) some capacities and not others. Comparative psychologists presume that the experimental organism is performing the same task (or a relatively similar task) as the ones that another species performs when it exhibits similar behaviors. The experimental task engages the subject in a behavior taken to indicate the operation (or absence) of the cognitive capacity. Tasks are thus indispensable anchors in this integrative project.

But how can we be sure that our task measures the capacity we think it measures? We use tasks because we cannot observe the capacities directly; we infer the capacity from task performance. In well-developed fields, this crucial choice is routinely taken for granted as a bit of the inherited practice one obtains in graduate training and post-doctoral research: We measure working memory with the n-back task or with the complex span test, and spatial memory is tested in a Morris water maze or in a scene recognition task. In new fields these tasks are more actively discussed. But it is open to question, in both instances, whether the task actually measures the capacity in question and, if so, how well it does so. Failures at this locus produce confounded experiments and conceptual confusion (Francken and Slors, 2014).

How must a task be related to a cognitive capacity for it to be used as an assay or measure of that capacity (Borsboom et al., 2004; Sullivan 2010)? In adopting a task, the researcher at least tacitly embraces a set of assumptions about how the stimulus elicits the capacity in question and how the capacity drives the task-related behavior. Call the collected set of these assumptions the *model of the task* performance. This model is based on a functional analysis of the capacity at issue, such as the analysis of memory mentioned in the introduction. While such an analysis describes in general terms what the capacity is supposed to do—how it transforms which inputs to what outputs—a task model describes in more or less detail how one thinks the more specific task conditions (stimulus and background conditions) are transformed into more specific task outputs (e.g., competent completion), revealing the stages and steps of a causal process/mechanism, perhaps associated with specific concrete structures and systems, one must traverse if one is to perform the task successfully.

A textbook example is the stop signal task (SST) which is routinely used as a measure of “response inhibition”. A coarse-grained functional analysis of response inhibition describes it, for example, as ‘the capacity to ‘intercept’ an upcoming action that is elicited by a given stimulus and to refrain from responding to the stimulus with that action.’ The family of stop signal tasks has been endlessly varied to study this cognitive capacity and disorders believed to involve deficits in response inhibition (Logan and Cowan 1984; Verbruggen & Logan 2008). It has also been varied to apply the task to adult, infant, and impaired humans, as well as to monkeys (Pani et al., 2018), rats (Eagle et al., 2008), and sheep (Knolle et al., 2017). In this task, a subject is instructed to perform an action (e.g., pressing a key when you see a face) unless a stop signal is presented prior to the moment of action. Researchers can vary the timing of the stop signal, for example, and determine the number of errors (in which the subject acts despite receiving the stop signal). Performance on the task is usually characterized in terms of an estimated value, called the *stop signal reaction time* (SSRT), which is taken to reflect the capacity to intercept responses—the less time we need to intercept an upcoming action, the better our response inhibition (but see Bissett et al., 2020). This capacity is thought by some to be involved in any cognitive act that requires volitional control over fleeting desires (e.g., going home after two glasses of wine or studying versus watching memes). Poor performance on such tasks is taken to indicate impulsivity, and is associated with attention deficit disorder and proneness to risk-taking and addictive behaviors (Dalley and Robbins 2017).

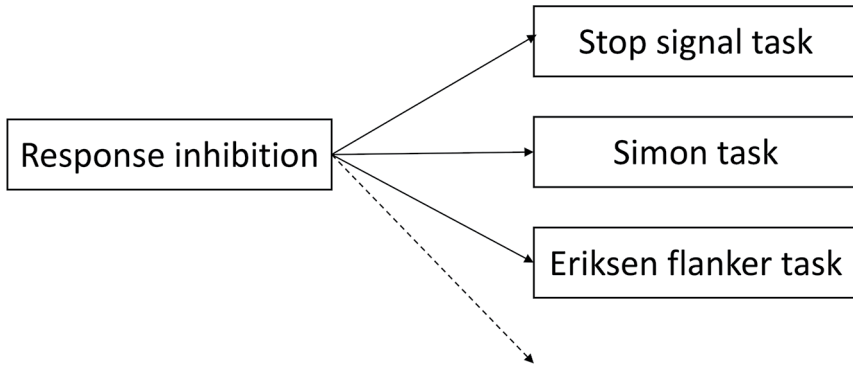
The precise task model for the stop signal family of tasks depends on its particular instantiation. It will include, minimally, the tendency of the subject to perform the dominant act (often acquired through pre-training), the ability to perceive the stop signal and to associate it with not acting (often acquired through training), and the ability to suppress the dominant behavioral tendency in light of that association. By shortening the time between the stop signal and the time for action, the go process (the process of responding to the stimulus by preparing the dominant action) finishes before the stop process (that is the process of intercepting the upcoming action response) and error frequency increases. It increases faster for those who are impaired in response inhibition. The ability to inhibit responses is both affected by the task conditions and influences response time, according to this model. That is why the task can be used to measure this capacity.

In perhaps the simplest kind of task model, the cognitive capacity in question is posited as necessary for task performance (or exhibiting the effect): Successful halting after the stop signal indicates that the response inhibition capacity is intact; subjects who require more time to stop are viewed as having less effective response inhibition. In this example of a task model, response inhibition is necessary to even perform the task, and the strength of that capacity is measured in the subject's probability of success. More generally, what matters is that the task conditions cause the cognitive capacity to be engaged and that the engagement of the capacity is a cause of the task output in such a way that the task output can be taken as an indicator of the capacity.

The Operationalization Problem arises because we cannot observe the cognitive capacity independently of our choice of task. To fully justify this choice, we would have to provide evidence for the fact that the cognitive capacity is necessary or otherwise involved in the performance of the task *independently* of the subject's performance on the task. We rarely (if ever) have such independent evidence; instead, the task itself is often taken, at least implicitly, as evidence for the capacity without independent justification.

When it is possible to use more than one task to measure a capacity, we might take some comfort in their consilience. But this comfort rests on having already decided that these tasks measure the same capacity; we also face the thorny question of when two tasks measure the same capacity (Francken and Slors, 2014). Typical tasks that are said to elicit response inhibition as well are the Simon task, the Stroop task and the Eriksen flanker task. The Simon task, to consider one example, involves trials where the spatial location of a stimulus (e.g., right) mismatches the spatial location of the required response, such as pressing a button, (e.g., left). Such mismatch typically yields slower responses, since we usually respond to things on the right with our right hands. In other words, the unusual combination of stimulus and response locations induces a conflict between those two drivers of action, and this requires time to resolve. The task model here assumes that participants will automatically start preparing a response on the side of the screen where the stimulus occurred, a preparation that needs to be interrupted when participants are instructed to respond by, e.g., pressing a button, on the other side of the screen. The SST (in its many variations) and the Simon task are supposed to engage the same cognitive capacity, response inhibition (Fig. 2). But it is open for discussion whether they measure *exactly* the same capacity or rather slightly different, related capacities. For example, the Simon task may also be described as involving decision making and response selection. This problem arises whenever researchers employ multiple tasks or make inferences from the behavior on one task to the behavior on another; and so it arises ubiquitously in cognitive science. The similarity among these tasks, however seriously these may be discussed in lab meetings and even in print, remains primarily intuitive and based, at least partly, on the phenomenological sense that these tasks require distinctive mental effort (Francken and Slors 2018). In short, to consider these tasks consilient requires a prior decision that they measure the same thing.

