3 Neuroscientific kinds through the lens of scientific practice

Jackie Sullivan

In this chapter, I argue that scientific practice in the neurosciences of cognition is not conducive to the discovery of natural kinds of cognitive capacities. The "neurosciences of cognition" include *cognitive neuroscience* and *cognitive neurobiology*, two research areas that aim to understand how the brain gives rise to cognition and behavior. Some philosophers of neuroscience have claimed that explanatory progress in these research areas ultimately will result in the discovery of the underlying mechanisms of cognitive capacities.[1] Once such mechanistic understanding is achieved, cognitive capacities purportedly will be relegated into natural kind categories that correspond to real divisions in the causal structure of the world. I provide reasons here, however, to support the claim that the neurosciences of cognition currently are not on a trajectory for discovering natural kinds. As I explain, this has to do with how mechanistic explanations of cognitive capacities are developed. Mechanistic explanations and the kinds they explain are abstract representational by-products of the conceptual, experimental and integrative practices of neuroscientists. If these practices are not coordinated towards developing mechanistic explanations that mirror the causal structure of the world, then natural kinds of cognitive capacities will not be discovered. I provide reasons to think that such coordination is currently lacking in the neurosciences of cognition and indicate where changes in these practices appropriate to the natural kinds ideal would be required if it is indeed the goal. However, an evaluation of current practices in these research areas is suggestive that discovering natural kinds of cognitive capacities is simply not the goal.

A primary aim of the neurosciences of cognition is to understand how the brain gives rise to cognitive capacities such as *vision*, *attention*, *working memory*, *face recognition*, *reward-based learning*, and *spatial memory*. Cognitive neuroscience investigates the neural basis of human cognitive capacities and aims to identify which brain regions subserve which capacities. Experiments typically involve human subjects performing cognitive tasks or being run through experimental paradigms designed to individuate different kinds of cognitive capacities. Functional neuroimaging and/or electrophysiological recording techniques are used to determine which areas of the brain are active during task performance. Non-invasive intervention techniques are sometimes used to interrupt brain activity to determine its impact on task performance. When such experiments are successful, cognitive neuroscience sheds light on the brain areas that underlie cognitive capacities.

Cognitive neurobiology, in contrast, investigates the synaptic, cellular and molecular mechanisms that underlie cognition. Neurobiologists work primarily with animal models (e.g., rodents, fruit flies) and combine the behavioral techniques of experimental psychology (e.g., classical and operant conditioning paradigms) with electrophysiological, pharmacological, genetic and protein analysis techniques. Intervention experiments are commonly used to determine the impact of synaptic, cellular and molecular changes on changes in behavior associated with a given cognitive capacity. To take an example, an experiment to determine whether activation of a specific protein kinase is requisite for a cognitive capacity like "spatial memory" typically involves training a rodent in a spatial memory paradigm (e.g., the Morris water maze) and blocking the activity of the protein kinase to determine the impact on performance in that paradigm.

One question that has arisen in recent philosophical debates about the nature of neuroscientific explanation is whether the neurosciences of cognition are poised to develop categories of cognitive capacities that track natural kinds. At least two responses to this question have emerged in the philosophical literature. The first, offered by Carl Craver and his collaborators, is that neuroscience not only *can* but also *ought* to aim to discover natural kinds of cognitive capacities or "mechanistic property clusters (MPCs)" – groups of properties that co-occur and whose "co-occurrence [. . .] from individual to individual is explained by causal mechanisms that regularly ensure these properties are instantiated together."[2] According to the MPC account, sciences like neuroscience typically begin by grouping phenomena together on the basis of detectable surface features. These preliminary groupings do not typically track mechanisms. In fact, scientists may wrongly "lump" together phenomena that are actually different or incorrectly "split" phenomena that should be grouped together.[3] As more is learned about the mechanisms that realize a phenomenon of interest, more appropriate ways to carve the phenomena are revealed, and scientists thus revise their groupings accordingly. To take an example, originally an entire class of behavioral phenomena was grouped together as memory phenomena. However, it was later discovered, via patients like Henry Molaison (H.M.), that medial temporal lobe structures underlie the ability to explicitly or consciously recall facts and events, whereas other structures underlie other forms of memory. This prompted "splitting" the category of memory. On the MPC view, as more is learned about the mechanisms of memory, more revisions are possible.

