Published in P. Humphreys (Ed.) *Oxford Handbook for the Philosophy of Science* (2016), Oxford: Oxford University Press, pp. 92–114
doi:10.1093/oxfordhb/9780199368815.013.5

mark.sprevak@ed.ac.uk

Philosophy of the psychological and cognitive sciences

Mark Sprevak
University of Edinburgh

13 November 2017

This chapter surveys work in philosophy of the psychological and cognitive sciences. The survey is organized by the type of task that philosophers have taken on. The author focuses on four types of task: (1) how we should interpret theories in cognitive science, (2) how we should precisify theoretical concepts in cognitive science, (3) how theories or methodologies in cognitive science cohere, and (4) how cognitive states, processes, and capacities should be individuated.

1 Introduction

Philosophy of the psychological and cognitive sciences is a broad and heterogeneous domain. The psychological and cognitive sciences are rapidly evolving, fragmented, and often lacking in theories that are as precise as one might like. Consequently, philosophers of science have a challenging subject matter: fast moving, variegated, and sometimes slightly fuzzy. Nevertheless, the psychological and cognitive sciences are fertile ground for philosophers. Philosophers can bring their skills to bear with productive effect: in interpreting scientific theories, precisifying concepts, exploring relations of explanatory and logical coherence between theories, and in typing psychological states, processes, and capacities.

In this chapter, I organize work in philosophy of the psychological and cognitive science by the kind of task that philosophers have taken on. The tasks on which I focus are:

1. How should we interpret theories in cognitive science?

2. How should we precisify theoretical concepts in cognitive science?
3. How do theories or methodologies in cognitive science fit together?
4. How should cognitive states, processes, and capacities be individuated?

None of these tasks is distinctively philosophical: not in the sense of being of interest only to philosophers nor in the sense that philosophers would be the best people to solve them. All of the tasks engage a wide range of inquirers; all are carried out to the highest standards by a diverse group of individuals. What marks these tasks out as special is that they are among the problems that tend to interest philosophers and to whose solution philosophers tend to be well placed to contribute. Philosophy of the psychological and cognitive sciences is not defined by questions that set it apart from other forms of inquiry. It is characterized by questions, shared with other researchers, that tend to suit the skills, and attract the interest of, philosophers.

Three qualifications before proceeding. First, this chapter does not attempt to cover non-cognitive research in the psychological sciences (e.g., differential or social psychology). For the purposes of this chapter, 'psychological science' and 'cognitive science' will be used interchangeably. Second, the term 'theory' will be used loosely, and looser than is normal in other areas of philosophy of science. 'Theory' may refer to a full-fledged theory or a model, description of a mechanism, sketch, or even a general claim. Finally, the tasks I discuss, and the particular examples I give, only sample work in philosophy of the psychological and cognitive sciences. I focus on a small number of cases that I hope are illustrative of wider, structurally similar projects. The examples are not intended to be a list of the best work in the field, but only examples by which to orientate oneself.

Let us consider the four tasks in turn.

2 Interpreting

A crude way to divide up scientific work is between *theory building* and *theory interpretation*. Theory building involves identifying empirical effects and coming up with a theory to predict, test, and explain those effects. Theory interpretation concerns how – granted the empirical prowess or otherwise of a theory – we should understand that theory. Interpretation involves more than assigning a semantics. The job of interpretation is to understand the import of the theory: What is the purpose of the descriptions and physical models used to express the theory? What are the criteria of success for the theory: truth, empirical adequacy, instrumental value, or something else? Which terms in the theory are referring terms? Which kinds of ontological commitments does a theory entail? Which aspects of a theory are essential?

Philosophers have tools to help with theory interpretation. Analogues to the questions above have been pursued for ethical, normative, mathematical, and other scientific discourses. A range of options have been developed concerning theory interpretation, amongst which are versions of realism and instrumentalism. An interpretation sets a combination of semantic, pragmatic, and ontological parameters regarding a theory. Our job is to see which setting results in the best interpretation of a psychological theory.

These concerns have recently played out with Bayesian models of cognition. Bayesian models predict and explain an impressive array of human behavior. For example, Bayesian models provide a good, predictive model of human behavior in sensory cue integration tasks (Ernst and Banks, 2002). In these tasks, subjects are presented with a single stimulus in two different sensory modalities (say, touch and vision) and asked to make judgments about that stimulus by combining information from the two modalities. For example, subjects may be presented with two ridges and asked to decide, using touch and vision, which ridge is taller. Ernst and Banks found that subjects' behavior could be predicted if we assume that the input to each sensory modality is represented by a probability density function and that these representations are combined using the Bayesian calculus to yield a single estimate. Probability density functions and Bayesian computational machinery do a good job of predicting human behavior in sensory cue integration. Bayesian models also appear to explain human behavior. This is because they tie human behavior to the optimal way in which to weigh evidence. Many aspects of human behavior have been modeled with this kind of Bayesian approach, including causal learning and reasoning, category learning and inference, motor control, and decision making (Pouget et al., 2013).

How should we interpret these Bayesian models?

One option is *realism*. This is known within psychology as the *Bayesian brain hypothesis*. Here, the central terms of Bayesian models – the probability density functions and Bayesian computational methods – are interpreted as picking out real (and as yet unobserved) entities and processes in the human brain. The brain 'represents information probabilistically, by coding and computing with probability density functions or approximations to probability density functions,' and those probabilistic representations enter into Bayesian, or approximately Bayesian, inference via neural computation (Knill and Pouget, 2004, p. 713). Bayesian models are *both* a theory of human behavior *and* of the neural and computational machinery that underpin the behavior. Realism about Bayesian models is usually qualified with the claim that current Bayesian models of cognition are only approximately true. What current Bayesian models get right is that the brain encodes information probabilistically and that it implements some form of approximate Bayesian inference.

The precise format of the probabilistic representations and the precise method of approximate Bayesian inference is left open to future inquiry (Griffiths et al., 2012).

Another interpretation option for Bayesian models is *instrumentalism* (Bowers and Davis, 2012; Colombo and Seriès, 2012; Danks, 2008; Jones and Love, 2011). According to the instrumentalist, Bayesian models do not aim to describe the underlying neurocomputational mechanisms (other than providing general constraints on their inputs and outputs). The central terms of Bayesian models – probability density functions and Bayesian computational machinery – should be understood, not as referring to hidden neural entities and processes, but as formal devices that allow experimenters to describe human behavioral patterns concisely. The instrumentalist allows that the underlying neural mechanisms could be Bayesian. But this possibility should be distinguished from the content of current Bayesian models. The aim of those models is to predict behavior. The success of Bayesian models in predicting behavior is not evidence that the mechanisms that generate that behavior are Bayesian. One might be inspired by the success of Bayesian models in predicting behavior to entertain the Bayesian brain hypothesis. But inspiration is not evidential support. Bayesian models should be understood as aiming at behavioral adequacy. Their aim is to predict behavior and specify human behavioral competences, not to describe neural or computational mechanisms.

Here we have two incompatible proposals about how to interpret Bayesian models of cognition. The difference between the two proposals matters. Do Bayesian models tell us how our cognitive processes work? According to realism, they do. According to instrumentalism, they do not (or, at least, they only provide constraints on inputs and outputs). How do we decide which interpretation option is correct?