A similar problem arises in transferring tasks across species or subjects: Does the same task model for stop signal tasks in humans apply to sheep? Or might humans and sheep have different mechanisms for controlling action under these conditions?



**Fig. 2** Schematic depiction of the Operationalization Problem. Different tasks putatively elicit “the same” cognitive capacity. But whether they in fact do so is a further, empirical question. On the mechanistic anchoring approach, the question should be answered by determining whether or not the various tasks elicit the same neural mechanism

As has been noted for the study of the measurement of time (Tal, 2016) and temperature (Chang, 2004), we seem to face a circle of reasoning: We (at least often) are asked to justify a judgment of consilience on grounds that cannot be established independently of decisions about whether the tasks measure the same capacity.<sup>1</sup>

Historically, operationalists had a radical solution to this problem: They *defined* cognitive capacities in terms of their tasks. On the strictest interpretation, which quickly faded from philosophic currency (e.g., Bridgman 1927; see Chang 2021; Feest forthcoming), no two different tasks can possibly measure the same cognitive capacity. But most researchers want to retain the logical distinction between cognitive capacities and task performance, if only to allow for the possibility that one and the same cognitive capacity might be involved in different tasks, such as tasks performed daily outside the laboratory.

If a judgment is made that two tasks involve the same cognitive capacity, the task models involved must overlap, at least partly. Task models are functionally characterized, so the claim that two tasks involve the same capacity implies that it is assumed that the participants involved in these tasks engage neural mechanisms that realize or implement the capacities in the relevant task models. Ascertaining that these mechanisms are indeed involved in the execution of the tasks would indeed provide the behavior-independent evidence required to solve the operationalization problem (see above). Such “reverse inference” is widely, if controversially, deployed in cognitive neuroscience (Poldrack, 2006). Less controversial is the assumption that differences in neural activity observed during task performance might well indicate differences in cognitive processing. For example, McDermott et al., (2009) show that “laboratory tasks” for remembering, such as memorization of word lists, activate partly different regions of the brain than do memory tasks such as recalling a childhood experience,

<sup>1</sup> Another problem at play here is that there might well be cross-subject variation in the ways tasks are performed (Ward, 2019; Viola, 2021; Seghier & Price, 2018). To address this problem would further complicate our analysis and, in so doing, obfuscate the circular structure we want to emphasize here.

though both of these tasks are routinely used to assay the capacity of remembering. Likewise, differential effects of local damage on performance of different tasks indicates independence of the capacities in question (as has been argued for declarative and non-declarative memory, episodic and semantic memory, etc.). Indeed, these are common strategies for lumping and splitting kinds in mechanistic sciences generally.<sup>2</sup>

The mechanistic anchoring approach in general and these strategies in particular use neural mechanisms as objective arbiters of cognitive similarity and so as a basis for revising our cognitive ontology. As we show in the next section, however, the judgment that two mechanisms are identical or different cannot simply be read off the causal structure of things; any definitive judgment requires definitive solutions to both the Abstraction Problem and the Boundary Problem, which we now consider in turn.

### 3 The abstraction problem

In order to *delineate* a cognitive capacity, we must indicate when two particular instances of some capacity are instances of the same capacity-kind, just as to delineate a species one must indicate when two individuals belong to it.

In virtue of what does a given object or capacity belong to scientific kind? Daniel Dennett, for example, asks: What makes a given chunk of matter a magnet, or a particular kind of magnet, such as a ferromagnet (Dennett, 1987, p. 43). He discusses two kinds of answer. One is broadly “externalist” (or analytical functionalist): Two objects are of the same kind when they are disposed to act and interact with other things in the same ways. A magnet is a ferromagnet because it behaves like a magnet: It attracts ferrous materials, repels like poles, orients north to south, etc. On this externalist view, lodestones, ceramic magnets and electromagnets all belong to the same kind.

Dennett also considers a second, “internalist” or mechanistic, answer: Two objects are of the same kind when they have the same or similar organizations of components. Magnets are magnets in virtue of the fact that the spins on neighboring electrons are aligned in an exchange interaction. But since the way in which this alignment is induced is different in lodestones (which are possibly magnetized by lightning (Wasilewski and Kletetschka 1999), ceramic magnets (where heated ferrite powder is compacted in the presence of a magnetic field), and electro magnets (where the magnetic properties of many tiny electric currents are combined in a coil of isolated copper wire), the mechanistic outlook would not count them as belonging to the same kind. These two answers about magnets offer an intuitive starting point for any effort to regiment our ontology of cognitive capacities.

<sup>2</sup> This does not mean that this is the only scientifically respectable way of lumping and/or splitting cognitive capacity-kinds (for an example of kind-splitting of a biological function that is not based on mechanistic anchoring, see Bollhagen 2021). On the ideal picture of mechanistic anchoring, however, such splitting consists of finding distinct finer-grained functional analyses of what seemed like one functional concept at first. This is part of the operationalization process. Mechanists would expect these distinct, finer-grained functional concepts to correspond to distinct biological mechanisms.



When are two cognitive capacities instances of the same kind? As with magnets, the externalist—or functionalist—will emphasize how the capacity acts and interacts with its environment. This strategy risks lumping things together that behave similarly but have different underlying explanations. Clever Hans gives the correct answers to math problems (etc.), but not in the way mathematically trained humans do.<sup>3</sup> Both humans and cephalopods have eyes and phototransducers, but cephalopods have a stunning variety and diversity of photoreceptors relative to us (Kingston et al., 2015). So, there is a temptation to look internally for relevant differences on the assumption that the mechanisms underlying these processes further distinguish which functionally similar processes belong in the same kind and which do not. Hans answers questions by tracking the subtle head movements of the questioner, not from memory and reasoning. Human eyes have different receptors for different colors, but cephalopods have receptors that only track differences in the brightness of light. However, when cephalopod receptors are placed under cells that can change color, so-called chromatophores, they are able to detect differences in color with the same brightness by changing the color of these chromatophores (Godfrey-Smith, 2016, 77–79; Desmond Ramirez and Oakley 2015). If (behaviorally) similar processes turn out to have different underlying mechanisms, then the mechanistic anchoring approach to delimiting kinds of capacity enjoins us to split the kind, i.e., to consider each as an instance of a distinct cognitive kind.

It is important to note here that even though the decision to split kinds in the above examples is based on differences in the mechanisms at play, this does not mean that the discovery of these mechanisms precedes the observation of behavioral differences. In fact, Clever Hans' alternative 'calculating' technique and the cephalopod's alternative style of color detection were discovered by behavioral measures. However, differences in behavior need not in and of themselves lead to splitting of cognitive capacity-kinds. The example from the previous section is a clear case in point: behavior in a Simon task and behavior in a stop signal response task are different, and yet both are thought to involve the same cognitive capacity. It is not *just* behavioral differences between Clever Hans and humans and between cephalopods and humans that lead to the splitting of kinds; it is these difference in conjunction with the knowledge about their radically different biological constitutions.