Proponents of the MPC view and "the ontic account" of mechanistic explanation downplay the role of scientific practice in shaping mechanistic explanations and the kinds scientists discover.[4] According to them, it is not scientists who explain but rather the mechanisms in the world that do the explaining. Scientists merely corral phenomena into groups; causal interventions inform this relegation.[5] Mechanistic property clusters or natural kinds are out there in the world to be discovered; scientists discover them or they do not. While the MPC account contains a normative prescription that scientists should strive to discover these clusters, its proponents are relatively silent with respect to how to better facilitate their discovery in instances in which a science may struggle to find them.

In contrast, another advocate for mechanistic explanation in the neurosciences of cognition, William Bechtel, points to the *cognitive labor* involved in developing mechanistic explanations.[6] He contends that "what [a] scientist advances is a representation of a mechanism – a construal of it – not the mechanism itself," and that appealing to such representations is fundamental for improving our understanding of a mechanism.[7] While he admits that these "representations [. . .] may not accurately represent the mechanisms operative in the world,"[8] he does seem to think that correct depiction is the goal. Although he is making a purely descriptive claim about the neurosciences of cognition, he believes that recognizing that mechanistic explanations are representations is the only way to make sense of Craver's normative claim that these areas of science ought to discover mechanisms (or MPC kinds). Because explanation cannot be liberated from a representational framework, at best what Craver can say, according to Bechtel, is that neuroscientists should aim to provide accurate representations of mechanisms. That is the closest that science can come to discovering natural or MPC kinds.

I am sympathetic to Bechtel's claim that representing is equally as important as intervening in discovering the mechanisms of cognitive capacities. Yet, representing and intervening are broad categories of practices that do not by themselves exhaust all of the possible types of practices that potentially shape the discovery process. According to Bechtel, neuroscientists also *decompose* systems like the brain into structural and functional subcomponents and *localize* specific cognitive functions in specific parts of the brain. In the process of engaging in these heuristic activities, neuroscientists *deploy concepts*, *design experiments*, *interpret data*, and *combine/integrate experimental findings* to support knowledge claims and build explanatory models of cognitive capacities. Furthermore, these practices take place in different contexts like scientific laboratories, research papers, review papers, websites, and conferences. It thus seems reasonable to characterize scientific practice in the neurosciences of cognition as involving many different kinds of practices that are relevant to or play a role in the discovery of natural kinds of cognitive capacities. Three types of practices that I want to consider here are

1 conceptual practices
2 experimental practices and
3 integrative practices.

By "conceptual practices," I mean to capture the ways in which investigators *generally* deploy concepts so as to pick out explanatory targets of interest, like cognitive capacities. Such concepts are theoretical constructs that are often only very loosely defined. In contrast, the designation "experimental practices" is intended to pick out the within-laboratory procedures – experimental paradigms or cognitive tasks – by which investigators specify how to produce, detect and measure a cognitive capacity of interest in the laboratory. While experimental paradigms do not exhaust the class of within-lab practices that potentially shape the kinds under study in the neurosciences of cognition, they are the strategies I will primarily focus on here. Finally, when investigators attempt to bring together data emanating from many different laboratories into a coherent model of the mechanisms involved in the production of a cognitive capacity, they may be said to engage in "integrative practices." Discussion sections of research papers and review papers are two contexts in which integrative mechanistic or explanatory models are developed. Another venue for integrating research study findings are published meta-analyses, which are becoming more commonplace in the neurosciences of cognition for bringing results from many labs together to bear on questions about functional localization and the mechanisms subserving cognitive capacities.