The primary rationale for instrumentalism is epistemic caution. Instrumentalism makes a strictly weaker claim than realism while remaining consistent with the data. There appear to be good reasons for epistemic caution too. First, one might worry about underdetermination. Many non-Bayesian mechanisms generate Bayesian behavior. A lookup table, in the limit, can generate the same behavior as a Bayesian mechanism. Why should we believe that the brain uses Bayesian methods given the vast number of behaviorally indistinguishable, non-Bayesian, alternatives? Second, one might worry about the underspecification of mechanisms in current Bayesian models. The Bayesian brain hypothesis is a claim about neurocomputational mechanisms. There are a huge number of ways in which a behaviorally-adequate Bayesian model could be implemented, both neurally and computationally. Current Bayesian models tend to be silent about their neural or computational implementation in actual brains. Absent specification of the neurocomputational implementation, we should most charitably interpret current Bayesian theories as simply not making a claim about neural or computational mechanisms at all.

What reasons are there for realism? A common inductive inferential pattern in science is to go beyond an instrumentalist interpretation if a theory has a sufficiently impressive track record of prediction and explanation (Putnam, 1975). Arguably, Bayesian models do have such a track record. Therefore, our interpretation of Bayesian models of cognition should be realist. The burden on the instrumentalist to show otherwise: to show that the brain does *not* use probabilistic representations or Bayesian inference. And current scientific evidence gives us no reason to think that the brain does not use Bayesian methods (Rescorla, 2016).

The disagreement between the realist and the instrumentalist is not epiphenomenal to scientific practice. The choice one makes affects whether, and how, experimental results bear on Bayesian models. For example, if instrumentalism is correct, then no neural evidence could tell in favor (or against) a Bayesian model. The reason is straightforward: the models are not in the business of making any claim about neural implementation, so there is nothing in the model for neural evidence to contest. If realism about Bayesian models is correct, then neural evidence *is* relevant to confirming the Bayesian models. If there is neural evidence that the Bayesian model's probability distributions or methods occur in the brain, then that is evidence in favor of the model. If there is evidence against, that is evidence against the Bayesian model. Bayesian models are evaluated not only by their fit to behavioral data, but also by their neural plausibility.

Bowers and Davis (2012) object that Bayesian theories of cognition are trivial or lacking empirical content. Their objection is that almost any dataset could be modeled as the output of some or other Bayesian model. Specific Bayesian models could be confirmed or disconfirmed by empirical data, but the general Bayesian approach is bound to succeed no matter what. Bayesianism is no more confirmed by behavioral data than number theory is confirmed. If the instrumentalist is right, then Bowers and Davis's objection has bite. Some or another Bayesian model will always fit a behavioral dataset. However, if the realist is right, then it is no longer clear that Bowers and Davis's objection succeeds. Realism raises the stakes of Bayesianism. This opens up the possibility that Bayesianism could be subject to empirical test. Precisely how to test it is not yet obvious. The reason is that currently there are no agreed proposals about the neural implementation of Bayesian models' theoretical posits (probability density functions and Bayesian computational processes). Nevertheless, realism at least opens the door to the possibility of testing Bayesianism. Suppose one were to have a theory of neural implementation in hand. If the brain's measured neural representations and computations – identifiable via the implementation proposal – really do have the properties ascribed by Bayesianism, then the general Bayesian approach would be vindicated. If not – for example, if the brain turns out to employ non-probabilistic representations or to manipulate its representations via a lookup table – then Bayesianism about cognition would

be found to be false. A theory of implementation plus neural evidence allows Bayesianism about cognition to be tested.

Interpreting a theory requires making decisions about the theory's goals (truth vs. instrumental accuracy) and how to interpret its theoretical terms (referring vs. formal devices). A final aspect of interpretation is the decision about *which* claims a theory includes. Which parts of a theory are essential, and which are explication or trimming?

Suppose that realism about Bayesian models is correct and therefore that the brain manipulates probabilistic representations. What does a Bayesian model require of these representations? Some things are clearly required: the representations must be probabilistic hypotheses, and they must encode information about the uncertainty of events as well as their truth conditions or accuracy conditions. But to *which* entities and events do these probabilistic hypotheses refer? Do they refer to distal objects in the environment (e.g., tables and chairs) or to mathematical entities (e.g., numerical values of a parameter in an abstract graphical model). In other words, does the correct interpretation of a Bayesian model of cognition make reference to distal objects in the organism's environment, or is the correct interpretation entirely formal and mathematical?

On either view, entertaining a probabilistic hypothesis could enable an organism to succeed. On the first, 'distal content', option, this could be because the probabilistic hypotheses are the organism's best guess about the state of its environment, and that appears to be useful information for an organism to consider when deciding how to act in its environment (Rescorla, 2015). On this view, neural representations of distal environment stimuli would be an essential part of a Bayesian story, not an optional extra. If this is correct, it would mean that Bayesianism about cognition is incompatible with eliminativism or fictionalism about neural representations of distal objects (Keijzer, 1998; McDowell, 2010; Sprevak, 2013).

On the second, 'mathematical content', option, the organism could reliably succeed because the probabilistic hypotheses describe a mathematical structure that is adaptive for the organism to consider in its environment. This need not be because the mathematical structure represents distal properties in the environment. There are other ways than representation in which inferences over a formal structure could reliably lead to successful action. Formal properties may track contingent nomological connections between the organism's environment, body, and sensory system. Consider that it is adaptive for an organism to consider the result of applying a Fourier transform to the values of incoming signals from the auditory nerve, or for an organism to consider the results of applying $\nabla^2 G$ to the values of the incoming signal from its retina. These are useful transformations for an organism to consider even if neither is a representation of a distal environmental property.

Egan (2010) argues for a 'mathematical content' interpretation of classical computational models in cognitive science. Her reasoning could be extended to Bayesian models. According to such a view, it is adaptive for an organism to consider the mathematical relations described by the Bayesian model even though the terms in that model do not represent distal properties or events in the environment. On this view, representations of distal objects would not be an essential part of a Bayesian story. Distal content could feature in a Bayesian model of cognition, but it is not an essential part of such a model.

My intention here is not to argue for one interpretation rather than another. My intention is only to illustrate that each theory in cognitive science requires interpretation. There are entailments between theory interpretation options. Just as in other areas of science, some aspects of psychological theories should be understood as essential, literal, fact stating, ontologically committing, whereas other aspects play a different role. It takes care and sensitivity to hit on the correct interpretation, or even to narrow the interpretation options down. This cannot be achieved simply by appealing to the utterances of theory builders because those utterances themselves need interpretation. Theories do not wear their interpretation on their sleeves.

3 Precisifying

Imprecision is bad, not only for its own sake, but also because it permits fallacious inference. If one did not know the difference between the two senses of 'bank,' one might wrongly infer that since a river has two banks, a river would be a good place to conduct a financial transaction. That no such confusion occurs reflects that the relevant precisification is known to every competent speaker of the language. Unfortunately, the same is not true of every term in psychology. The correct usage of many terms in psychology – 'consciousness', 'concept', 'module' – is murky, even for experts. One task to which philosophers have contributed is to clarify our theoretical terms so that those terms better support reliable, non-trivial, inductive inference. This may involve distinguishing between different things that fall under a term, redefining a term, or sometimes removing a term entirely. This is not mere semantic busy work. Concepts are the building blocks of scientific theories. Fashioning precise and inductively powerful concepts is essential to scientific progress (for more on this, see Machery (ms)).