The mechanistic anchoring approach is familiar in scientific arguments for revising cognitive ontology (or "reconstituting the phenomenon", see Bechtel & Richardson 2010). Here's an example from work on predictive coding. The brain makes top-down predictions about the visible world to facilitate the processing of visual stimuli, specifically to help filter out unambiguous pictures from obfuscatory noise. The question is: What brain mechanisms implement that capacity? One possibility is that higher-level predictions about the visual world act on low-level processing regions to suppress information consistent with the prediction, allowing only the error signal to flow forward through the system (Murray et al., 2002). Another pos-

---

<sup>3</sup> Clever Hans was a horse in early 20th -century Germany who allegedly was able to demonstrate striking knowledge of arithmetic and other facts about the world. It was later discovered that he managed to perform these tasks by observing the unconscious responses of the questioner when the correct answer was presented or completed.

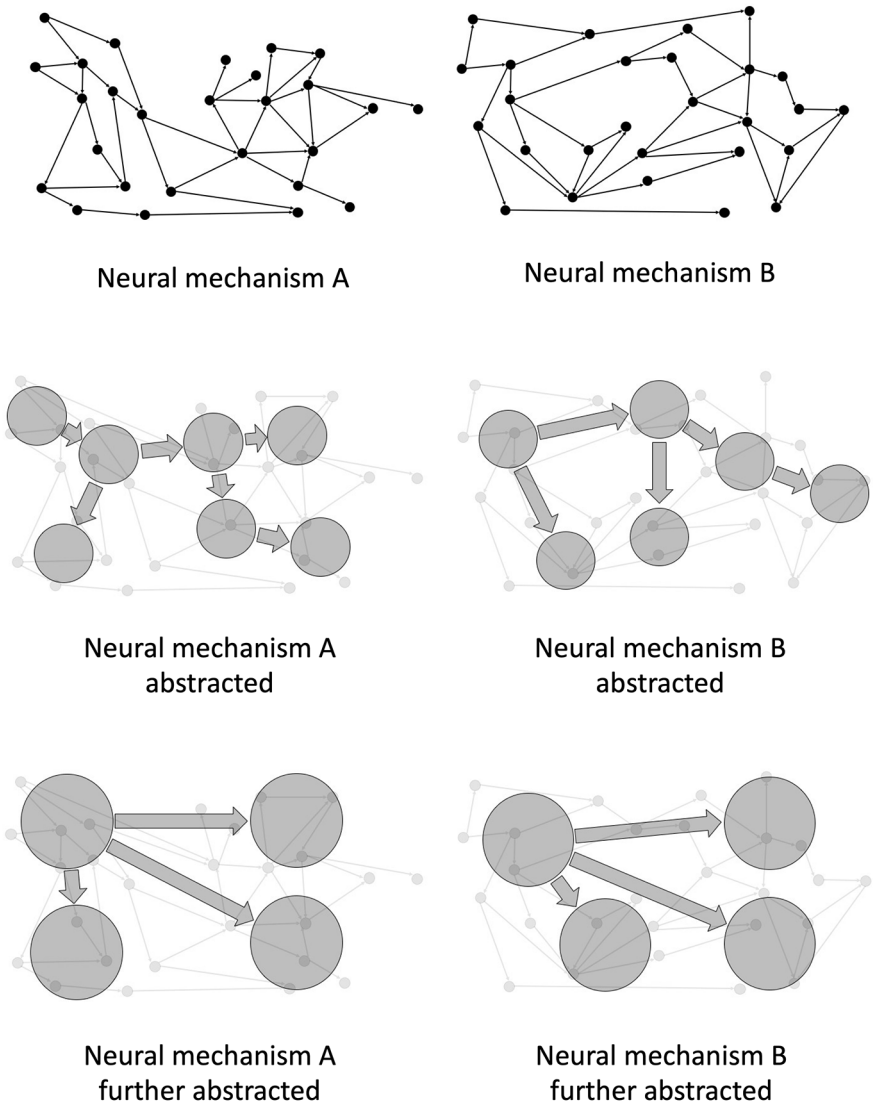
sibility is that the higher-level predictions act on low-level processing regions to sharpen representations with which it is consistent (Lee & Mumford, 2003). This in turn can be done either by suppressing incongruent lower-level information or by strengthening the signal of the predicted representations so they outcompete alternative representations. Suppose, not implausibly, that both mechanisms are at work in a single organism or that different organisms use different such mechanisms.<sup>4</sup> The correct ontology should recognize two systems in our cognitive ontology, not (or in addition to) the one described by the capacity alone.

This splitting is an important kind of ontological progress. But notice that it is predicated on a prior understanding of when two mechanisms are mechanisms of different kinds (see Craver 2009). If we are to follow the rule that we should split higher-level kinds when we discover that the same phenomenon is produced by two distinct kinds of mechanism, we need a further set of rules telling us when two mechanisms belong to the same or different kinds.<sup>5</sup> This is the same type of question with which we started this section. When are two mechanisms mechanisms of the same kind? Perhaps when they have the same kinds of parts, activities and organizing relations. But when are parts, activities and organizational relations of the same kind?

To judge two mechanisms to be similar or different, we have to decide on an appropriate grain of abstraction for describing those mechanisms (see Craver 2009; Levy 2018). At a high level of abstraction, both predictive mechanisms are using prediction to enhance perception. If we add in more detail, they begin to diverge. But here's the general philosophical problem: No two instances of 'the same' biological mechanism are physically—cell for cell, atom for atom—identical. There is inevitable biological variation from one person to the next, and even one instant to the next in the same person. To see any two mechanisms as of the same kind is necessarily to abstract away from these internal differences. Further, any physical difference between two mechanisms is also a causal (and so perhaps a functional) difference between them. So, when we lump two particular mechanisms under the same kind despite causal differences, we necessarily gloss over causal differences between them. At the other end of the spectrum, if we assign them to different kinds on the basis of only minor differences, every mechanism-instance becomes a kind unto itself, and the concept of a scientific kind ceases to be useful for putting like with like. There appears to be no uniquely correct degree of abstraction for describing any given system. Sometimes the differences matter; sometimes they do not (Fig. 3). The unfiltered causal structure of the world, therefore, lacks the resources required to specify the appropriate degree

<sup>4</sup> Kok et al., (2012) use an ingenious experiment to argue in favor of the 'sharpening-mechanism'. It may well be that this means that only one of the two sub-varieties of the capacity to use top-down predictions to disambiguate noisy visual inputs is actually used by the human brain. But the possibility that in different experimental circumstances it will be found that the other variety is used as well cannot be ruled out.

<sup>5</sup> We are silent here about the various methods that can be used to make such discoveries. Another problem that will complicate our analysis is presented by the fact that different methods of measuring the brain may yield different hypotheses about the mechanisms at play during the execution of a given task. Alternatively, it may not be clear what inferences to draw about neural mechanisms from e.g. fMRI data (e.g., Logothetis et al. 2001). Once again (see footnote 1) we acknowledge that this problem will further complicate our analysis, but we will not discuss it further because it does not alter our analysis of the 'cycle of kinds'—see below—and because we do not want to overcomplicate our discussion.