What I aim to show is that the relationship between these three kinds of practices is dynamic and this has specific implications for the kinds of cognitive capacities that the neurosciences of cognition discover. To do this, I want to appeal to a conceptual framework I introduced previously for thinking about what I refer to simply as "the experimental process" (Figure 3.1a).[9] Briefly, this process begins with an empirical question about a cognitive capacity/function of interest (Figure 3.1a). Before an investigator attempts to go in search of the neural basis or cellular and molecular mechanisms of a cognitive capacity, she typically has grouped together instances of what she takes to be the same capacity under a concept or construct. She may rely on how other investigators define the concept, but she may also define the term slightly differently. To take an example, rats successfully navigating mazes, birds correctly remembering where their nests are located and taxi drivers being able to navigate London in the absence of maps or GPS are often grouped together under the construct "spatial memory." Other constructs that designate cognitive capacities include *working memory*, *language*, *attention*, *face recognition* and *procedural memory* (to name only a handful). Such constructs originate with a concept that investigators associate with certain observations that serves as basis for theory building and experimental task/paradigm design and construction. When investigators deploy such constructs in the introduction of research papers in order to motivate a set of experiments, for example, they are engaged in such conceptual practices.

Once an investigator has selected a cognitive capacity of interest, which is designated by a construct, she then engages in various *experimental practices* directed at investigating that capacity in the laboratory (Figure 3.1a). Most importantly, she develops an experimental paradigm – a set of procedures for producing, measuring and detecting an instance of that capacity in the laboratory – which is used in conjunction with imaging or recording technologies (e.g., fMRI, electrophysiology) and intervention techniques (e.g., TMS, pharmacological inhibitors) to identify the brain areas or cellular and molecular mechanisms that underlie that capacity. For example, an experimental paradigm used to investigate a cognitive capacity such as spatial memory will include a set of production procedures that specify the stimuli (e.g., distal and local cues) to be presented, how those stimuli are to be presented/arranged (e.g., spatially, temporally) and how many times each stimulus is to be presented during phases of pre-training, training and post-training/testing. The paradigm will also include measurement procedures that specify the response variables to be measured in the pre-training and post-training/testing phases of the experiment and how to measure them using apparatuses designed for such measurement. Finally, a set of detection procedures

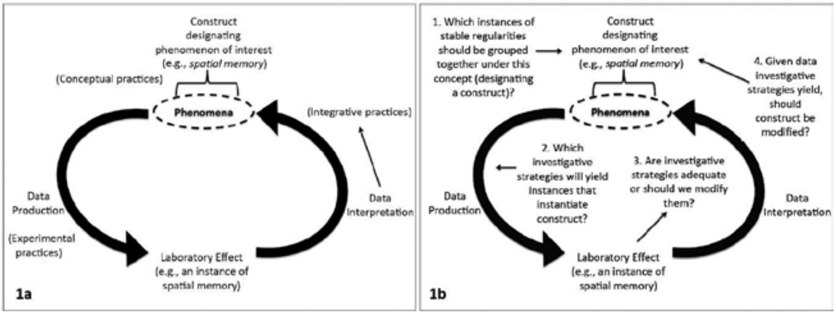

Figure 3.1 The experimental process: conceptual, experimental and integrative practices (3.1a) and construct explication (3.1b).

specifies what the comparative measurements of the response variables from the different phases of the experiment must equal in order to ascribe the cognitive capacity of interest to the organism and/or the locus of the function to a given brain area or neuronal population.

An investigator will, in the ideal case, aim to design an experimental paradigm that produces an instance of the kind of capacity that she intends to detect and measure. She ought to want the match between the effect she produces in the laboratory and the phenomena she takes to be grouped together under the general construct to be valid. Another way to put this is that she aims for the experimental paradigm she selected to have a high degree of "construct validity." Construct validity "is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not operationally defined."[10] It "involves making inferences from the sampling particulars of a study to the higher-order constructs they represent."[11] Experimental paradigms or cognitive tasks may have anywhere from a low to high degree of construct validity. The higher the degree of construct validity, the closer the match between the effect under study in laboratory and the cognitive phenomena designated by the construct.

The final stage of the experimental process is data interpretation (Figure 3.1a), in which the investigator determines which hypothesis of the set of competing hypotheses about the effect produced in the laboratory the statistically analyzed data discriminate. This hypothesis is then taken to be true of the effect produced in laboratory. It may then be extended back to the original effect of interest in the world that prompted the empirical question about the phenomenon of interest in the first place. However, another fundamental component of every research study that occurs either within the context of data interpretation or on the heels of it involves researchers situating their results within the broader context of the literature. These comparative practices often combine results from different laboratories to arrive at conclusions about the brain areas that underlie (cogneurosci), or the cellular and molecular mechanisms that produce (cogneurobio), a give cognitive capacity. These practices may be described as "small-scale integrative" practices.