I discuss here two examples of concepts that philosophers have helped to precisify: the concept of consciousness and the concept of a cognitive module.

The term 'consciousness' has its origin in folk use. We might say, 'she wasn't conscious of the passing pedestrian,' 'he was knocked unconscious in the boxing ring,' or speak of the 'conscious' experience of smelling a rose, making love, or hearing a symphony,

making life worth living. Scientific and philosophical work on consciousness only started to make progress when it distinguished different things that fall under the folk term. Precisifying the concept of consciousness has enabled researchers to have a fighting chance to discover the purpose, functional mechanism, and neural basis of consciousness.

A preliminary precisification of consciousness is *arousal*. When we say that someone is conscious, we might mean that she is alert and awake; she is not asleep or incapacitated. When in dreamless sleep, or in a pharmacological coma, a person is unconscious. This sense of 'consciousness' is usually accompanied by the assumption that the predicate '… is conscious' is a monadic predicate. Someone is simply conscious or unconscious; they need not be conscious *of* something specific. Someone may also be conscious in the sense of being aroused without being capable of consciousness in the sense of being aware of particular stimuli. Patients in a vegetative state show sleep-wake cycles, hence arousal, but they are not aware of particular stimuli. The neural mechanisms that govern consciousness-as-arousal also appear distinct from those that govern consciousness-as-awareness. Arousal is regulated by neural systems in the brainstem, notably, the reticular activating system. In contrast, the neural basis of consciousness-as-awareness appears to be in higher cortical regions and their subcortical reciprocal connections. Consciousness-as-arousal and consciousness-as-awareness have different purposes, relate to different aspects of functional cognitive architecture, and have different neural implementations (Laureys et al., 2009).

Ned Block's concept of access consciousness is one way to further precisify consciousness-as-awareness (Block, 1995). A mental representation is defined as access conscious if and only if 'it is poised for free use in reasoning and for direct "rational" control of action and speech' (Block, 1995, p. 382). One indicator of access consciousness is verbal reportability – whether the subject can say he or she is aware of a given mental episode. Reportability is, however, neither necessary nor sufficient for access consciousness. Block's precisification of consciousness-as-awareness highlights a number of other properties of consciousness-as-awareness. First, access consciousness is attributed with a relational predicate: an individual is conscious *of* something. Second, the object of that individual's consciousness is determined by a representation encoded in their brain. Third, access consciousness requires that this representation be 'broadcast widely' in their brain: it should be available to central reasoning processes and able to cause a wide variety of behavior, including verbal report. Fourth, the representation need not have actual behavioral effects; it need only have the disposition to cause appropriate behavioral effects. This catalogue provides a partial functional specification of consciousness-as-awareness. Empirical work has focused on identifying which, if any, neural properties answer to this description (Baars, 1997; Dehaene and Changeux, 2004).

Despite its virtues, there are idiosyncrasies in Block's precisification of consciousness-as-awareness: Why is rational control necessary for access consciousness? What does it mean for a neural representation to be 'poised' to have effects but not actually have them? What does it mean for a representation to 'directly' control behavior given that all control of behavior is mediated by other neural systems? The best way to see Block's account is as a stepping stone along the way to an appropriate concept of consciousness-as-awareness.

Forging the right notion of consciousness-as-awareness is not a task for armchair reflection. Precisification does not proceed prior to, or independently of, empirical inquiry. Precisification involves a two-way interaction between empirical hypotheses and consideration of how changes to the concepts that make up the empirical hypotheses would better capture the patterns relevant to scientific psychology. Precisification of a concept must be informed by the utility of the resulting concept to scientific practice. A precisification of consciousness-as-awareness proves its worth by whether the way it groups phenomena pays off for achieving goals in scientific psychology. Groupings that yield reliable inductive inferences or explanatory unification are those that should be favored. For example, a precisification of consciousness-as-awareness should aim to pick out shared facts about purpose, cognitive functional architecture, and neural implementation of consciousness-as-awareness. Recent work suggests that consciousness-as-awareness should be split into smaller concepts as no one concept meets all three of these conditions (Dehaene, Changeux et al., 2006; Koch and Tsuchiya, 2006).

The concepts of consciousness-as-arousal and consciousness-as-awareness are distinct from the concept of phenomenal consciousness. The concept of phenomenal consciousness picks out the qualitative feel – 'what it is like' – associated with some mental episodes. It feels a certain way to taste chocolate; it feels a certain way to taste mint; and those two feelings are different. Phenomenal consciousness is characterized purely ostensively and from a subjective, first-person point of view. Consider your mental life, pay attention to the qualitative feelings that accompany certain episodes – those are phenomenally conscious feelings. Given that our concept of phenomenal consciousness is rooted in first-person reflection, it is not surprising that this concept has proven hard to place in relation to scientific concepts related to brain function. Dehaene (2014) suggests that the concept of phenomenal consciousness should for the moment be set aside in pursuing a science of consciousness; phenomenal consciousness is not currently amenable to scientific explanation.

In his introduction to modularity, Fodor (1983) lists nine features that characterize a cognitive module: domain specificity, mandatory operation, limited central accessibility, fast processing, informational encapsulation, shallow outputs, fixed neural architecture, characteristic and specific breakdown patterns, and characteristic onto-

genetic pace and sequencing. Whether a mechanism counts as a module depends on whether it meets a weighted sum of these features to an 'interesting' extent (Fodor, 1983, p. 37). What counts as interesting, and how different features from the list should be weighed, is left largely unspecified. As one might imagine, there is room for precisifying the concept of modularity in different ways. Fodor claimed that the nine listed properties typically co-occur. Subsequent work has shown that they do not. Moreover, even if they did co-occur, their co-occurrence would not necessarily be of interest to scientific psychology (Elsabbagh and Karmiloff-Smith, 2006; Prinz, 2006). The concept of a psychological module has since been precisified in different ways and for different purposes. This has been done by giving priority to different properties associated with modularity from Fodor's list.

Fodor himself gives highest priority to two properties from the list: domain specificity and informational encapsulation (Fodor, 2000). These properties are distinct. Domain specificity is a restriction on the inputs that a mechanism may receive. A mechanism is domain specific if only certain representations turn the module on, or are processed by the module. For example, in the visual system, a domain-specific module might only process information about retinal disparity and ignore everything else. Informational encapsulation is different. A mechanism is informationally encapsulated if, once the mechanism is processing an input, the information that the mechanism may then draw on is less than the sum total of information in the cognitive system. For example, in the visual system an informationally encapsulated module that processes information about retinal disparity might not be able to draw on the system's centrally held beliefs. Illusions like the Müller–Lyer illusion appear to show that the visual system is, to some extent, informationally encapsulated.