**Fig. 3** Schematic depiction of the Abstraction Problem. Two neural mechanisms are abstracted twice. On the first degree of abstraction, the two mechanisms are not of the same kind. When abstracted further, however, both fall under the same kind of mechanism

of abstraction and so to specify on its own when internal differences do and do not warrant splitting the kind. This is what we label the Abstraction Problem.

The practically-minded might not find the abstraction problem all that perplexing; they might see it simply as a reason to be pluralists about kinds, especially in the

special sciences. Pluralists acknowledge that the world contains many overlapping and at times cross-cutting kinds. Pluralism is well suited to the wide range of our actual and possible practical needs and to the character of contemporary science. The pluralist will insist that the boundaries of kinds are not completely arbitrary—as radical constructivists might hold. The legitimate causal kinds have to respect the causal structure of things; but that causal structure can be described in many ways (abstracting more or less, and here rather than there) each yielding a possibly legitimate way of carving the taxonomy of kinds, depending on one's needs. For instance, instead of splitting, it might make more sense to lump different types of predictive brain mechanisms if our focus is on similar varieties of behavior (de Lange, Heilbron, and Kok 2018). But different categorizations may be more or less useful in addressing competing needs and interests (see Chang 2004; Dupré 1993). This form of principled pluralism is, in our view, unobjectionable.

Once we take this pluralist implication on board, it turns out that two mechanisms will be mechanisms of the same kind when we see no relevant differences between them. But relevant to what? Well, relevant to the phenomenon we hope to control, explain, or predict (see Craver 2009; Levy and Bechtel 2013, p. 256). But notice that we have now run ourselves in a circle. The causal structure of the world was supposed to tell us which phenomena to include in or exclude from the mechanisms underlying cognitive capacities. But the judgment of which causal structures count as relevant parts of those mechanisms depends on how we have specified the capacity to be explained and the type of explanation required—a conceptual decision made at the beginning rather than discovered within the causal order of things. This, of course, is the opposite of the direction of fit the mechanistic anchoring strategy of kind delineation exploits: Whether two mechanisms are of the same kind depends on what phenomenon they are called upon to explain and how that phenomenon is characterized.

In summary, the hope that the Operationalization Problem can be solved by identifying cognitive capacities with neural mechanisms and sorting kinds of mechanisms faces the equally fraught challenge of sorting kinds of mechanism: To decide when two tasks measure the same capacity, we appeal to the sameness of underlying mechanisms, but sameness of underlying mechanisms depends on sameness of capacity, and judgments of the latter depend, as we have argued, upon how capacities are operationalized in tasks. This is one loop in the cycle of kinds.

## 4 The Boundary Problem

In addition to the Abstraction Problem, the mechanistic anchoring strategy for fixing cognitive kinds also faces a Boundary Problem. The Boundary Problem is the problem of saying which parts are in the mechanism and which are not. A solution to the Boundary Problem is required, first, to say where one mechanism ends, and another begins—both in space and in time. This problem might arise in a sequential mechanism, such as memory: Where does the encoding mechanism end and the storage mechanism begin? Secondly, a solution to the boundary problem is required to distinguish mechanisms from their environments and background conditions. Why are some of the entities, activities, and organizational features in the mechanism (in

the relevant causal structure) and others not? While this problem might be understood both in terms of spatial containment (what is “in” the mechanism) and in terms of temporal inclusion (what lies “between” input and output (Prychitko, 2019)) the more fundamental matter here is relevance (Craver, 2007; Craver et al., 2021). To be in the mechanism is to be relevant to how it works.<sup>6</sup>

The Boundary and Abstraction Problems are not completely independent; what counts as a component of a given mechanism may sometimes depend on how far we have abstracted away from certain causal difference between instances of ‘the same’ mechanism. Yet these problems are logically distinct. Return to the magnet example: The Abstraction Problem is the problem of deciding whether lodestones, ceramic magnets and electromagnets all belong to the same kind. The Boundary Problem is the problem of determining whether the power source of an electromagnet is part of the magnet (without power the thing will not attract iron and hence not be a magnet) or merely a background condition for the magnet to function or, perhaps more problematically, whether the ambient temperature is part of the magnet, given that its attractive force changes with temperature.

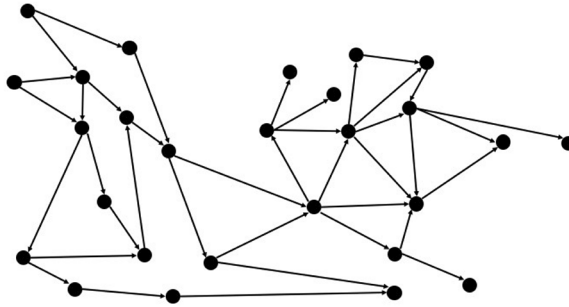
The problem for the mechanistic anchoring answer to the question of cognitive kinds arises from the fact that the world does not come packaged into neatly delineated mechanisms. The unfiltered causal structure of the world is blooming, buzzing confusion. It takes considerable insight to carve away enough of the obfuscatory and irrelevant detail to see the kind of orderly mechanistic structure depicted in the call-out boxes of our biology textbooks. (This always requires abstracting away from the unfiltered causal structure of the world—hence the connection between the Abstraction Problem and the Boundary Problem). This is why mechanism-discovery is an achievement and not merely a matter of reading off what causes what. When we consider mechanical effects, diffusion of molecules, heat transfer, metabolic exchange, waste production, electrical effects, etc., the causal structure of our bodies is bewilderingly complex and interwoven. And these are simply the occurrent mechanisms. Which of these mechanisms do we foreground and background in our search for kinds (Fig. 4)?

The problem is that one can be led to lump or split differently depending on which entanglements one decides to include in the mechanism. We can use top-down prediction in visual processing as an example again. One of the possible mechanisms underlying the use of predictions to disambiguate noisy information is ‘sharpening’ certain lower-level visual representations through excitatory feedback, allowing it to outcompete alternative representations in the fight for downstream processing. Are the lower-level alternative representations part of the mechanism that ultimately lead the brain to select the predicted visual image? Or are they the background from which the predicted image is selected? Both options are defensible; neither answer is uniquely highlighted by the world’s causal structure.

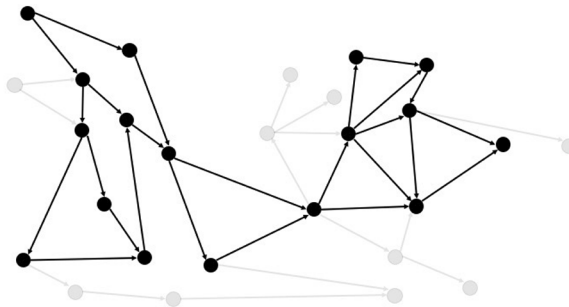
This, again, emphasizes a degree of liberty in this mechanistic anchoring approach (allowing for many possible, equally correct reasons for lumping and splitting

---

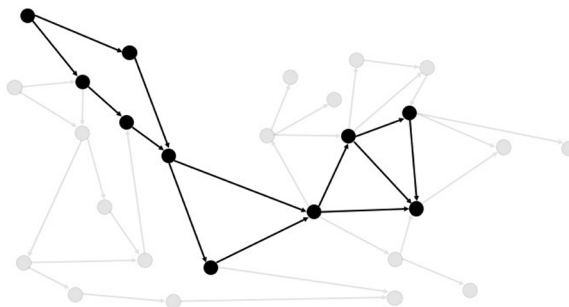
<sup>6</sup> For more on using the character of the phenomenon to limn the boundaries of mechanisms, see (Sullivan 2009; Irvine 2013; and Stinson 2016). Craver (2009; 2013) defends the view that the boundaries of mechanisms are in many cases driven by pragmatic decisions of the sort discussed in this section.