However, there are also “large-scale integrative” practices in which investigators bring together multiple different kinds of data from different research studies, that occur most commonly in the context of (literature) review papers. One reason this is necessary is that data pertaining to the mechanisms of a cognitive capacity are supposed to be replicable across different laboratories and different investigators to rule out the possibility of experimental artifacts. Adina Roskies puts the point aptly with respect to cognitive neuroscience in claiming that “convergence across multiple experiments is the key to epistemic warrant when it comes to attributing function to anatomical regions” and “results from any single study are viewed by scientists as providing evidence for one way of parceling out one cognitive function over another, but not as conclusive evidence.”[12] A second reason why *integrative practices* are necessary for discovering the mechanisms of cognitive functions is that mechanistic explanations of a cognitive capacity will not arise in the context of a single laboratory; rather, they are based on results emanating from many laboratories situated at the same and different levels of analysis.[13]

While review papers are a primary context in which results from many different laboratories are brought together often in diagrammatic models that depict mechanisms, meta-analyses are becoming an increasingly popular tool for bringing data emanating from multiple different laboratories to bear on the question of the anatomical location of a single cognitive function in the brain[14] and the cellular and molecular mechanisms of a given cognitive function. Databases such as the *Cognitive Atlas* (http://www.cognitiveatlas.org) have been created so that specific information from research papers that use fMRI and structural MRI may be entered into the database and mined for correlations or interesting patterns in the data across research studies. The hope is that such meta-analyses will be revelatory with respect to functional localization. Some neuroscientists and philosophers of neuroscience recently have advocated for a similar approach with respect to cognitive neurobiological experiments.[15]

To appreciate the dynamical relationship between conceptual, experimental and integrative practices, it is important to note that the experimental process within any given laboratory is rarely one-shot. Oftentimes, an investigator and/or her critics wonder whether the investigative procedures she has used in the laboratory satisfy the criterion of construct validity. Such worries prompt a process known as “construct explication” (Figure 3.1b). This process may be understood in terms of a series of questions that also in the ideal case become a fundamental part of the experimental process. Specifically, an investigator asks at the relevant stages of this process:

1 Which instances of worldly phenomena should be grouped together under the concept designating the construct?
2 Which investigative strategies will yield instances that instantiate the construct?
3 Are the investigative strategies I have used adequate or should they be modified?

4 Given the data that these investigative strategies yield, should the construct itself be revised to exclude phenomena from the category that do not belong or include additional phenomena that I deemed were not part of the original category?

Historically, a primary aim of efforts in psychology and the social sciences to define the concepts of construct validity and construct explication was to improve the correspondence between theoretical constructs and the phenomena in the world they are intended to designate. Investigators engaging in the conceptual practices of construct validation and construct explication was considered the best strategy for an area of science to progress towards categories that track natural kinds. I argue that it still is. However, construct validity is not the only constraint shaping the experimental process; it is not the only desirable feature of experimental paradigms. Investigators also want to ensure that their experiments are reliable, and that the data supports claims that have *predictive validity* or *external validity*.[16] In other words, there is not one specific constraint on experimental practices, but many different possible and even competing desiderata. Investigators are at liberty to decide which is/are most fundamental.

Using the aforementioned conceptual distinctions we can identify and analyze different kinds of *conceptual*, *experimental* and *integrative* practices across laboratories, investigators, research studies and review/meta-analysis papers in the neurosciences of cognition and assess whether they are likely to promote the discovery of natural kinds of cognitive capacities. Consider, for example, *conceptual practices*. One relevant question we might ask is, Are constructs designating cognitive capacities deployed consistently across investigators? Although providing a thorough answer to this question requires assessing constructs on a case-by-case basis, there is good evidence in the literature that the usage of theoretical terms designating cognitive capacities differs across investigators.[17] In cognitive neurobiology, for example, a lexicon to broaden the meanings of terms designating different kinds of memory has been put forward because "terms in memory research are occasionally used in more than one way."[18] Broadening the class of phenomena to which a given construct refers does not, however, fix the lack of coordination across investigators with respect to how they use theoretical terms. Carrie Figdor[19] puts the point nicely in claiming that the terms used to designate kinds of cognitive capacities do not have stable meanings; even if different investigators use the same term to refer to a kind of cognitive function or a kind of experiment, it does not mean that they intended to designate the "same" cognitive function by means of the term.