This gets us in the right ball park for precisifying the concept of a cognitive module, but domain specificity and information encapsulation need to be more carefully characterized. Informational encapsulation requires that something like the following three further conditions be met (Samuels, 2005). First, the informational encapsulation should not be short-lived; it should be a relatively enduring characteristic of the mechanism. Second, informational encapsulation should not be the product of performance factors, such as fatigue, lack of time, or lapses in attention. Third, informational encapsulation need not shield the mechanism from every external influence. Processing may, for example, be affected by attentional mechanisms (Coltheart, 1999). Informational encapsulation requires 'cognitive impenetrability.' Roughly, this means that although the mechanism may be modulated by certain informational factors (e.g., attention), it cannot be modulated by others (e.g., the high-level beliefs, goals, or similar representational states of the organism). Pinning this down more precisely requires work (see Machery (2015); Firestone and Scholl (2016)).

Both domain specificity and informational encapsulation admit of degrees. Not just any step down from complete informational promiscuity produces domain specificity or information encapsulation. Moreover, the step down is not merely numerical, but also a matter of kind. The input domain and the informational database should be, in some sense, unified. One should be able to characterize the mechanism as a module for X, where X is some task or process that makes sense as a single unit in the light of concerns about the purpose and cognitive architecture of the organism. Illuminating the precise nature of this constraint on modularity is non-trivial.

The concepts of domain specificity and informational encapsulation offer scope for the development of a palette of distinct precisifications of the concept of a cognition module. If one turns attention to other properties associated with modularity in Fodor's list, more scope for divergent precisifications emerges. One could think of Fodor's criteria as conceptually separable parameters that are important in theorizing about the mind or brain (Elsabbagh and Karmiloff-Smith, 2006). Some criteria may be more important for capturing certain kinds of pattern – computational, anatomical, developmental, and so on – than others. Which kind of pattern one wishes to capture depends on the interests of the investigator. Different concepts of modularity may include various combinations of these criteria. How we should separate out and sharpen the parameters of Fodor's characterization into one or more useful working concepts depends on the kinds of pay-off previously described. A precisification of modularity should help inquirers achieve their scientific goals. It should allow us to capture empirical patterns relevant to scientific psychology. Achieving this requires a two-way interaction between empirical inquiry and reflection on how changes to the concepts that make up the empirical hypotheses would allow us to better pick up on significant empirical patterns.

Before moving on, it is worth noting that there is also benefit in having imprecise concepts. Lack of precision in one's concepts may sometimes result in a fortuitous grouping of properties by a hypothesis. And it is not always bad for different research teams to be working with different understandings of theoretical concepts or for their inquiry to proceed with underspecified concepts. The promiscuous inferences that result may sometimes be helpful in generating discovery. Nevertheless, *pace* heuristic benefits, at some stage we need to be aware of exactly which claims are being made, and to understand when a conflict between say, two research teams, is genuine or merely verbal. Eventually, it is in everyone's interest to precisify.

4 Understanding how things hang together

Sellars described the task of philosophy as 'to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term' (Sellars, 1963, p. 1). One way to do this is to understand the explanatory and logical relations between theories and the ways in which they offer, or fail to offer, each other epistemic support. In this section, we examine two ways in which this is done in philosophy of the psychological and cognitive sciences. First, we examine the relationship between computational and dynamical systems theories in the psychological sciences. Second, we examine the relationship between different levels of inquiry in the psychological sciences.

Computational theories and dynamical systems theories both attempt to explain cognitive capacities. Both aim to explain how, in certain circumstances, we are able to do certain tasks. However, the two theories appear to explain this in different ways.

According to the computational approach, a cognitive capacity should be explained by giving a computational model of that capacity. Pick a cognitive capacity – for example, the ability to infer the three-dimensional shape of an object from information about its two-dimensional shading. An advocate of the computational approach might offer a computation that is able to solve this problem and suggest that this computation, or something like it, is implemented in the brain and causally responsible for the capacity in question (Lehky and Sejnowski, 1988). Computational explanations are characterized by appeal to subpersonal representations and formal transformations by mechanisms built from simple components.

Dynamical systems theorists also aim to explain cognitive capacities, but their explanations appear to work differently. A dynamical systems theory involves differential equations relating variables that correspond to abstract parameters and time. The dynamical systems theorist claims that these parameters also correspond to some aspects of neural and bodily activity. Differential equations describe how these parameters interact over time to generate the behavior. Dynamical systems theory explains a cognitive capacity by narrowing in on a dynamical property of the brain or body that is causally responsible for the capacity (Schöner, 2008).

Both computational theories and dynamical systems theories have had various successes in explaining cognitive capacities. Aspects of decision-making, such as production of the A-not-B error in infants – the phenomenon of an infant persevering in reaching for box A even though she just saw the experimenter place a desired toy in box B – are well modeled by dynamical systems theory (Thelen et al., 2001). Other cognitive capacities, such as those involved in inferring three-dimensional shape from two-dimensional information recorded by the retina, are well modeled

by computation (Qian, 1997). Our question is: How these two theories relate? If psychology employs both, how do the theories fit together?

On the face of it, they appear to be rivals (van Gelder, 1995). Each seems to instantiate a rival bet about the nature of the mind: either a cognitive capacity is produced by a computation or it is produced by a dynamical causal relation.

On reflection, this seems too strong. The computational and dynamical systems approach agree on a great deal. They agree that the brain is a dynamical system. It is no part of a computational theory to deny that cognitive capacities could, or should be, explained by a time-based evolution of physical parameters; indeed, the computational approach proposes one class of paths through the space of physical parameters. Computational theories might appear to differ from dynamical systems theories in that only computational theories employ subpersonal representations. However, when explaining a psychological capacity, there are often reasons – independent of the decision to use a computational or dynamical systems theory – to introduce subpersonal representations (Bechtel, 1998; van Gelder, 1995, p. 376). Some psychological capacities are 'representation hungry'; for example, some capacities require the system to keep track of absent stimuli (Clark and Toribio, 1994). Explaining these capacities motivates the introduction of subpersonal representations, no matter whether one places those representations in a computational or dynamical systems context. Furthermore, it is unclear whether subpersonal representations are really a necessary commitment of a computational approach. Not every state in a computation needs to be representational. It is an open question how much, if any, of a mechanism need have representational properties in a computational approach (Piccinini, 2008). Appeal to subpersonal representations does not show that computational and dynamical systems theories are incompatible.

Dynamical systems theory appears to differ from a computational approach in that dynamical models give a special role to time. Dynamical models offer descriptions in which time is continuous and the transitions between states are governed by explicitly time-based differential equations. Computational models – for example, Turing machines – use discrete time evolution and are governed by rules that do not mention time. However, not all computational models are like Turing machines. Some computational models involve continuous time evolution and have time-based differential equations as their transition rules. Within this class are connectionist models (Eliasmith, 1997) and more recent realistic computational models of neural function (Eliasmith, 2013). These computational models assume continuous time evolution, contain parameters that map onto a subset of neural and bodily properties, and use, as their transition rules, differential equations that specify how the parameters evolve over time. These models have all the signature properties of dynamical systems theories.

The distance between computational theories and dynamical systems theories is not as great as it may first appear. In some cases, the two theories converge on the same class of model. This is not to say that every dynamical systems theory is a computational theory or vice versa. It is rather that, in the context of explaining our cognitive capacities, a computational approach and a dynamical systems approach may converge: they may employ the same elements related in the same way – representations, abstract parameters, mapping from those parameters to neural and bodily properties, continuous time evolution, and time-based differential equations. The right way to see the relationship between a computational theory and a dynamical systems theory is not as two rivals, but as two possibly compatible alternatives.