Neural mechanism including background conditions



Neural mechanism distinguished from background conditions (1)



Neural mechanism distinguished from background conditions (2)

**Fig. 4** Schematic depiction of the Boundary Problem. Different ways of distinguishing between mechanism and background conditions imply different mechanisms

depending on the mechanism to which one attends and on how one attends to it). And it again illustrates the fact that which mechanisms are relevant for building our

taxonomy depends on what we want our taxonomy to do. And here is the key point: This direction of fit is opposite of what we hoped to achieve when we turned to the mechanistic structure of the world to provide an objective anchor for our taxonomic decisions.

In practice in cognitive neuroscience, the Boundary Problem, while not often explicitly recognized as such, is reduced by methodological contrasts or subtractions. In neuropsychology, people with brain damage and cognitive deficits are compared to those without to identify the parts that do and do not contribute to the cognitive function in question; the removal of irrelevant parts is (all things equal) inconsequential for the capacity. In functional neuroimaging<sup>7</sup>, such as task-based fMRI, one compares activation profiles during task performance to those profiles obtained during rest; activation during task vs. control indicates relevance. In each case, the methods determine whether or not a part is relevant by reference to the behavioral task. Relevance is determined by deficits in task performance in neuropsychology and by activations during task performance in neuroimaging. In such studies of mechanistic relevance, task selection is necessary for and prior to the determination of relevance.

So we have again completed a loop in the Cycle of Kinds: Whether we have one or two kinds of mechanism will depend on which task conditions we choose for the purposes of studying it and on which control task conditions we believe properly exclude confounds.

## 5 Ontological progress: moving through the cycle of kinds

The slow and iterative nature of developing cognitive ontologies has been highlighted by different authors (e.g. Sullivan 2014, 2016; Francken & Slors 2014). In the above sections, we have identified and highlighted for consideration and criticism a pair of cyclical inferential structures that appear to be operating behind the scenes of this iterative process. The cognitive ontology project, the important effort to bring some order to the taxonomy of parts and operations that we take, collectively, to compose the mind, is defined in part by the need to find a simultaneous solution to three interlocking problems. Our discussion of these problems is aimed at making that core problematic explicit with the goal of exposing it thereby to more careful critical reflection.

The first challenge, the Operationalization Problem, is to provide sound reasons (as opposed to stipulations) for associating a given cognitive task with a given cognitive capacity. To solve this problem, we might look at the neural realizers of these capacities and compare the neural activity in different tasks that elicit the same capacity. However, this strategy leads us to two further problems that ultimately cycle back on themselves. First, we cannot determine whether two cognitive capacities are of the same kind by looking at their neural mechanisms. For we cannot decide when two mechanisms are mechanisms of the same kind by merely looking at their internal causal structures. Looking more or less abstractly at the mechanisms, we might

---

<sup>7</sup> For a look at how the contrastive method is used in neuroimaging, see Poldrack (2010). For a look at how it is used in neuropsychology, see Shallice 1988; Caramazza 1984.

lump and split one and the same mechanism differently. Likewise, purely anatomical approaches will earn their merit when and only when differences in, e.g., cell type or differences in laminar structure in fact represent functionally distinct regions of the brain. This is the Abstraction Problem, and it requires for its solution commitments as to how properly to characterize the capacity one is trying to explain. That is the first loop in the cycle.

The second loop in the cycle concerns boundaries. Because neural mechanisms do not come neatly packaged for us (and because there is— and should be— no consensus about what constitutes neat packaging), we have to decide which components and activities are part of the mechanism and which are instead e.g., background conditions that do not fall within the mechanistic boundaries or bits of the flotsam and jetsam floating in nearby spacetime. Indeed, although cognitive neuroscientists try to escape the Boundary Problem with their experimental controls and subtractions, in fact this only leads us again through the Cycle of Kinds, back to the Operationalization Problem.

If the above analysis is correct, there is something broadly circular in the core project of cognitive ontology—at least when we look to the neural mechanisms underlying cognitive capacities to anchor our ontological decisions. But ultimately, we do not think this approach is *viciously* circular. It would be viciously circular if the Operationalization Problem we end up with when trying to solve the Boundary Problem and/or the Abstraction Problem is exactly the same problem we wanted to solve by letting neural mechanisms anchor our taxonomic decisions in cognitive ontology. But that need not be the case. Although deciding on the appropriate boundaries and levels of abstraction in delineating mechanisms cannot be done without having recourse to functional, task-related considerations about cognitive capacities, this does not mean that the tasks involved in these considerations are the same original tasks we started out with. One should expect in this project an iterative correction-process in which tasks, capacities, and mechanisms adjust, scanning the possible ways of making sense of the mind and its parts. The aims of this apparent wandering are to reduce predictive error, to maximize explanatory reach, and to inspire effective practical applications (such as treatments and other remedies).

Return again to the use of higher-level predictions about the visual world to filter out a picture from noisy visual inputs. Experiments with the perception of random lines and 2D/3D shapes (made up of similar lines ordered slightly differently), suggest that higher-level predictions suppress redundant information in lower-level visual information processing (Murray et al., 2002). Another study, using computational models of predictive coding combined with experimental evidence on the perception of illusory contours, suggests that predictions sharpen expected lower-level visual representations (Lee & Mumford, 2003). If we abstract away far enough from these proposed mechanisms, they are similar because both suggest that higher-level predictions are used to filter out unambiguous pictures from noisy visual inputs. But if we zoom in more, the mechanisms are different because one is based on excitatory feedback to lower-level expected visual representations while the other suggests inhibition. Kok et al., (2012) have used a *third* behavioral experiment, involving the perception of grating stimuli *and* a clever manipulation of expectations (using auditory



inputs that are connected with visual stimuli in familiarization trials) to settle (for now) this abstraction problem (in favor of the finer-grained, excitatory mechanism).

The part of the Boundary and Abstraction Problem that needs to be settled by going back to functional, task-related considerations about cognitive capacities, then, need not overlap completely with the functional-level question we started with when deciding to let neural mechanisms anchor our taxonomic decisions. The cycle can be iterative, fine-tuning answers to each problem under the constraint of the others, without being viciously circular (see Chang 2004). And the outcome of the search for such an conceptual equilibrium is bound to drive us to find, to the extent possible, mutually satisfying and reinforcing mappings among the solutions to our three problems.