We encounter a similar lack of coordination when it comes to the experimental practices involved in the study of the brain areas that underlie cognitive capacities and the cellular and molecular mechanisms productive of those capacities. If different investigators begin with different ideas about what phenomenon a given construct designates, it likely shapes the experimental strategies they use to investigate that phenomenon. In fact, experimental paradigms and the protocols used in conjunction with them to produce instances of cognitive capacities in the laboratory often do differ, sometimes in subtle ways, across investigators who claim to be studying the same phenomenon.[20] Thus, it is possible that different investigators who claim to be investigating the same phenomenon are actually investigating different phenomena produced by different mechanisms.

Additionally, not all investigators engage in the process of construct explication. In cognitive neuroscience, for example, some investigators, especially those with backgrounds in cognitive psychology, do engage in a form of construct explication called "task analysis."[21] However, many cognitive neurobiologists rely instead on "intuitive judgments regarding the cognitive processes engaged by a given task."[22] Still others are more concerned with ensuring the reliability of their experiments rather than with construct validity. Thus, construct explication is not something that currently occurs across laboratories and investigators in either a consistent or coordinated way.[23] In other words, the purported best strategy according to social scientists for developing taxonomies of kinds that correspond to real divisions in the causal structure of the world is not a common strategy in the neurosciences of cognition.

The aforementioned features of conceptual and experimental practices in the neurosciences of cognition have specific implications for *integrative practices*. Philosophers and scientists alike agree that mechanistic explanations of cognitive capacities come about exclusively via the "integration" or combination of findings emanating from many different laboratories. However, the integrative accounts or mechanistic models of cognitive capacities that we find in review papers require investigators to abstract away from specific details of the various experimental contexts in which the data being integrated into these models was produced. In other words, insofar as these models abstract away from within-lab conceptual and experimental practices, the details of which are relevant to the mechanisms and types of capacities under study in a given laboratory, these mechanistic models are at best abstract representations of the mechanisms productive of an abstract cognitive capacity. Meta-analyses have similar limitations since data-mining techniques abstract away from specific features of the experimental paradigms used to produce cognitive capacities in order to arrive at general conclusions about the neuroanatomical locations of cognitive capacities or their cellular and molecular mechanisms.

The lack of coordination in scientific practice across investigators and laboratories is suggestive that investigators working in the neurosciences of cognition are not aiming to discover natural kinds of cognitive capacities and their mechanisms. For, if they were, conceptual, experimental and integrative practices would be coordinated across investigators so as to achieve this goal. From the perspective of experimental practice, then, the neurosciences of cognition currently do not appear to be interested in identifying natural kinds or providing accurate explanatory representations of cognitive capacities that carve nature at its joints. Conceptual, experimental and integrative practices would have to change radically across investigators if such accuracy were to be the goal.[24] Thus, for the time being, there appear to be good grounds for a healthy anti-realism about the kinds of cognitive capacities and their mechanisms currently on offer in the neurosciences of cognition.