This does not explain how classical, discrete, symbolic-rule-governed computational models relate to the dynamical systems approach. One suggestion is that the latter reduce to the former as a limiting case (for an attempt to show this, see Feldman (2012)). And to understand how things hang together generally across theories in the psychological sciences, one needs to understand the relationship between computational theories, dynamical systems theories, and a wide range of other theories: statistical, Bayesian, enactivist, and others. Under which conditions are two theories rivals, compatible alternatives, or do they reduce to one another? Under which conditions do they offer each other epistemic support?

A second way in which to understand how things hang together in the psychological sciences is to understand how investigation at different levels of inquiry coheres. What is the relationship between inquiry at different levels in the psychological sciences?

Here, we focus on only two kinds of level: levels of spatial organization inside mechanisms and Marr's levels of computational explanation.

First, let us consider levels of spatial organization inside mechanisms. Different mechanisms exist at different spatial scales. 'Higher' level mechanisms involve larger systems and components. 'Lower' level mechanisms involve smaller systems and components. We normally explain by describing mechanisms at several levels. For example, we explain the cognitive capacity that a mouse exhibits when navigating a maze to find a reward by appeal to mechanisms inside and around the mouse at multiple spatial scales. The top-level mechanism is the entire mouse engaged in a spatial navigation task. Within this mechanism is a mechanism that involves part of the mouse's brain, the hippocampus, storing a map of locations and orientations inside the maze. Within this mechanism, inside the mouse's hippocampus, is a smaller mechanism storing information by changing weights in the synapses between pyramidal cells. Within this mechanism is a mechanism that produces long-term changes at a synapse by modifying the synapse's N-methyl-D-aspartate (NMDA) receptors. Those NMDA receptors undergo change because of yet smaller

mechanisms governing their functioning. A mechanistic explanation of the mouse's cognitive capacity involves appeal to multiple mechanisms at multiple spatial scales and showing how they work in concert to produce the cognitive capacity (Craver, 2007). The relationship between higher and lower levels goes beyond the mere mereological whole–part relation. Higher level mechanisms are not just part of, but also are *realized* by lower level mechanisms. Lower level mechanisms form the component parts of higher level mechanisms. Cognitive capacities are explained by showing how mechanisms at different spatial scales, integrated by the mechanistic realization relation, produce the cognitive capacity in question.

Descriptions of mechanisms at different levels of spatial scale have a degree of autonomy from each other. A description of a mechanism at one spatial level may be silent about how that mechanism's component parts work, or about the larger system in which the mechanism is embedded. One might, for example, describe how cells in the mouse hippocampus store a map of locations while remaining silent about the lower level mechanism that produces synaptic change. One might also describe how cells in the hippocampus store a map of locations while remaining silent about how those cells are recruited by the entire mouse to solve the task.

This partial autonomy between descriptions at different spatial levels is not full-blown independence. The autonomy arises because mechanistic realization allows for the (logical) possibility of multiple realization. It is possible for the component parts of a mechanism to be realized in multiple ways, within the constraints that the performance of the mechanism dictates. It is also possible for the mechanism to be embedded in multiple larger contexts, within the constraints that the context should support the operation of the mechanism. Description at a particular level of spatial scale places some constraints on higher or lower level descriptions, but leaves some degree of freedom. This degree of freedom allows different teams of inquirers to focus on discovering mechanisms at different spatial scales in a partially autonomous manner. It allows scientific inquiry in the psychological sciences to split into different fields: biochemistry, cellular biology, neurophysiology, cognitive neuroscience, and behavioral ecology.

Oppenheim and Putnam (1958) claimed that scientific disciplines are structured into autonomous levels corresponding to the spatial scale of their subject matter. We are now in a position to see what is right about this idea. There is no necessity for scientific inquiry to be structured by levels of spatial scale. Nevertheless, structuring scientific inquiry in the psychological sciences by spatial scale is permissible. It is permissible because the ontologically layered structure generated by the mechanistic realization relation, and the partial autonomy between levels of spatial scale that this provides, allows scientific inquiry to proceed at different spatial scales along relatively separate tracks. The structuring of scientific disciplines by levels of spatial

scale that Oppenheim and Putnam describe is a consequence of the ontological layering, and partial autonomy, generated by the mechanistic realization relation.

Marr (1982) introduced a distinct kind of level of inquiry into the psychological sciences. Marr argued that the psychological and cognitive sciences should be divided into three levels of explanation.

Marr's first level is the 'computational' level. The aim of inquiry at the computational level is to describe *which* task an organism solves in a particular circumstance and *why* that task is important to the organism. A task should be understood as an extensional function: a pattern of input and output behavior. In order to discover which function an organism computes, we need to understand the ecological purpose of computing that function – why the organism computes this function and what computation of this function would allow the organism to achieve. Without a guess as to the ecological purpose of computing this function, there would be no way of picking out from the vast number of things that the organism does (its patterns of input–output behavior) which are relevant to cognition.

Marr's second level is the 'algorithmic' level. The aim of inquiry at the algorithmic level is to answer *how* the organism solves its task. The answer should consist in an algorithm: a finite number of simple steps that take one from input to output. Many different algorithms compute the same extensional function. Therefore, even if we were to know which extensional function the organism computes, the algorithm would still be left open. In order to discover which algorithm an organism uses, researchers look for indirect clues about the information-processing strategies exploited by the organism, such as the organism's reaction times and susceptibility to errors.

Marr's third level is the 'implementation' level. The aim of inquiry at the implementation level is to describe how steps in the algorithm map onto physical changes. Even if we were to know both the extensional function and the algorithm, that would still leave open how the algorithm is implemented in physical changes of the organism. The brain is a complex physical system. Without some guide as to which parts of the brain implement which parts of an algorithm, there would be no way to know how the brain enables the organism to solve its task. The implementation level identifies which physical parts are functionally significant: which parts are relevant, and in which ways, to the computation that the organism performs. In the case of an electronic PC, electrical changes inside the silicon chips are functionally significant; the color of the silicon chips or the noise the cooling fan makes are not. Researchers look for implementation level descriptions by using techniques such as magnetic resonance imaging, electroencephalograms, single-cell recording, and testing how performance is affected when physical resources are damaged (e.g., by stroke) or temporarily disabled (e.g., by drugs).

Marr's three levels are not the same as the levels of spatial organization in mechanisms described previously. The levels of spatial organization in mechanisms involve positing an ontological layering relation: psychological mechanisms are related smaller to larger by mechanistic realization; component parts are realized by increasingly smaller mechanisms. One consequence of this ontological layering is that scientific inquiry at different spatial scales can be structured into partially autonomous domains (from biochemistry to behavioral ecology). Structuring scientific inquiry into levels is a consequence, not the principal content, of the claim. In contrast, the principal content of Marr's claim is that scientific inquiry, not ontology, should be structured. Marr divides work in the psychological sciences into three types of inquiry: computational, algorithmic, and implementational. There is no assumption that this division is accompanied by a division in the ontology. Marr's claim is simply that the psychological sciences should pursue three types of question if they are to explain psychological capacities adequately.