Progress in this iterative practice involves reducing the incongruities among tasks, cognitive ontologies, and our understanding of mechanisms. We cycle from a functional description of capacities, to the neural implementation, to the task choice and task model, and back again. Repetition of this circular process need not involve stagnation but may yield increasingly more refined functional concepts and informed decisions about what to count as and in a mechanism. It would, after all, be a tremendous achievement to bring our solutions to these three component problems in the cycle of kinds (tasks, capacities/functional roles, and mechanisms) into unforced alignment for any given practical project. Such alignment is the stop signal in the search for cognitive kinds, halting the cycle of accommodation among the solutions to its constituent problems.<sup>8</sup>

Our proposal aligns well with key features of Boyd's "homeostatic property cluster" theory of natural kinds (Boyd, 1989). According to Boyd, analytical functionalist (externalist) and mechanistic anchoring (internalist) approaches to kind delineation can be conjoined: An object or capacity is a member of a kind in virtue of the similarities both in how they regularly behave (the functionalist answer) and in the mechanisms that explain why they regularly behave that way. Boyd once embraced a naturalistic "principle of accommodation" that enjoins us to split kinds whenever we find inductively relevant differences; and these differences might be found either in the cognitive function or in the mechanism. Boyd's historic view expresses concisely the idea that scientists should populate their models and theories with kinds that best systematize our knowledge of the world's causal structure (Salmon, 1984) and that therefore offer the most "bang for the buck," maximizing predictive leverage and instrumental control in the most economical way (Strevens, 2008). If one fails to recognize real distinctions between kinds, one's model or theory necessarily suffers in some prediction, instrumental application, and explanatory task (reducing the bang). On the other hand, if one distinguishes in one's models and theories functionally and mechanistically identical kinds of objects and capacities, one necessarily introduces

---

<sup>8</sup> Of course, it is possible that the selective force of science has driven us only to a local minimum of error in coordinating the answers to these three basic questions. In that case, it would be little exaggeration to say that revolutionary changes in our understanding of the world's kinds, of the techniques for detecting them, and for our chunking of the causal structures that render them stable will be required to reach the global error minimum. We confront the exciting possibility that the stability we now seek might afford only shallow coherence relative to other conceptual schemes we have not yet considered or that we have for the time-being excluded from our imaginative vision.

predictively and instrumentally irrelevant and otiose detail (increasing buck). Boyd is not explicitly concerned with the experimental tasks used to operationalize the capacities in question. But if we are to understand the process by which our cognitive ontology will be sharpened, we have to consider the tools and methods by which we engage and measure these capacities. These tools and methods cannot simply be taken for granted but are themselves part of the intellectual background that embodies our ontology in material practices. These practices are especially worthy of attention in the cognitive sciences in part because there is as much difference of opinion (even outright controversy) over the adequacy of different tasks as there is about the correct ontology. The ability to construct any coherent and economical picture of how tasks, capacities, and mechanisms relate, is itself a scientific achievement that deserves to be taken seriously. It is a further question how different ways of bringing our answers to these three questions into equilibrium with one another are and should be evaluated relative to one another and whether it is permissible to have, in a science, more than one such stable arrangement of solutions.

The iterative cycle we describe here is not unique to cognitive neuroscience but has analogues in even more “basic” sciences, especially in the early days of concept formation. For example, Hasok Chang describes a similar process of “epistemic iteration” for the concept of temperature, its mechanisms, and its measures in the study of temperature in the Eighteenth and Nineteenth Centuries (Chang, 2004). Knowledge accumulation is possible even where we can point to no secure and indubitable foundations, but it is measured in its coherence, its predictive adequacy, and its stability. The feeling of hotness acts as the first, intuitive and roughly-hewn touchstone guiding the search for the thing, “heat”, that might explain it, even when it is utterly unclear what, if anything, the feeling of hotness detects in things. Likewise, our intuitive interactions with remembrance serve as the anchor point for scientific investigation, which involves the development of tasks, controls, and ontologies that can lead us productively away from that intuitive home. It is therefore not a conceptual failing of contemporary cognitive science that its taxonomy of kinds is in flux; this is in keeping both with a healthy pluralism, as described above, and with the way that other sciences have developed. A certain looseness in kind definitions and matters of ontology, especially when such matters are (most would agree) far from settled, provides space for scientific research programs to live and breathe. Yet their work, if it is to make the iterative progress seen in other sciences, must be guided by an underlying recognition of the task and the task-model: To make progress by moving through the cycle of kinds, bringing our tasks, concepts, and explanations into stable equilibrium.

So here is a practical take-home message for cognitive neuroscience: We have argued that progress in cognitive ontology will be iterative and cyclic, even in the best of conditions. In practice, this means that there is an important, additional stage after data analysis and interpretation: Going back to, and possibly correcting the cognitive ontology, the external (functional) description, one’s understanding of the mechanisms, and one’s model of the task. In the words of Krakauer et al., (2017) neuroscience needs behavior; we should not expect thoroughgoing bottom-up reform of cognitive ontologies, precisely because the development of such ontologies is cyclical, not a one-way affair. The current emphasis on neural mechanisms should be balanced by a renewed interest in the mind and behavior and how we study them

experimentally. The cycle of kinds is deep and unbroken, but it is the ineliminable tie that binds our experimental practices to our ontological categories for parsing mind and brain alike.

**Acknowledgements** Jolien Francken and Marc Slors would like to thank Eelke Spaak for helpful comments. Carl Craver would like to thank Russel Poldrack, Adina Roskies and Joe McCaffrey as well as the participants in the 2018 WUSTL Cognitive Ontology workshop and the 2020 Pitt HPS Cognitive ontology workshop for helpful input. All authors would like to thank two anonymous referees for this journal for very insightful comments that have helped us to improve this paper considerably.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Anderson, M. L. (2015). Mining the Brain for a New Taxonomy of the Mind. *Philosophy Compass*, 10(1), 68–77. <https://doi.org/10.1111/phc3.12155>
- Baluška, F. & Levin, M. (2016). On Having No Head: Cognition throughout Biological Systems. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00902>
- Bédécarrats, A., Chen, S., Pearce, K., Cai, D., Glanzman, D.L. (2018). RNA from Trained Aplysia Can Induce an Epigenetic Engram for Long-Term Sensitization in Untrained Aplysia. *eNeuro* 5(3). <https://doi.org/10.1523/ENEURO.0038-18.2018>
- Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Taylor & Francis Group
- Bechtel, W., & Richardson, R. C. (2010). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. MIT Press
- Bissett, P. G., Jones, H. M., Poldrack, R. A., & Logan, G. D. (2020). Severe and Pervasive Violations of Independence in Response Inhibition Tasks. *PsyArXiv* 7 (12). Online publication. <https://doi.org/10.1126/sciadv.abf4355>
- Bollhagen, A. (2021). The inchworm episode: reconstituting the phenomenon of kinesin motility. *European Journal of Philosophy of Science*. Online publication preprint. <http://philsci-archives.pitt.edu/18803/>
- Borsboom, D., Cramer, A. O. J., Kievit, R. A., Zand Scholten, A., & Franic, S. (2009). The End of Construct Validity. In R. W. Lissitz (Ed.), *The Concept of Validity: Revisions, New Directions and Applications* (pp. 135–170). Information Age Publishing
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, 111(4), 1061–1071. Doi:<https://doi.org/10.1037/0033-295X.111.4.1061>
- Boyd, R. (1989). What Realism Implies and What It Does Not. *Dialectica*, 43(1–2), 5–29. <https://doi.org/10.1111/j.1746-8361.1989.tb00928.x>
- Bridgman, P. W. (1927). *The Logic of Modern Physics*. New York: Macmillan
- Caramazza, A. (1984). The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language*, 21, 9–20. DOI: [https://doi.org/10.1016/0093-934x\(84\)90032-4](https://doi.org/10.1016/0093-934x(84)90032-4)
- Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. New York: Oxford University Press
- Chang, H. (2021). Operationalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021). Online publication. <https://plato.stanford.edu/archives/fall2021/entries/operationalism/>