Notes

1 See: Craver, C. (2007) *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press; Craver, C. (2009) 'Mechanisms and Natural Kinds', *Philosophical Psychology*, 22:5, pp. 575–594; Kendler, K., Zachar, P. and Craver, C.F. (2011) 'What Kinds of Things Are Psychiatric Disorders', *Psychological Medicine*, 41, pp. 1143–1150; Piccinini, G. and Craver, C. (2011) 'Integrating Psychology and Neuroscience: Functional Analysis as Mechanism Sketches', *Synthese*, 183:3, pp. 283–311.
2 Kendler, K., Zachar, P. and Craver, C.F. 'What Kinds of Things Are Psychiatric Disorders', pp. 1147.
3 See: Craver, C., *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*; Craver, C.F., 'Mechanisms and Natural Kinds'.
4 See: Craver, C.F. (2014) 'The Ontic Conception of Scientific Explanation', In: Hütteman, A. and Kaiser, M. (eds.) *Explanation in the Biological and Historical Sciences*. Dordrecht, Netherlands: Springer, pp. 27–52.
5 While Craver does acknowledge conventional factors, namely the "interests and objectives" of scientists "contribute partly, but ineliminably, to the kinds of mechanisms [they] find in the world," for him, such factors merely constrain the class of MPC kinds neuroscientists actually discover. Craver, C.F., "Mechanisms and Natural Kinds", p. 592.
6 Bechtel, W. (2008) *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York: Routledge.
7 Bechtel, W., *Mental Mechanisms*, p. 18.
8 Bechtel, W., *Mental Mechanisms*, p. 19.
9 Sullivan, J. (2009) 'The Multiplicity of Experimental Protocols: A Challenge to Reductionist and Non-reductionist Models of the Unity of Neuroscience', *Synthese*, 167, pp. 511–539.
10 Cronbach, L. and Meehl, P. (1955) 'Construct Validity in Psychological Tests', *Psychological Bulletin*, 52, pp. 281–302, p. 282.
11 Shadish, W., Cook, T. and Campbell, D. (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company, p. 65.
12 Roskies, A. (2010) 'Saving Subtraction: A Reply to Van Orden and Paap', *British Journal of the Philosophy of Science*, 61:3, pp. 635–665, p. 641.
13 See for example: Dudai, Y. (2002) *Memory from A to Z: Keywords, Concepts and Beyond*. Oxford: Oxford University Press; Piccinini, G. and Craver, C., 'Integrating Psychology and Neuroscience: Functional Analysis as Mechanism Sketches', pp. 283–311.
14 See: Anderson, M. (2015) 'Mining the Brain for a New Taxonomy of the Mind', *Philosophy Compass*, 10:1, pp. 68–77.
15 Silva, A.J., Bickle, J. and Landreth, A. (2013) *Engineering the Next Revolution in Neuroscience: The New Science of Experiment Planning*. Oxford: Oxford University Press.
16 See for example: Sullivan, J., 'The Multiplicity of Experimental Protocols: A Challenge to Reductionist and Non-Reductionist Models of the Unity of Neuroscience'; Sullivan, J. (2010) 'Reconsidering Spatial Memory and the Morris Water Maze', *Synthese*, 177:2, pp. 261–283.
17 Stinson, C. (2009) 'Searching for the Source of Executive Attention', *Psyche*, 15:1, pp. 137–154; Irvine, E. (2012) *Consciousness as a Scientific Concept*. Dordrecht: Springer; Sullivan, J., 'Reconsidering Spatial Memory and the Morris Water Maze'.
18 See: Dudai, Y., *Memory from A to Z: Keywords, Concepts and Beyond*, p. 1.
19 Figdor, C. (2011) 'Semantics and Metaphysics in Informatics: Toward an Ontology of Tasks', *Topics in Cognitive Science*, 3, pp. 222–226.

20 See: Sullivan, J., 'The Multiplicity of Experimental Protocols: A Challenge to Reductionist and Non-Reductionist Models of the Unity of Neuroscience'.

21 Carter, C., Kerns, J. and Cohen, J. (2009) 'Cognitive Neuroscience: Bridging Thinking and Feeling to the Brain, and Its Implications for Psychiatry' In: Charney, D. and Nestler, E. (eds.) (3rd edition) *Neurobiology of Mental Illness*. Oxford: Oxford University Press, pp. 168–178, p. 169.

22 Poldrack, R.A. (2010) 'Subtraction and Beyond: The Logic of Experimental Designs for Neuroimaging' In: Hanson, S. J. and Bunzl, M. (eds.) *Foundational Issues in Human Brain Mapping*. Cambridge: MIT Press, p. 149.

23 See: Sullivan, J. (2014a) 'Is the Next Decade in Neuroscience a Decade of the Mind' In: Wolfe, C. (ed.) *Brain Theory: Critical Essays in Neurophilosophy*. London: Palgrave MacMillan; Sullivan, J. (2014b) *Stabilizing Mental Disorders: Prospects and Problems in Classifying Psychopathology: Mental Kinds and Natural Kinds*. Edited by H. Kincaid and J. Sullivan. Boston, MA: MIT Press.

24 See: Sullivan, J., 'Is the Next Decade in Neuroscience a Decade of the Mind'.