Computational, algorithmic, and implementational questions concern a single system at a single level of spatial scale. They also concern a single capacity: which extensional function that capacity instantiates, which algorithm computes that function, and how that algorithm is physically implemented. In contrast, each level of the mechanistic scale concerns a different physical system: the entire organism, the brain, brain regions, neural circuits, individual neurons, and subcellular mechanisms. Each level of scale concerns a different capacity: the capacity of the whole organism to navigate a maze, the capacity of the hippocampus to store spatial information, the capacity of pyramidal synapses to undergo long-term potentiation (LTP), and so on. At each level of spatial scale, and for each corresponding capacity and system, one can ask Marr's questions: Which function does the physical system compute and why? Which algorithm does it use to compute this function? How is that algorithm implemented by physical changes in the system? Marr's questions cut across those concerning mechanisms at different levels of spatial scale.

Marr claimed a degree of autonomy, but not full independence, between his levels. The autonomy that exists between Marr's levels derives from two properties of computation. First, the same extensional function can be computed by different algorithms. Second, the same algorithm can be implemented in different ways. The first property means that proposing a particular function at the computational level does not restrict algorithmic-level inquiry to a particular algorithm. The second property means that proposing a particular algorithm at the algorithmic level does not restrict implementation-level inquiry to a particular physical implementation. Similar degrees of freedom do not hold in reverse. If one were to propose a particular implementation – a mapping from physical activity in the system to steps of some algorithm – then the algorithm that the system employs would be thereby fixed. Similarly, if one were to propose that the system uses a particular algorithm,

then the extensional function that the system computes would be thereby fixed. The autonomy between Marr's levels is downwards only: lower levels are partially autonomous from upper levels, but not vice versa.

Downwards autonomy in Marr's scheme, like the autonomy between levels of inquiry about mechanisms at different spatial scales, is only present as a logical possibility. The degree of autonomy is likely to be attenuated in practice. Downwards autonomy for Marr derives from the two logical properties of computation described earlier. However, the psychological sciences are constrained by more than what is logically possible. Their concern is what is reasonable to think given all we know about the brain and agent. The numbers of permissible algorithms and implementations are likely to be significantly less than those that are logically possible when constraints are added about the resources that the brain can employ, the time the system can take to solve its task, assumptions made in other areas of the psychological sciences are taken into account.

The autonomy between Marr's levels of inquiry is different from that between levels of inquiry concerning mechanisms at different spatial scales. The autonomy between Marr's levels of inquiry derives from two properties of computation: that many algorithms compute the same function and that there are many ways to implement the same algorithm. The autonomy between levels of inquiry concerning mechanisms at different spatial scales derives from two properties of the mechanistic realization relation: that it is possible for different mechanisms to produce the same causal power, and that the same mechanism could be embedded in different contexts. The latter two properties give rise to an in principle upward and downward autonomy between levels of spatial scale. Positing a mechanism at a particular level of spatial scale does not fix how its smaller component parts work, nor does it fix how that mechanism is embedded in a larger system. Autonomy between levels of inquiry concerning mechanisms at different spatial scales is bi-directional. This is formally distinct from the downwards-only autonomy associated with Marr's levels of computational explanation.

5 Individuating

Disagreements within philosophy of the psychological sciences are sometimes not about the causal flow involved in cognition, but how to individuate that causal flow. Two philosophical camps may agree on the basic causal relations, but disagree about which of the elements in the causal structure are cognitive, representational, perceptual, sensory, doxastic, gustatory, olfactory, and so on. These disagreements may give outsiders the appearance of being merely verbal. But this is rarely the case. The competing sides agree on the meanings of their words. What they disagree

about is how cognitive capacities, processes, and states should be individuated.

In this section, I look at two examples of this kind of dispute. The first is the disagreement about how to individuate the senses. The second is the disagreement about whether human mental states and processes extend outside our brains and bodies.

The senses are different ways of perceiving, such as seeing, hearing, touching, tasting, and smelling. What makes two senses different? How many senses are there? Under which conditions would an organism have a new sense? The psychological sciences provide data that appear to challenge folk ideas about the senses. Novel sense modalities seem to exist in non-human animals, including magnetic senses, electric senses, infrared senses, and echolocation. Humans seem to have senses for pressure, temperature, pain, balance, and their internal organs in addition to their traditional five senses. Neural processing of sensory information in humans is multimodal; visual areas in the brain are not exclusively visual and integrate information from sound and other stimuli. Blindsight patients appear to have vision without associated visual phenomenology or consciously accessible beliefs. Tactile-visual sensory substitution (TVSS)-equipped patients appear to see via touch. Based on this information, should we revise the folk view that humans have five senses? If so, how? In order to answer this question, we need a way to individuate the senses. Let us look at four contenders.

The first is representation-based. Suppose that each sense has an object or property that is exclusively detected and represented by that sense – its 'proper sensible'. The proper sensibles of hearing, tasting, smelling, and seeing are sound, flavor, odor, and color respectively. According to the representation-based view, the representations of these proper sensibles – representations that are not generated in any other way – individuate the senses. A sense is individuated by the characteristic representations that the sense produces. A challenge for the view is that it lands us with a new problem: How do we individuate the sensory representations? It is not clear that this is substantially easier than the original problem of how to individuate the senses.

The second approach is experience-based. Hearing, tasting, smelling, seeing, and touch are each associated not only with distinct representations, but also with distinct subjective phenomenal experiences. The phenomenal experiences tend to be similar within sensory modalities and different between sensory modalities. According to the experience-based view, it is because sensory experiences are phenomenally similar to and different from each other that we have distinct senses. A sense is individuated by the types of phenomenal experience to which it gives rise. A challenge for the view is to say what are the relevant similarities and differences in phenomenal experience. Experiences within a sensory modality are not all alike and those between sensory modalities are not all dissimilar. Which phenomenal

similarities and differences matter for individuating the senses, and why are they important?

The third approach is stimulus-based. Different senses involve responding to proximal stimuli of different physical kinds. Seeing involves reacting to electromagnetic waves between 380 nm and 750 nm. Hearing involves reacting to air pressure waves in the ear canal. Smelling involves reacting to airborne chemicals in the nose. According to the stimulus-based view, the reason why the senses are distinct is because they involve responses to different physical types of proximal stimulus. A sense is individuated by type of proximal stimulus to which the organism reacts. A challenge for the view is that the same proximal stimulus could be associated with different senses. For example, the same pressure wave in the air may be processed by an organism for hearing and for echolocation.

The final approach is organ-based. Different senses tend to be associated with different sense organs. Seeing involves the eyes, hearing involves the ears, smelling involves the nose. Each sense organ contains physiologically distinct receptor cells. According to the organ-based view, the reason why the senses are distinct is because they employ distinct sense organs. A sense is individuated by its associated sense organ. A challenge for the view is that the same sense organ (e.g., the ear) could be used for two different senses (e.g., hearing and echolocation).

The four proposals prescribe different revisions to folk assumptions about the senses in light of scientific data. The task facing philosophers is to determine which, if any, of these views is correct. Nudds (2004) argues that we should not endorse any of them. His claim is that individuation of the senses is context-dependent. No single account individuates the senses across all contexts. Different criteria apply in different contexts depending on our interests. Macpherson (2011) argues that the senses should be individuated context-independently. She claims that all of the criteria above matter. All contribute jointly to individuating the senses. The four proposals can be used as a multidimensional metric on which any possible sense can be judged. The clustering of an organism's cognitive capacities across multiple dimensions, rather than on a single dimension, determines how its senses should be individuated.