- Colaço, D. (2018). Rip It up and Start Again: The Rejection of a Characterization of a Phenomenon. *Studies in History and Philosophy of Science Part A*, 72, 32–40. Doi:<https://doi.org/10.1016/j.shpsa.2018.04.003>
- Colaço, D. (2020). Recharacterizing Scientific Phenomena. *European Journal for Philosophy of Science* 10(2), 1–19. Online publication. Doi:<https://doi.org/10.1007/s13194-020-0279-z>
- Craver, C. F. (2007). *Explaining the Brain*. Oxford: Oxford University Press
- Craver, C.F. (2009) Mechanisms and natural kinds, *Philosophical Psychology*, 22:5, 575–594. <https://doi.org/10.1080/09515080903238930>
- Craver, C. F. (2013). Functions and mechanisms: A perspectivalist view. In P. Huneman (Ed.), *Functions: Selection and mechanisms* (pp. 133–158). Springer
- Craver, C. F., & Darden, L. (2001). Discovering Mechanisms in Neurobiology: The Case of Spatial Memory. In P. K. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and Method in the Neurosciences* (pp. 112–137). University of Pittsburgh Press
- Craver, C. F., Glennan, S., & Povich, M. (2021). Constitutive relevance & mutual manipulability revisited. *Synthese*, 199, 8807–8828. <https://doi.org/10.1007/s11229-021-03183-8>
- Cronbach, L. J., & Meehl, P. E. (1995). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Dalley, J. W., & Robbins, T. W. (2017). Fractionating Impulsivity: Neuropsychiatric Implications. *Nature Reviews Neuroscience*, 18(3), 158–171. <https://doi.org/10.1038/nrn.2017.8>
- Danziger, K. (1997). *Naming the Mind: How Psychology Found Its Language*. SAGE Publications
- Dennett, D. C. (1987). *The Intentional Stance*. MIT Press
- Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Harvard University Press
- Eagle, D. M., Baunez, C., Hutcheson, D. M., Lehmann, O., Shah, A. P., & Robbins, T. W. (2008). Stop-Signal Reaction-Time Task Performance: Role of Prefrontal Cortex and Subthalamic Nucleus. *Cerebral Cortex*, 18(1), 178–188. DOI: <https://doi.org/10.1093/cercor/bhm044>
- Feest, U. (2005). Operationism in Psychology: WHAT the Debate Is about, What the Debate Should Be About. *Journal of the History of the Behavioral Sciences*, 41(2), 131–149. Doi:<https://doi.org/10.1002/jhbs.20079>
- Feest, U. (2010). Concepts as Tools in the Experimental Generation of Knowledge in Cognitive Neuropsychology. *Spontaneous Generations: A Journal for the History and Philosophy of Science* 4 (1). Doi:<https://doi.org/10.4245/sponge.v4i1.11938>
- Feest, U. (2011). What Exactly Is Stabilized When Phenomena Are Stabilized? *Synthese* 182 (1), 57–71. Doi:<https://doi.org/10.1007/s11229-009-9616-7>
- Feest, U. (2017). Phenomena and Objects of Research in the Cognitive and Behavioral Sciences. *Philosophy of Science* 84 (5), 1165–76. Doi:<https://doi.org/10.1086/694155>
- Feest, U. (2020). Construct Validity in Psychological Tests – the Case of Implicit Social Cognition. *European Journal for Philosophy of Science* 10 (1). Online publication. Doi:<https://doi.org/10.1007/s13194-019-0270-8>
- Feest, U. (forthcoming). Operationism and the Epistemology of Discovery in Experimental Psychology.
- Francken, J. C., & Slors, M. (2014). From Commonsense to Science, and Back: The Use of Cognitive Concepts in Neuroscience. *Consciousness and Cognition*, 29, 248–258. Doi: <https://doi.org/10.1016/j.concog.2014.08.019>
- Francken, J. C., & Slors, M. (2018). Neuroscience and Everyday Life: Facing the Translation Problem. *Brain and Cognition* 120, 67–74. Doi: <https://doi.org/10.1016/j.bandc.2017.09.004>
- Godfrey-Smith, P. (2016). *Other Minds. The Octopus, the Sea, and the Deep Origins of Consciousness*. Earrar, Strauss and Giroux
- Irvine, E. (2013). *Consciousness as a Scientific Concept*. Springer
- Janssen, A., Klein, C., & Slors, M. (2017). What Is a Cognitive Ontology, Anyway? *Philosophical Explorations*, 20(2), 123–128. <https://doi.org/10.1080/13869795.2017.1312496>
- Khalidi, M. A. (2017). Crosscutting Psycho-Neural Taxonomies: The Case of Episodic Memory. *Philosophical Explorations*, 20(2), 191–208. <https://doi.org/10.1080/13869795.2017.1312501>
- Kingston, A. C. N., Wardill, T. J., Hanlon, R. T., Cronin, T. W. (2015). An Unexpected Diversity of Photoreceptor Classes in the Longfin Squid, *Doryteuthis Pealeii*. *PLoS ONE* 10(9): e0135381. <https://doi.org/10.1371/journal.pone.0135381>
- Klein, C. (2012). Cognitive Ontology and Region- versus Network-Oriented Analyses. *Philosophy of Science*, 79(5), 952–960. Doi: <https://doi.org/10.1086/667843>