Let us turn to our second example of a dispute about individuation. The hypothesis of extended cognition (HEC) asserts that human mental life sometimes extends outside the brain and takes place partly inside objects in the environment, such as notebooks or iPhones (Clark and Chalmers, 1998). Disagreements about whether HEC is true have taken the form of a disagreement about the individuation conditions of human mental states and processes.

The best way to understand HEC is to start with the weaker claim known as *dis-*

tributed cognition (Hutchins, 1995). An advocate of distributed cognition claims that human cognitive capacities do not always arise solely in, and should not always be explained exclusively in terms of, neural mechanisms. The mechanisms behind human cognitive capacities sometimes include bodily and environmental processes. The brain is not the sole mechanism responsible for our cognitive abilities, but is only part – albeit a large part – of a wider story. The brain recruits environmental and bodily resources to solve problems. Recruiting these resources allows humans to do more than they could otherwise, and to work more quickly and reliably. Brains off-load work onto the body and environment. Sometimes the off-loading is under conscious control: for example, when we consciously decide to use a pen, paper, or our fingers to solve a mathematical problem. Sometimes the off-loading is not under conscious control: for example, when we use our eye gaze to store information (Ballard et al., 1997; W. D. Gray and Fu, 2004). Distributed cognition is the claim that *distributed* information-processing strategies figure in the best explanation of, and causal story behind, some human cognitive accomplishments.

HEC is a stronger claim than distributed cognition. According to HEC, not only do brains recruit environmental resources to solve problems, but those non-neural resources, when they have been recruited, also *have mental properties*. Parts of the environment and the body, when employed in distributed cognition strategies, have just as much claim to mental or cognitive status as any neural process. Against this, the hypothesis of embedded cognition (HEMC) accepts the distributed-cognition claim about the exploitation of bodily and environmental resources, but rejects HEC's assertion about the body and environment having mental properties (Rupert, 2004; Rupert, 2013). According to HEMC, non-neural resources, despite figuring in the explanation and causal story behind some cognitive accomplishments, do not have mental properties. Only neural processes have mental or cognitive properties.

How is this about individuation? HEC is, in essence, a claim about the individuation of mental kinds. HEC claims that the causal flow in human cognition should be individuated in such a way that the neural and non-neural parts instantiate a single kind – a mental kind. This is not to say that there are no differences relevant to psychology between the neural and non-neural parts. HEC's claim is merely that the neural and non-neural parts jointly satisfy a condition sufficient for them to instantiate a single mental kind. In contrast, HEMC's claim is that the causal flow in human cognition should be individuated so that the neural and non-neural parts fall under different kinds – mental and non-mental respectively. This is not to say that there are no kinds of interest to psychology that both instantiate. Rather, it is to say that whatever kinds they instantiate, they jointly fail to meet the condition required for them to instantiate a single mental kind. HEC and HEMC disagree, it cases of distributed cognition, about how to individuate mental properties across the causal flow.

Which view is right: HEC or HEMC? To answer this, we need to agree on the minimal condition, mentioned earlier, for a physical process to instantiate a mental kind. There are two main proposals on this score. The first is functionalist. On this view, a physical state or process is mental provided it has the right functional profile. I have argued elsewhere that the functionalist proposal decisively favors HEC (Sprevak, 2009). If one does not accept HEC granted functionalism, one concedes the chauvinism about the mind that functionalism was designed to avoid. The second proposal is based on explanatory pay-off to cognitive science. On this view, a physical state or process is mental just in case it fits with best (current or future) cognitive science to treat it as such. I have argued elsewhere that considerations of explanatory value regarding cognitive science are toothless to decide between HEC and HEMC. Cognitive science could continue to be conducted with little or no loss either way (Sprevak, 2010). The dispute between HEC and HEMC is a case in point for which a question about individuation of mental states and processes cannot be answered by a straightforward appeal to scientific practice. Work in philosophy of the psychological and cognitive sciences needs to draw on a wide range of considerations to settle such questions about individuation.

6 Conclusion

We have surveyed four types of task in philosophy of the psychological and cognitive sciences: How should we interpret our scientific theories? How should we precisify our theoretical concepts? How do our theories or methodologies fit together? How should our cognitive states, processes, and capacities be individuated? We have focused on only some of current work in philosophy of the psychological and cognitive sciences. Work we have not covered includes proposals for psychological mechanisms and architectures (for example, Apperly and Butterfill, 2009; Grush, 2004); analysis of scientific methodology in the psychological sciences (Glymour, 2001; Machery, 2013); examination of key experimental results in the psychological sciences (Block, 2007; Shea and Bayne, 2010); and analysis of folk psychological concepts (H. M. Gray, K. Gray and Wegner, 2007; Knobe and Prinz, 2008).

Acknowledgements

I would like to thank faculty and graduate students at the University of Pittsburgh, as well as David Danks, for helpful suggestions on this entry.

Bibliography

Apperly, I. A. and S. A. Butterfill (2009). "Do humans have two systems to track belief and belief-like states?" In: *Psychological Review* 116, pp. 953–970.

Baars, B. (1997). *In the Theater of Consciousness*. Oxford: Oxford University Press.

Ballard, D. H., M. M. Hayhoe, P. Pook and R. Rao (1997). "Deictic codes for the embodiment of cognition". In: *Behavioral and Brain Sciences* 20, pp. 723–767.

Bechtel, W. (1998). "Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science". In: *Cognitive Science* 22, pp. 295–318.

Block, N. (1995). "On a confusion about a function of consciousness". In: *Behavioral and Brain Sciences* 18, pp. 227–247.

— (2007). "Consciousness, accessibility, and the mesh between psychology and neuroscience". In: *Behavioral and Brain Sciences* 30, pp. 481–548.

Bowers, J. S. and C. J. Davis (2012). "Bayesian just-so stories in psychology and neuroscience". In: *Psychological Bulletin* 128, pp. 389–414.

Clark, A. and D. J. Chalmers (1998). "The extended mind". In: *Analysis* 58, pp. 7–19.

Clark, A. and J. Toribio (1994). "Doing without representing?" In: *Synthese* 101, pp. 401–431.

Colombo, M. and P. Seriès (2012). "Bayes on the brain—On Bayesian modelling in neuroscience". In: *The British Journal for the Philosophy of Science* 63, pp. 697–723.

Coltheart, M. (1999). "Modularity and cognition". In: *Trends in Cognitive Sciences* 3, pp. 115–120.

Craver, C. F. (2007). *Explaining the Brain*. Oxford: Oxford University Press.

Danks, D. (2008). "Rational analyses, instrumentalism, and implementations". In: *The Probabilistic Mind: Prospects for Rational Models of Cognition*. Ed. by N. Chater and M. Oaksford. Oxford University Press, pp. 59–75.

Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. London: Penguin Books.

Dehaene, S. and J.-P. Changeux (2004). "Neural mechanisms for access to consciousness". In: *The Cognitive Neurosciences, III*. Ed. by M. Gazzaniga. Cambridge, MA: MIT Press, pp. 1145–1157.