- Knolle, F., McBride, S. D., Stewart, J. E., Goncalves, R. P., & Morton, A. J. (2017). A Stop-Signal Task for Sheep: Introduction and Validation of a Direct Measure for the Stop-Signal Reaction Time. *Animal Cognition*, 20(4), 615–626. DOI: <https://doi.org/10.1007/s10071-017-1085-7>
- Kok, P., Rahnev, D., Jehee, J. F. M. M., Lau, H. C., & De Lange, F. P. (2012). Attention Reverses the Effect of Prediction in Silencing Sensory Signals. *Cerebral Cortex*, 22(9), 2197–2206. DOI: <https://doi.org/10.1093/cercor/bhr310>
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, 93(3), 480–490. DOI: <https://doi.org/10.1016/j.neuron.2016.12.041>
- de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, 22(9), 764–779. <https://doi.org/10.1016/j.tics.2018.06.002>
- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America*, 20(7), 1434–1448. DOI: <https://doi.org/10.1364/josaa.20.001434>
- Lenartowicz, A., Kalar, D. J., Congdon, E., & Poldrack, R. A. (2010). Towards an Ontology of Cognitive Control. *Topics in Cognitive Science*, 2(4), 678–692. <https://doi.org/10.1111/j.1756-8765.2010.01100.x>
- Levy, A. & Bechtel, W. (2013). Abstraction and the Organization of Mechanisms. *Philosophy of Science* 80(2), 241–261. <https://doi.org/10.1086/670300>
- Levy, A. (2018). Idealization and abstraction: Refining the distinction. *Synthese*, 198, 5855–5872. <https://doi.org/10.1007/s11229-018-1721-z>
- Logan, G. D., & Cowan, W. B. (1984). On the Ability to Inhibit Thought and Action: A Theory of an Act of Control. *Psychological Review*, 91(3), 295–327
- Logothetis, N., Pauls, J., Augath, M. et al. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157. <https://doi.org/10.1038/35084005>
- McCaffrey, J. B., & Machery, E. (2016). The Reification Objection to Bottom-up Cognitive Ontology Revision. *The Behavioral and Brain Sciences*, 39, e125. DOI: <https://doi.org/10.1017/S0140525X15001594>. Online publication
- McDermott, K. B., Szpunar, K. K., & Christ, S. E. (2009). Laboratory-Based and Autobiographical Retrieval Tasks Differ Substantially in Their Neural Substrates. *Neuropsychologia*, 47(11), 2290–2298. DOI: <https://doi.org/10.1016/j.neuropsychologia.2008.12.025>
- Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduced activity in human primary visual cortex. *PNAS*, 99(23), 15164–15169. DOI: <https://doi.org/10.1073/pnas.192579399>
- Pani, P., Giarrocco, F., Giamundo, M., Montanari, R., Brunamonti, E., & Ferraina, S. (2018). Visual Salience of the Stop Signal Affects the Neuronal Dynamics of Controlled Inhibition. *Scientific Reports*, 8, 14265. DOI: <https://doi.org/10.1038/s41598-018-32669-8>
- Piccinini, G. (2020). *Neurocognitive Mechanisms*. Oxford University Press
- Poldrack, R. A. (2010). Mapping Mental Function to Brain Structure: How Can Cognitive Neuroimaging Succeed? *Perspectives on Psychological Science*, 5(6), 753–761. DOI: <https://doi.org/10.1177/1745691610388777>
- Poldrack, R. A., & Yarkoni, T. (2016). From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure. *Annual Review of Psychology*, 67, 587–612. DOI: <https://doi.org/10.1146/annurev-psych-122414-033729>
- Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y. ... Bilder, R. M. (2011). The cognitive atlas: Toward a knowledge foundation for cognitive neuroscience. *Frontiers in Neuroinformatics*, 5, 17. <https://doi.org/10.3389/fninf.2011.00017>
- Poldrack, R. A. (2006). Can Cognitive Processes Be Inferred from Neuroimaging Data? *Trends in Cognitive Sciences*, 10(2), 59–63. <https://doi.org/10.1016/j.tics.2005.12.004>
- Price, C. J., & Friston, K. J. (2005). Functional Ontologies for Cognition: The Systematic Definition of Structure and Function. *Cognitive Neuropsychology*, 22(3–4), 262–275. DOI: <https://doi.org/10.1080/02643290442000095>
- Prychitko, E. (2019). The Causal Situationist Account of Constitutive Relevance. *Synthese* 198(2). Online publication. DOI: <https://doi.org/10.1007/s11229-019-02170-4>
- Ramirez, D. M., & Oakley, T. H. (2015). Eye-Independent, Light-Activated Chromatophore Expansion (LACE) and Expression of Phototransduction Genes in the Skin of Octopus Bimaculoides. *Journal of Experimental Biology*, 218(10), 1513–1520. DOI: <https://doi.org/10.1242/jeb.110908>
- Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press

- Seghier, M. L., & Price, C. J. (2018). Interpreting and utilizing subject variability in brain function. *Trends in Cognitive Science*, 22(6), 517–530. <https://doi.org/10.1016/j.tics.2018.03.003>
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press
- Stinson, C. (2016). Mechanisms in Psychology: Ripping Nature at Its Seams. *Synthese*, 193(5), 1585–1614. Doi:<https://doi.org/10.1007/s11229-015-0871-5>
- Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press
- Sullivan, J. A. (2007). *Reliability and Validity of Experiment in the Neurobiology of Learning and Memory*. Dissertation, University of Pittsburgh. <https://philarchive.org/archive/ANNRAV>
- Sullivan, J. A. (2009). The Multiplicity of Experimental Protocols: A Challenge to Reductionist and Non-Reductionist Models of the Unity of Neuroscience. *Synthese* 167 (3), 511–39. Doi: <https://doi.org/10.1007/s11229-008-9389-4>
- Sullivan, J. A. (2010). Reconsidering ‘spatial Memory’ and the Morris Water Maze. *Synthese* 177 (2), 261–83. <https://doi.org/10.1007/s11229-010-9849-5>
- Sullivan, J.A. (2014). Is the Next Frontier in Neuroscience a ‘Decade of the Mind’?. In: Wolfe, C.T. (eds) *Brain Theory*. Palgrave Macmillan, London. [https://doi.org/10.1057/9780230369580\\_4](https://doi.org/10.1057/9780230369580_4)
- Sullivan, J. A. (2015). Kendig, C. (Ed.). (2015). *Natural Kinds and Classification in Scientific Practice* (pp. 47–56). Routledge. Doi:<https://doi.org/10.4324/9781315619934>
- Sullivan, J. A. (2016). Construct Stabilization and the Unity of the Mind-Brain Sciences. *Philosophy of Science* 83 (5), 1–21. Doi:<https://doi.org/10.1086/687853>
- Sullivan, J. A. (2017). Coordinated Pluralism as a Means to Facilitate Integrative Taxonomies of Cognition. *Philosophical Explorations* 20 (2), 129–45. <https://doi.org/10.1080/13869795.2017.1312497>
- Tal, E. (2016). Making Time: A Study in the Epistemology of Measurement. *British Journal for the Philosophy of Science*, 67(1), 297–335. Doi: <https://doi.org/10.1093/bjps/axu037>
- Uttal, W. R. (2001). *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. MIT Press
- Verbruggen, F., & Logan, G. D. (2008). Response Inhibition in the Stop-Signal Paradigm. *Trends in Cognitive Sciences*, 12(11), 418–424. DOI: <https://doi.org/10.1016/j.tics.2008.07.005>
- Viola, M. (2021). Beyond the Platonic Brain: facing the challenge of individual differences in function-structure mapping. *Synthese*, 199, 2129–2155. <https://doi.org/10.1007/s11229-020-02875-x>
- Ward, Z. B. (2019). Registration Pluralism and the Cartographic Approach to Data Aggregation Across Brains. *The British Journal for the Philosophy of Science*. Online publication. <http://philsci-archive.pitt.edu/id/eprint/16159>
- Wasilewski, P., & Kletetschka, G. (1999). Lodestone: Natures Only Permanent Magnet-What It Is and How It Gets Charged. *Geophysical Research Letters*, 26(15), 2275–2278. <https://doi.org/10.1029/1999GL900496>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.