Dehaene, S., J.-P. Changeux, L. Naccache, J. Sackur and C. Sergent (2006). "Conscious, preconscious, and subliminal processing: A testable taxonomy". In: *Trends in Cognitive Sciences* 10, pp. 204–211.

Egan, F. (2010). "Computational models: a modest role for content". In: *Studies in History and Philosophy of Science* 41, pp. 253–259.

Eliasmith, C. (1997). "Computation and dynamical models of mind". In: *Minds and Machines* 7, pp. 531–541.

— (2013). *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford: Oxford University Press.

Elsabbagh, M. and A. Karmiloff-Smith (2006). "Modularity of mind and language". In: *The Encyclopedia of Language and Linguistics*. Ed. by K. Brown. Vol. 8. Oxford: Elsevier, pp. 218–224.

Ernst, M. O. and M. S. Banks (2002). "Humans integrate visual and haptic information in a statistically optimal fashion". In: *Nature* 415, pp. 429–433.

Feldman, J. (2012). "Symbolic representation of probabilistic worlds". In: *Cognition* 123, pp. 61–83.

Firestone, C. and B. J. Scholl (2016). "Cognition does not affect perception: Evaluating the evidence for "top-down" effects". In: *Behavioral and Brain Sciences* 39, E229.

Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.

— (2000). *The Mind Doesn't Work That Way*. Cambridge, MA: MIT Press.

Glymour, C. (2001). *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. Cambridge, MA: MIT Press.

Gray, H. M., K. Gray and D. M. Wegner (2007). "Dimensions of mind perception". In: *Science* 315, p. 619.

Gray, W. D. and W.t.- Fu (2004). "Soft constraints in interactive behavior". In: *Cognitive Science* 28, pp. 359–382.

Griffiths, T. L., N. Chater, D. Norris and A. Pouget (2012). "How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012)". In: *Psychological Bulletin* 138, pp. 415–422.

Grush, R. (2004). "The emulator theory of representation: Motor control, imagery, and perception". In: *Behavioral and Brain Sciences* 27, pp. 377–442.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Jones, M. and B. C. Love (2011). "Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition". In: *Behavioral and Brain Sciences* 34, pp. 169–231.

Keijzer, F. A. (1998). "Doing without representations which specify what to do". In: *Philosophical Psychology* 11, pp. 269–302.

Knill, D. C. and A. Pouget (2004). "The Bayesian brain: the role of uncertainty in neural coding and computation". In: *Trends in Neurosciences* 27, pp. 712–719.

Knobe, J. and J. Prinz (2008). "Intuitions about consciousness: experimental studies". In: *Phenomenology and Cognitive Science* 7, pp. 67–85.

Koch, C. and N. Tsuchiya (2006). "Attention and consciousness: Two distinct brain processes". In: *Trends in Cognitive Sciences* 11, pp. 16–22.

Laureys, S., M. Boly, G. Moonen and P. Maquet (2009). "Coma". In: *Encyclopedia of Neuroscience* 2, pp. 1133–1142.

Lehky, S. R. and T. J. Sejnowski (1988). "Network model of shape-from-shading: neural function arises from both receptive and projective fields". In: *Nature* 333, pp. 452–454.

Machery, E. (2013). "In defense of reverse inference". In: *The British Journal for the Philosophy of Science* 65, pp. 251–267.

— (2015). "Cognitive penetrability: A no-progress report". In: *The Cognitive Penetrability of Perception*. Ed. by J. Zeimbekis and A. Raftopoulos. Oxford: Oxford University Press, pp. 59–74.

— (ms). "Philosophy Within its Proper Bounds". manuscript.

Macpherson, F. (2011). "Individuating the senses". In: *The Senses: Classic and Comtemporary Philosophical Perspectives*. Ed. by F. Macpherson. Oxford: Oxford University Press, pp. 3–43.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

McDowell, J. (2010). "Tyler Burge on disjunctivism". In: *Philosophical Explorations* 13, pp. 243–255.

Nudds, M. (2004). "The significance of the senses". In: *Proceedings of the Aristotelian Society* 104, pp. 31–51.

Oppenheim, P. and H. Putnam (1958). "Unity of science as a working hypothesis". In: *Concepts, theories, and the mind–body problem*. Ed. by H. Feigl, M. Scriven and G. Maxwell. Vol. II. Minnesota studies in the philosophy of science. Minneapolis, MN: University of Minnesota Press, pp. 3–36.

Piccinini, G. (2008). "Computation without representation". In: *Philosophical Studies* 137, pp. 205–241.

Pouget, A., J. M. Beck, W. J. Ma and P. E. Latham (2013). "Probabilistic brains: Knows and unknowns". In: *Nature Neuroscience* 16, pp. 1170–1178.

Prinz, J. (2006). "Is the mind really modular?" In: *Contemporary Debates in Cognitive Science*. Ed. by R. Stainton. Oxford: Blackwell, pp. 22–36.

Putnam, H. (1975). *Mathematics, Matter and Method, Philosophical Papers, volume 1*. Cambridge: Cambridge University Press.

Qian, N. (1997). "Binocular disparity and the perception of depth". In: *Neuron* 18, pp. 359–368.

Rescorla, M. (2015). "Bayesian perceptual psychology". In: *The Oxford Handbook of Philosophy of Perception*. Ed. by M. Matthen. Oxford University Press, pp. 694–716.

— (2016). "Bayesian sensorimotor psychology". In: *Mind and Language* 31, pp. 3–36.

Rupert, R. D. (2004). "Challenges to the hypothesis of extended cognition". In: *The Journal of Philosophy* 101, pp. 389–428.

— (2013). "Memory, natural kinds, and cognitive extension; Or, Martians don't remember, and cognitive science is not about cognition". In: *Review of Philosophy and Psychology* 4, pp. 25–47.

Samuels, R. (2005). "The complexity of cognition: Tractability arguments for massive modularity". In: *The Innate Mind: Vol. I, Structure and Contents*. Ed. by P. Carruthers, S. Laurence and S. P. Stich. Oxford: Oxford University Press, pp. 107–121.

Schöner, G. (2008). "Dynamical systems approaches to cognition". In: *Cambridge Handbook of Computational Psychology*. Ed. by R. Sun. Cambridge: Cambridge University Press, pp. 101–126.

Sellars, W. (1963). *Science, Perception and Reality*. London. Routledge & Kegan Paul.

Shea, N. and T. Bayne (2010). "The vegetative state and the science of consciousness". In: *The British Journal for the Philosophy of Science* 61, pp. 459–484.

Sprevak, M. (2009). "Extended cognition and functionalism". In: *The Journal of Philosophy* 106, pp. 503–527.

— (2010). "Inference to the hypothesis of extended cognition". In: *Studies in History and Philosophy of Science* 41, pp. 353–362.

— (2013). "Fictionalism about neural representations". In: *The Monist* 96, pp. 539–560.

Thelen, E., G. Schöner, C. Scheier and L. B. Smith (2001). "The dynamics of embodiment: A field theory of infant perseverative reaching". In: *Behavioral and Brain Sciences* 24, pp. 1–86.

Van Gelder, T. (1995). "What might cognition be, if not computation?" In: *The Journal of Philosophy* 91, pp. 345–381.